When scientists play: how toy models in science help us understand the world

Dominik Hangleiter

Ludwig-Maximilians-Universität München

Bachelorarbeit

vorgelegt an der Fakultät für Philosophie, Wissenschaftstheorie und Religionswissenschaft

Ludwig-Maximilians-Universität München

Januar 2014

ABSTRACT

In science toy models are developed and explored ubiquitously. Their study aims either at making a puzzling phenomenon comprehensible or at learning about the core mechanisms behind more complex models. Model building is a core skill of the theoretician. Not *any* model is a good tool, it is only those models that meet certain standards. Models should contain the essence of a physical problem, and be mathematically tractable. Toy models meet these standards. In this thesis I argue that with a toy model at hand a theorist may set out to understand the phenomenon under study. But *how* do toy models function in order to help theorists gain understanding?

To answer this question I take three steps. The first one is an illustrative one: I analyse three case studies, the Lotka-Volterra model of predator-prey dynamics, Schelling's chequerboard model of residential segregation and the Ising model of second-order phase transitions. From these case studies I extract four functions that toy models can perform, namely representation of a type of phenomenon, theoretical exploration, pedagogical usage, and substantiating arguments that effect hypothesis revision. Second, as a preliminary argument I analyse in virtue of what toy models are particularly apt to perform these functions. I argue that the core properties in furnishing the representative functions are their high level of idealisation and their mathematical structure. The exploratory and pedagogical functions require the mathematical tractability (exact solvability) and the open domain of application that are characteristic of toy models. Thus the properties and functions of toy models are linked.

The functions toy models and my analysis of how their properties afford these functions form the basis of my main thesis: by studying a toy model a scientist can understand the type of phenomenon it represents. Thus she gains partial understanding of the elements of the type. Prior to the argument I need to clarify the notions of explanation, understanding and intuition. With the notion of understanding as qualitative knowledge at hand the thesis follows. In order to make sense of how the sharable representation of a toy model results in understanding as a private cognitive category I draw on cognitive science. In particular, I show how the model of mental models for toy modelling as a cognitive process suits the problem perfectly and is able to connect the core aspects of understanding: recognition of a pattern, and association of a model with other concepts. With the distinction between the sharable external and the private internal representation of a toy model can we make sense of how toy models function as guides for scientific progress.

ACKNOWLEDGEMENTS

First and foremost, I want to thank Stephan Hartmann for his patient and demanding supervision and guidance not only in the process of writing my thesis but already in the search for a topic. Stephan has the unequalled talent to make me see errors or gaps of thought by asking ever-so-simple down-to-earth questions. Thus he has challenged me to push myself further and further toward higher standards of precision in the formulation of my thoughts. On the other hand, his open and uncomplicated nature made me feel welcome at the MCMP. The many talks, conferences and activities on a huge variety of topics sparked my enthusiasm for the approach to philosophy taken at the MCMP. In particular, from the conference on "Reduction and Emergence in the Sciences" at the CAS did I gain much inspiration and insight.

I also want to thank Lorenzo Casini, who guided my path through the djungle of philosophy in search for a topic. Lorenzo was always open to discussion, and provided me with helpful comments and literature. My proofreaders Lorenzo, Robert Darius, Tobias and Andreas Hangleiter have done a great job at putting their fingers on bad formulation and inconsistencies.

As always, my parents sustained the emotional aspects of thesis writing - the highs and lows, frustration and enthusiasm. They are mirror and emotional cushion for me. My gratitude for their ever-lasting support cannot be put into words.

Last but not least I want to thank my flatmate Matthias, who takes me down to earth in daily discussions, and in his reflecting and clear manner influences my view on the world. Special thanks go to Cléo on the other side of the world, and Andreas Reichel on the other side of the academic spectrum. The emails and conversations with them always put a smile on my face and let me dive into another world.

Dominik Hangleiter, München, Januar 2014

CONTENTS

1. INTRODUCTION

Ever since Galileo have scientists constructed models. By building models scientists render the diverse phenomena that we observe in the world tractable in systematic investigation. Besides experimentation and theorising, model building is one of the primary activities in science. Models are constructed across all sciences. In physics the spectrum ranges from the simple model of the ideal gas to highly complex models of solids, the model of a physical pendulum to models of star formation. In ecology there are models for predator-prey interaction, in economics models of the ideal market or agent-based models, a method of investigation that is even used in philosophy. In every subject scientists have a different approach to modelling, different criteria according to which models are assessed, and a different language in which it is shared.

Models have also been a “hot topic” in the philosophy of science literature for a long time, starting with Pierre Duhem and continuing up to today. There has been an abundance of case studies and general analyses of models, especially in economics (e.g. Gibbard and Varian, 1978, Morgan, 2001) and physics (e.g. Bokulich, 2011, Cartwright, 1983, Hartmann, 1995, 1998, 2001, Hughes, 1997, 1999) regarding different kinds of models[1]. While there has been much progress, the studies are manifold with respect to the questions they ask. The term “model” is ubiquitously and often handwavingly used; the concept of a model is blurry. This is why an overarching systematic account of their role and function is nearly impossible. In particular, this concerns their distinction from other scientific activities, theorising and experimenting. While theories have the aim to unify as many phenomena as possible under as few principles as possible, experiments are performed to provide empirical data to justify theories. Models navigate somewhere between the two; they *mediate* between the realm of theory and the world (Morgan and Morrison, 1999). Moreover, they can perform functions of both theory and experiment. As a minimal characterisation let us take a scientific model to be a set of assumptions about a concrete system or phenomenon in the world (the target) that are inspired by, isolated from, or abstracted from the target’s properties.

The most intriguing and so-far almost overlooked class of models is the “scientist’s favourite”, minimal or toy models, that is, models that are highly idealised and based on very few assumptions. Toy models are omnipresent in the literature of all sciences and hard to be distinguished sharply from other types of models. Loosely, the name suggests the function: in a toy model scientists “play around” with parameters, assumptions, and possible interactions between components of the

[1] A long list of what kinds of models there are is given in (Frigg and Hartmann, 2012).

model without constraint or regard to what every single change may correspond to in the world[2]. Sometimes they are used as an illustrative "ideal" study in order to understand a related more complex model (which may not be susceptible to analytic treatment). Sometimes a toy model is proposed and studied on its own to argue for the relevance of its core mechanism to understanding a type of phenomenon in the world. They can even become "Drosophilae" of their respective sciences, i.e. paradigmatic examples for the application of a theory. So on the one hand toy models are relevant heuristically, on the other hand they are tools to learn about another more complex theoretical construct.

What makes toy models both intriguing and elusive is precisely the tension between their unclear relation to the world and their pervasive use in all sciences, which suggests that they *are* regarded as useful. This tension is due to their near-absurd level of idealisation on the one hand, and the unimportance of empirical data in their manipulation on the other hand. Would not a "perfect" model that exactly matches the features of an observed phenomenon yield a much better explanation than a toy model that only crudely captures aspects of a phenomenon, a model whose relation to the target phenomenon is not at all clear? Indeed, this view was held in science until the middle of the 20th century; call it the "traditional approach" to theorising. However, the more complicated the system to be described becomes, the harder it will be to construct a model that fits *all* of its features perfectly well. Moreover, the more detail is added the less comprehension over the target phenomenon is attained; we cannot *grasp* the implications of the model anymore. Therefore, in the modern approach a theorist is rather expected to "*understand* what is going on and to elucidate which are the crucial features of the problem" (Fisher, 1983, p. 47, emph. in the original). Models are *the* crucial tool to do so. In particular, scientists "should be prepared to look even at rather crude models" that are so highly simplified that they are "'mere caricature[s]' of reality" (*ibid.*, p. 47). A good scientist is considered one who is ready "to construct or adapt a model or a toy [model] to suit the problem on hand" (Ziman, 1965, p. 1192). According to the view expressed by these two theorists, scientific theorising is all about extracting the *essential* features of a phenomenon by studying models. Because they are so central to science, a philosophical account of toy models is pressing.

In this thesis I want to give a more precise characterisation of toy models and understand aspects of their role in science. In particular, I investigate which functions toy models perform in the scientific enterprise by analysing examples that are uncontroversially considered to be toy models. Subsequently I aim to understand how they can form and furnish our understanding of the world. In particular, this amounts to answering the questions what it is about toy models that lets them perform the various functions, and how the functions they perform afford an understanding of the world. The answer will tell us why it is that toy models are used so diversely, across all disciplines and in an ever increasing intensity *although* their connection to the world is (at the very least) vague and dubious. In a nutshell, this

[2] Cf. (Hausman, 1992).

thesis is about how the toys of science become the vehicles of our understanding of the world. Answering this question will be a further step toward understanding the relation between the external world and our scientific description of it.

1.1 Characterisation of toy models

Before beginning to answer these questions let me flesh out some of the terms that are crucial to the discussion of toy models. As already stated above, I take a *scientific model* to be a set of assumptions about a concrete system or phenomenon (the target), which are approximations, idealisations, or abstractions of a subset of the target's properties. Models fall into two classes: models of theory, that is, models that are concrete applications of a theory, and phenomenological models that are independent of theories (Hartmann and Bailer-Jones, 1999).

Toy models Few suggestions have been made for how to characterise toy models (Arnold, 2010, Frigg and Hartmann, 2012, Hartmann, 1995), but the authors agree that toy models do not directly relate to a particular target system. I comply with this feature, but want to give a more refined characterisation, which will turn out to be natural in the course of this thesis. A *toy model* is a model that is (1) highly idealised, and (2) exactly solvable, or, if not mathematically formulated, tractable without approximations. It is (3) constructed and manipulated to gain *qualitative* knowledge about the model behaviour, but not to *quantitatively* predict the behaviour of a target, and (4) has no specified domain of application. In many cases, (5) a toy model also comprises new hypotheses or variations of old ones (Ziman, 1965, p. 1191). This characterisation is vague, and must be so to comply with the usage of the term in the scientific literature. Some toy models are phenomenological models, some models of theory. As I will show below, two main functions of toy models are the representation of types of phenomena as their targets, and the exploration of theoretical tools through their use. They are particularly important as tools to obtain understanding about their targets.

Idealisation, abstraction and approximation Some more words on the terms that I have used in the above characterisation are in order here. The three processes that are involved in all model building activities are approximation, idealisation and abstraction. Conforming with McMullin (1985) I take an *idealisation* to be the "deliberate simplifying of something complicated (a situation, a concept, etc.) with a view to achieve at least a practical understanding of that thing" (p. 248). Idealisation is closely connected to model building. In particular, I take idealisation to be what McMullin calls "Galilean idealisations" being either "construct idealisations" or "causal idealisations". Construct idealisations are deliberate simplifications or omissions of features in the *construction* of a model that is then studied as a substitute for the real system. Construction is possible both in thought and in the real world[3]. Causal or subjunctive idealisation is the isolation of individual lines of causes that are relevant to a phenomenon. Again, they can be both performed in the real world, as experiments, or in thought as counterfactual suppositions (subjunctives).

The relation between idealisation and isolation is unclear: one might say that in order to omit a feature in a model we need to have already isolated that feature in thought. From this perspective, having constructed a model is necessary for isolation. The question whether idealisation or isolation is "prior" comes down to the problem with the chicken and the egg. Mäki (2009) recognises the problem and concludes that "*both* construction *and* isolation are involved in modelling" (p. 32, emph. in the original), and with that in the process of idealisation. Thus idealisation, model construction, and isolation are not separable but interrelated processes.

Though not uncontroversial, *abstraction*[4] according to Cartwright (1989), is the "stripping away" of features from an object in our imagination. It may be achieved in a process that Husserl (1985) characterised as "eidetic reduction". He vividly described it: we find the characteristic features of an object by varying it in our imagination. By comparing all the variations we may extract its essential features as their common denominator. Another view on abstraction is defended by Morrison (2009b), who takes an abstraction to be a the introduction of a (mathematical) feature to the model that is *necessary* to obtain results about the model, as opposed to idealisation, which distorts features in the construction process. However, it is unclear whether such features exist at all[5]. Therefore, I will comply with Cartwright's and McMullin's view on abstraction and idealisation, respectively, and take idealisation to be the *positing of features* in the construction of a model, and abstraction the *stripping away* of inessential features of a kind of object in order to obtain its characteristic features that are common to the set of particulars.

By an *approximation* I mean both the relation between a model and its target, and a mathematical operation that is performed with the aim to make a model tractable. A model can be said to "approximately resemble its target", or it may be necessary to "perform approximations for its solution". Approximation in the latter sense has a pragmatic function. It can be contrasted with idealisation, which has an additional cognitive function, namely to extract, or isolate a feature from a phenomenon and explore its consequences (Hartmann, 1998). Approximation and idealisation are also related in that the more idealisation is performed in the construction process of a model, the less approximation will be necessary for its solution (Batterman, 2002). On the other hand, the higher the level of idealisation of a model is the cruder it approximates its target.

[3] Think of scale models.

[4] Abstraction is also called "Aristotelian idealisation" (Frigg and Hartmann, 2012).

[5] According to Morrison (2009b) an example of an abstract feature of a model is the thermodynamic limit in which the number of particles approaches infinity. The thermodynamic limit appears to be necessary for the two-dimensional Ising model to exhibit a phase transition.

Phenomena and representation Approximation, idealisation and abstraction are performed in order to model *phenomena* or *systems* in the real world[6]. In order to understand the relation of a model to its target phenomenon, it must be clarified what a phenomenon is. On the one hand the term "phenomenon" can refer to a *particular* fact or event. On the other hand it can refer to a *general* fact or pattern of events (Batterman, 2002). For example, we can observe the particular event of an apple falling from a tree, but there is also the general fact that apples always fall from the tree they hang on. In science, model building is mainly concerned with phenomena as general facts about the world. That is, science aims to abstract from the particular properties of individual phenomena and give an account of a phenomenon in general. Historical and social research on the other hand is interested in the specifics of a particular phenomenon (Weber, 1991b). For example, a Historian might be interested in what it was about the particular course of history that made World War I come about. In both cases, however, researchers search for the essential features that bring about the phenomenon in question. For a general phenomenon these are features that obtain in all of its instances and bring it about repeatedly. For a particular phenomenon these are the specific features that distinguish this phenomenon from all others. Unless specified differently I use "phenomenon" in its general sense, and take it to be the set of its particulars.

After another step of abstraction or idealisation one may extract *generic features* of a class of phenomena, that is, features that are shared by a set of phenomena (i.e. a set of sets of particular events). For example, it might be said that a generic feature of apple trees, pear trees and plum trees is their common property that their fruits fall to the ground. We may abstract further to a level such that still the essential features of the phenomenon (fruits falling to the ground) obtain, but the particulars of the individual instances (shape of the tree, weight, colour, taste of the fruits, etc.) have become irrelevant. A generic feature F defines a *type of phenomenon* P_F, the set of all phenomena that exhibit F. Of course, the distinction between a phenomenon (the set of particulars) and a type of phenomenon (a set of phenomena) is vague. It depends on the conventions of what scientists individuate or study as individual phenomena. Consequently, there can be *types* of types of phenomena, too. For example, ferromagnetism and critical opalescence are phenomena, a phase transition is a type of phenomenon that contains ferromagnetism and critical opalescence. Ferromagnetism, in turn, may be divided up into the phenomena of ferromagnetism in different materials[7].

By saying that a model *represents* its target, I mean that the components of a model are interpreted by, that is, identified with, properties of a real-world system or phenomenon. In particular, a representation cannot be true or false, but only

[6] It might be argued that what is modelled are systems. If the reader agrees that a system is the specific arrangement of a set of objects, or the general pattern or such an arrangement, all of my claims and analyses hold true for systems, too. Therefore I will stick with the term "phenomenon" without loss of generality.

[7] Ferromagnetism and critical opalescence are represented by the Ising model as discussed in Section 2.3.

adequate to a certain degree. A feature A_1 of a system A is *represented* well by a system B, if system B exhibits a feature B_1 that *resembles* A_1 to large enough degree by some similarity relation (Mäki, 2009, Weisberg, 2007). If all features $A_1, \ldots, A_n$ of A are represented adequately by B, B may be said to fully represent A.

In particular, in order to establish whether or not a model adequately represents its target, the behaviour of the model has to be compared to an adequately (e.g. mathematically) formulated dynamics of the target system. By "behaviour" I mean the dynamics of the model. A model exhibits dynamics in that some model parameter of interest depends non-trivially on a parameter or variable assigned to a particular property of the model, which follows from one or more of its assumptions.

1.2 Structure of the thesis

In this thesis I will show that toy models are crucial in the modern approach to science in that they provide understanding. Specifically, they are tools to extract the essential features of phenomena. *How* are they used thusly and how are they different from "normal models"? Can we not reduce the study of toy models cannot be reduced to the study of just any models. The key aspect that distinguishes toy models from "normal" models, is the way they are studied. While "normal models" are idealised and may be exactly solvable, too, the way toy models are *used* sharply distinguishes them. Their usage aims principally at theoretical exploration not regarding concrete application. This function is particularly important in two contexts: on the one hand toy models are very often proposed in the context of more complex models in order to provide some insight into the mechanisms involved in these complex models[8]On the other hand toy models are proposed for phenomena for which there is no theoretical account at all[9]. In this latter sense a toy model are a heuristic tool, a first yardstick used to gain some comprehension of a phenomenon and put it in a context. Here I will put the focus of my thesis.

Previous studies, in particular the debate on whether models can be regarded as fictions (e.g. Suárez, 2009) have focused on the representative aspects of modelling. In my thesis I want to take seriously the dynamical aspect of modelling, that is, the *usage* of models as tools. Thereby I hope to better understand their role in science. To get a grip on this aspect in Section 2 I will study three well known toy models from different sciences that are used in the yardstick-sense: the Lotka-Volterra model of predator-prey dynamics in biology, Schelling's chequerboard model of residential segregation in economics, and the Ising model of critical-point phenomena in physics. Not only will I present the motivation, assumptions, formulation, and solution of these models, but also provide rough sketches of their pervasive uses in the respective subjects.

[8] An example of this usage of toy models is provided by Mendoza-Arenas et al. (2013). They devote a good part of their study of dephasing-enhanced transport in quantum systems to the study of a toy model and use it to show that the mechanism behind this effect obtains generically and is thus "expected to persist in more realistic driven systems" (p. 1).

[9] Detailed examples for toy models in this context are given in Section 2.

Subsequently, I will analyse these models in Section 3. First, I will extract and illustrate the functions that these toy models perform. Thereafter, I will analyse what is characteristic about these toy models that lets them perform these functions. In particular, I show that two key functions of toy models are the representation of a type of phenomenon and exploration of theoretical tools. I argue that they can perform their representative function only because of their high level of idealisation. Their exploratory function additionally requires the exact solvability of toy models. Moreover, the low constrains both with respect to how well they resemble their target and what this target is are conducive to exploration. The questions I ask in this section are therefore “Which functions do toy models perform?” and “What is it about toy models in virtue of which they can perform these functions?”.

In Section 4 I will then use the answers to the questions posed in Section 3 to argue for my main thesis: by their exploratory and representative functions can toy models be used to obtain understanding. Specifically, I begin by briefly reviewing the literature on the epistemic import of minimal models in Section 4.1. In a comprehensive account of how toy models are used to obtain understanding both their representative and the exploratory functions must be accommodated and made sense of. Such an account requires a clarification and distinction of the notions of explanation, understanding, and intuition, which I provide in Section 4.2. This clarification will pave the way to understand “understanding with toy models” in Section 4.3. I argue that by studying a toy model one can gain understanding of the type of phenomenon that toy model represents. This thesis is based on the assumption that through the exploration of a toy model one can gain intuition about the model, that is, acquire the ability to recognise counterfactual patterns that are characteristic for this model. In order to flesh out this assumption I will draw from cognitive science. I argue that mental models are a good model for the way toy models act both as a sharable external interface on which an entire community can perform manipulations and as a “manual” for the construction of an internal private representation in the mind of individual scientists. Within this model I also show how in a process of repeated analysis and adaptation of toy models we can improve this understanding. Hence, in this section I ask: “What is understanding, and how can toy models provide understanding?”

Finally, in Section 5, I conclude and hint at possible implications of the account presented, in particular regarding the epistemic import of quantum simulations.

2. CASE STUDIES

2.1 Lotka-Volterra model

The Lotka-Volterra model was proposed by Lotka (1910), and independently by Volterra (1926b) as a model of reaction kinetics and predator-prey dynamics, respectively[1]. Volterra was inspired to construct his model by a peculiar fact that was noticed after the first world war: during the war the population of predators in the Adriatic Sea (sharks) had increased, and the population of prey (cod, squid, lobster) decreased, while the level of fishing during the war had been lower than usual. How were these two facts connected? To explain this phenomenon, he proposed a mathematical model for the dynamics of interacting populations. In particular, he motivated his investigation by stating that these dynamics were "important theoretically, but often [...] also [of] practical importance" (Volterra, 1931, p. 4) for fisheries, in the context of agriculture, but also infective diseases. He aimed at extracting the external environmental (such as the changing of the seasons), and internal parameters (such as reproduction rates of the fish populations) that bring about the periodic dynamics of the populations that was observed.

Volterra explicitly *constructed* the mathematical model based on three assumptions about species (Volterra, 1931):

(i) The population of a species is a continuous time-dependent variable.

(ii) Hatchings and deaths are distributed evenly (and continuously) over time and proportional to the population size.

(iii) Populations are homogeneous (without age and size).

Starting from these assumptions he could then add and adapt the mathematical representations of the different factors (i.e. terms in the differential equations) that he might be interested in.

The particular case that related to his original question is that of one predator and one prey species. In this case the model under the above a mathematical description of their populations is given in terms of the two coupled differential equations (Volterra, 1931, p. 9),

$$\frac{dN}{dt} = (\epsilon_1 - \gamma_1 P)N, \qquad \frac{dP}{dt} = (-\epsilon_2 + \gamma_2 N)P, \tag{2.1}$$

[1] I will base the discussion of this model on the translation of (Volterra, 1926b) in (Volterra, 1931) and the international research article (Volterra, 1926a).

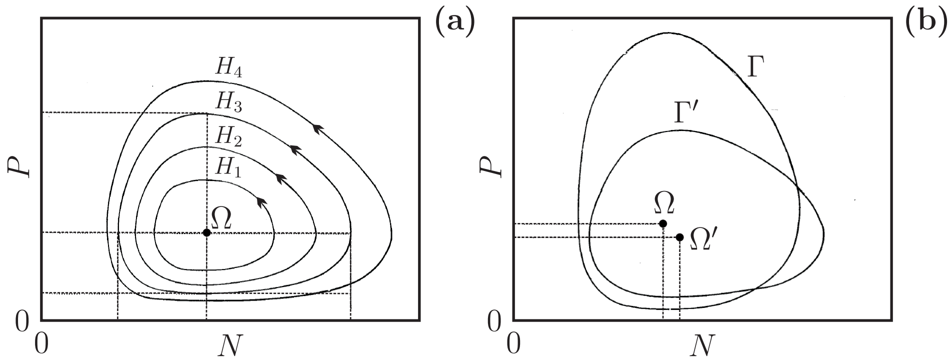


Fig. 2.1: **(a)** (N,P) phase plane trajectories from Equation 2.2 with $1 + \epsilon_2/\epsilon_1 < H_1 < H_2 < H_3 < H_4$. Their centre $\Omega = (\bar{N}, \bar{P})$ remains invariant under changes of H and determines the average populations. Arrows indicate the direction of movement along the trajectories with time. **(b)** A change in the reproduction rates of the two species results in a shift of the averages $\Omega \to \Omega'$ and the curves $\Gamma \to \Gamma'$. Specifically, for $\epsilon_1' < \epsilon_1$ and $\epsilon_2' > \epsilon_2$, which corresponds to externally reducing the numbers of predators and prey proportional to their population sizes, the averages behave as $\bar{N}' < \bar{N}$ and $\bar{P}' > \bar{P}$. Both figures are adapted from (Volterra, 1931).

where N (P) correspond to the populations of prey (predators), ϵ_i and γ_i are strictly positive constants. The *interpretation* of ϵ_i is the rate of population increase (decrease), while γ_i parametrise the interaction between predator and prey species, and thus correspond to "the aptitude of the [prey] to defend itself" (Volterra, 1931, p. 10) (the efficiency of the predators to kill the prey) all for $i = 1\,(2)$.

Volterra himself applied simple calculus to analyse Eq. 2.1 and with that determine the model behaviour. It can easily be seen that the solutions must be periodic functions. Their phase space solution is (Murray, 2002, p. 80),

$$\frac{\gamma_2 N}{\epsilon_1} + \frac{\gamma_1 P}{\epsilon_2} - \ln\left[\left(\frac{\gamma_2 N}{\epsilon_2}\right)^{\epsilon_2/\epsilon_1} \frac{\gamma_1 P}{\epsilon_1}\right] = H, \tag{2.2}$$

where $H > 1 + \epsilon_2/\epsilon_1$ is a constant. These solutions are stable and periodic as shown in Fig. 2.1(a). In particular, the average populations $\bar{N} = \epsilon_2/\gamma_2$ and $\bar{P} = \epsilon_1/\gamma_1$ are conserved and only depend on the interaction and reproduction constants. Volterra now went on to consider the impact of heavy fishing represented by a perturbation of the constants ϵ_i, namely in this case ϵ_2 increases and ϵ_1 decreases. We can easily see that this results in an increase of the prey population and a decrease in the predator population: the cycles in the phase plane are shifted as shown in Fig. 2.1(b). From a comparison of the empirical data to his mathematical predictions he concluded that he had answered the initial question about the connection of fishing and population sizes as he found "the results of the statistics [data] (...) to be in accord with the mathematical predictions" (Volterra, 1931, p. 21).

This is not the end of the story: having answered his initial question Volterra himself, but also many more scientists up to today went on to study his model. Here I want to give a brief overview over how the model was studied and what it was used for thereafter. In his article of 1926 Volterra proceeded to study variants of the initial model, all based on the initial assumptions. Specifically, he analysed different types of *internal* interactions systematically by allowing the signs of ϵ_i and γ_i to change[2], and then proceeded to change the nature of an *external* perturbation. Finally an arbitrary number of species with various types of interactions, and time-dependent external perturbations was considered by adapting the equations adequately. Volterra concluded his analysis by deducing generalisations about the dependency of the behaviour of the populations on the model parameters, such as the number of species or the type of their interaction. However, in his investigation he never compared his results to any empirical data, but only analysed them with respect to their (intuitive) plausibility.

Interestingly, after Lotka had found that the model did not represent its designated target (chemical reaction kinetics) adequately, he still found its dynamics intriguing in themselves. Already in his 1910 paper he notes

> No reaction is known which follows the above law (...). It seems interesting, however, (...) to note that in a system in which consecutive reactions take place in the presence of an autocatalytic decomposition-product, we have the requisite conditions for the occurrence of a periodic process. (Lotka, 1910, p. 274)

Accordingly he generalised the model and *searched* for applications, both of which he presented in his more detailed account in 1925 (Lotka, 1925). In cases such as population growth in the United States, or an individual organism in general (for a one-species model), the spreading of immunising diseases in a population (for two-species) he found the model to be adequate in that the model behaviour qualitatively agreed with the relevant data. He concluded his study of the model by systematising the different ways in which species can be related to one another according to the cases that occur in the model.

In the same vein, scientists subsequent to Lotka and Volterra searched for further real-world phenomena, in particular concrete sets of data, to which the model might be applicable. One important example is the application of the model to the data on lynx-hare interaction in a national park in Canada. It was found that according to the model hare were eating lynx! To render the model more realistic modifications were introduced, that is, Volterra's strong assumptions were individually relaxed. Specifically, assumption (i) was changed to allow only integer populations, and discrete timesteps. Assumption (ii) was modified in that growth rates depending on both predator and prey densities were introduced (rather than allowing unbounded growth). Akçakaya et al. (1995) summarise the studies of (two-species) predation models in terms of a functional response (of prey to predators) encoded in the func-

[2] These combinations result in three effective types of interaction, namely (i) competition, (ii) symbiosis and (iii) predator-prey interaction.

tion $g(N, P)$ (called the trophic function),

$$\frac{dN}{dt} = f(N)N - g(N, P)P, \qquad \frac{dP}{dt} = eg(N, P)P - \mu P, \tag{2.3}$$

where $f(N)$ is the prey growth rate, μ is the predator death rate, and e the predator trophic efficiency[3]. Numerous forms of $g(N, P)$ have been suggested, but Akçakaya et al. argue that due to its empirical adequacy a relationship $g(N, P) = g(N/P)$ is *in general* the most appropriate and simple one. Hence, they generalised all kinds of possible assumptions in the Lotka-Volterra model and its descendants in a unifying formalism.

Which were the lessons of the Lotka-Volterra model? Given its lack of quantitative empirical adequacy the main impact of this highly idealised model was pointing to the delicate and intricate relationship between interacting species. In particular, showing the significant impact of external perturbations on the ecological equilibrium is a main result of the model. It triggered a large group of scientists to become involved in the subject and provided the basis of subsequent more detailed studies of population dynamics.

2.2 Schelling's chequerboard model

As an example from economics I present Schelling's famous chequerboard model of residential segregation (Schelling, 1971, 1978). Schelling's motivation to construct his model was that in all areas of societal life segregation phenomena occurred. For example, he was interested in the mechanisms behind segregation between blacks and whites in U.S. cities. His model is targeted at elucidating the relation between the macroscopic phenomenon of segregation and the microscopic level of individuals, i.e., "understand what kinds of segregation (...) may result from individual choice." (Schelling, 1978, p. 142). Furthermore, he aimed to assess the extent to which inferences from the actual macro-level phenomenon (segregation) to the micro-level details (individual preferences) can be made. With several examples of segregation by race, sex and religion Schelling motivates his hypothesis: pronounced collective behaviour can come about as the un*intended* result of un*coordinated*, *individual* choices and/or mild preferences. In his view, in order to understand the phenomenon or residential segregation, one has to study the behaviour of the relevant process, and how it changes under different initial conditions.

He constructs the simplest conceivable model of segregation that can be interpreted by any "twofold, exhaustive and recognisable" (Schelling, 1978, p. 138) distinction between autonomous individuals, but is not applicable to organised action or economic processes. The model consists of a square grid (that has at least the size of a chequerboard) on which two types of coins (pennies and dimes) are distributed randomly such that the grid is not entirely filled. The coins can then move about according to previously specified rules. The two types of coins correspond to two

[3] In terms of Volterra's formulation: $f(N) = \epsilon_1$, $\mu = \epsilon_2$, $g(N, P) = \gamma_1 N_1$ and $e = \gamma_2/\gamma_1$.

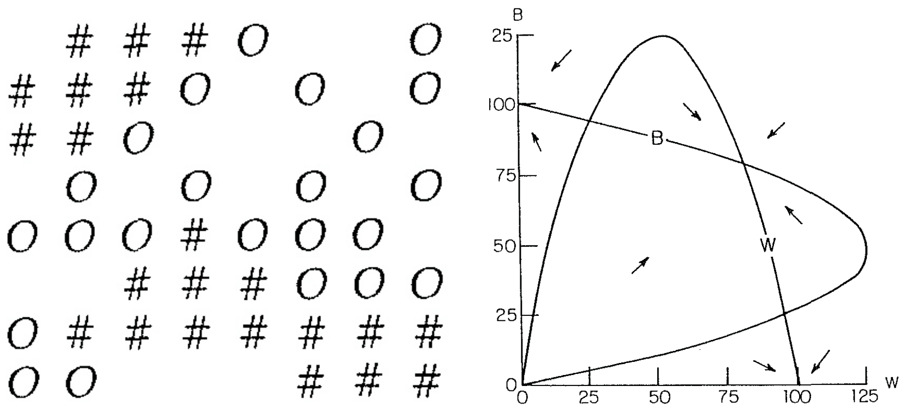
groups $\mathcal{A}$ and $\mathcal{B}$. The rules that determine the dynamics of their movement depend on the individuals' preferences $P_{\mathcal{A}}, P_{\mathcal{B}} \in [0, 1]$ on what the ratio between the number of individuals $N_{\mathcal{A}}$ of $\mathcal{A}$ and $N_{\mathcal{B}}$ of $\mathcal{B}$ in a certain area of the chequerboard (the neighbourhood) must be for every individual to remain where they are. Thus $P_{\mathcal{A}}$ and $P_{\mathcal{B}}$ are lower or upper bounds to the ratios $N_{\mathcal{A}}/N_{\mathcal{B}}$ and $N_{\mathcal{B}}/N_{\mathcal{A}}$, respectively. The total numbers of $\mathcal{A}$- and $\mathcal{B}$-coins can be changed, as can be the preferences of all individuals.

In the simplest case, the "self-forming neighbourhood model", the neighbourhood is relative to each coin and consists of the eight squares around it. The modeller will then "sweep" across the grid and move the coins depending on what the actual ratio in relation to the preferences is. If a coin's preference is met, it will stay, if not, it will move to the closest grid point at which its preference is met. In this situation the movements are readily interpreted by the individuals decisions, whether they should stay in, or leave their neighbourhood.

In the "bounded neighbourhood model" *the* neighbourhood is defined absolutely as a certain area on the grid. Every coin's preferences now regard the same area, where before they only regarded their own surroundings. However, now every coin is allowed to have a different preference, which is modelled by a distribution of preferences. Schelling adds the assumptions that (i) all preferences are upper bounds on the respective ratios (tolerances), (ii) there is flux of people both into and out of the neighbourhood, and (iii) there is perfect information about the numbers $N_{\mathcal{A}}$ and $N_{\mathcal{B}}$. Schelling suggests to interpret this situation as "membership or participation in a job, office, university, church, voting bloc, [sic!] restaurant, or hospital." (Schelling, 1978, p. 155).

Schelling himself analyses the self-forming neighbourhood model physically, by moving around the coins, and observing their behaviour, in particular, which equilibrium state the coins on the grid reach and when it obtains. Likewise he advises the reader to do so, too: "I cannot too strongly urge you to get the dimes and pennies and do it yourself. (...) there is nothing like tracing it through for yourself and seeing the thing work itself out." (Schelling, 1978, p. 150). It is found (and confirmed in computer simulations) that already moderate preferences (such as lower bounds $P_{\mathcal{A}} = 1/3$, $P_{\mathcal{B}} = 1/2$) can lead to pronounced segregation irrespective of the initial distribution of the two groups on the grid. A particularly interesting phenomenon is the effect of differences between group sizes or preferences on the *population densities*: if there is an asymmetry between either of these factors, i.e., if the individuals of $\mathcal{A}$ are less demanding than those of $\mathcal{B}$, or if their total number is larger than that of group $\mathcal{B}$, they will be less densely distributed over the grid, while group $\mathcal{B}$ will be close together as shown in Fig. 2.2(b).

Schelling analyses the second model graphically, drawing the tolerance distributions of each group (in terms of absolute numbers) in a single coordinate system as shown in Fig. 2.2(a). The dynamics of movement are then shown in terms of the direction of change on each point of the coordinate system. The system is in equilibrium at an intersection point of the two curves or at a root of one of the two curves, but its stability must be analysed separately. By changing group sizes



(a) (b)

Fig. 2.2: **(a)** A possible outcome distribution of the self-forming neighbourhood model, if one group (hashs) is less discriminatory than the other (circles). **(b)** In the bounded-neighbourhood model there is a distribution of preferences (plotted here in terms of total numbers). Here the total numbers of the two groups are equal. The arrows show where the population in the neighbourhood will move to. (Both figures reproduced from Schelling (1978).)

and distribution curves liberally different situations can be modelled. In that way instruments or policies such as restrictions of entry into the neighbourhood can be modelled.

Schelling's analysis of segregation was the first example of a whole branch of economics, namely the study of agent-based models (Epstein and Axtell, 1996, Winsberg, 2013). In these models a large number of agents interact according to *local* rules according to which the system is then evolved in simulations.

On the one hand, a range of scenarios in response questions of the type "What would happen if . . . ?" have been studied[4]. It has been found that Schelling-type segregation is strongly robust under a range of agent rules and neighbourhood setups. These include perfect integrationist preference (Epstein and Axtell, 1996, Pancs and Vriend, 2007, Zhang, 2004a), finite agent lifetimes (Epstein and Axtell, 1996), policy interventions (Muldoon et al., 2012, Pancs and Vriend, 2007), and weighted preferences (social distance), constraints on agent movement, forced agent movement (Fossett, 2006, Macy and van de Rijt, 2006). Lately, even as general properties as the effect of different levels of information about the neighbourhood have been studied Muldoon et al. (2012)[5]. It was found that partial information about the agent's environment is sufficient for segregation to arise. Furthermore, it has been shown that segregation arises in all kinds of different topologies (e.g. triangular lattices),

[4] Although a complete survey is outside the scope of my investigation, these few examples should stand as representatives. A review is given by Aydinonat (2008).

[5] See also (Laurie and Jaggi, 2003)

and dimensions of the neighbourhood (Epstein and Axtell, 1996, Fagiolo et al., 2007). From these analyses one can conclude that segregation is merely a result of the agent-environment interaction (Epstein and Axtell, 1996), clearly too strong a statement given that there are well integrated cities (Muldoon et al., 2012).

On the other hand, the literature on Schelling's model is also instructive from a methodological point of view. The short-term *dynamics* of the model has been studied using agent-based simulations on computers (Epstein and Axtell, 1996, Fossett, 2006, Macy and van de Rijt, 2006). Recently mathematicians have also moved on to show rigorous results about the *long-term* behaviour of the model, increasingly in the framework of graph theory (Fagiolo et al., 2007, Henry et al., 2011, Pollicott and Weiss, 2001), but also using methods from stochastic evolutionary game theory (Zhang, 2004b).

Schelling himself concludes his study of the self-forming neighbourhood model by asserting that the mechanisms that drive the microscopic dynamics of his model *may* be *compatible* with the observed macroscopic phenomena. In line with this rather weak claim he acknowledges the shortcomings of the model, namely that it is based on speculation, there are time lags in the behaviour, that it neglects organised action, and only contains a single area. However, according to Schelling "it can be built on to accommodate some of those enrichments" (*ibid.*, p. 166). In the same vein, 40 years after their inception, and in spite of their clear shortcomings, Muldoon et al. (2012) conclude that Schelling-like models "remain important tools not only for a fundamental understanding of population dynamics but also for thinking about the effects of potential policy interventions." (p. 60).

2.3 Lenz-Ising model

To cover the realm of physics, I consider the "Drosophila of statistical mechanics"[6], the Ising model. This model was proposed by Lenz (1920) and solved in one dimension by his student Ising (1925). It was constructed as a model of ferromagnetism and studied with the hope that it might exhibit the (hitherto) not theoretically understood phase transition from para- to ferromagnetism at a critical temperature T_c.

The assumptions of the model can be understood as a direct response to Weiss's theory of ferromagnetism in which Weiss assumed (long-range) dipole interactions between elementary magnets in a solid. On the contrary Lenz and Ising assumed that the elementary magnets (i) interact only via a short-range interaction, and (ii) that the elementary magnets that can only be oriented in discrete directions according to the solid's crystal structure.

Specifically, the magnetic medium is modelled as an array Λ of N lattice sites, where on each site $i \in \Lambda$ there is a spin σ_i, $i \in \{1, \ldots, N\}$ each of which can point either down or up $\sigma_i \in \{-1, 1\}$ (cf. Fig. 2.3. These spins interact only via a next-neighbour interaction strength J that is negative for repulsive, and positive

[6] as my professor for statistical mechanics, Ulrich Schollwöck, calls it.

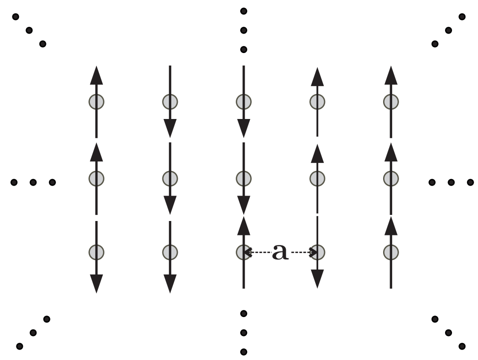


Fig. 2.3: Setup of the two-dimensional Ising model on a square lattice with spacing **a**.

for attractive interactions. The (classical or quantum-mechanical) Hamiltonian that describes this system is given by,

$$\mathcal{H} = -\sum_{\langle i,j \rangle} J\sigma_i\sigma_j - H\sum_i \sigma_i,$$

where $\langle i, j\rangle$ denotes all next-neighbour combinations $i, j \in \Lambda$ and H is an external magnetic field.

The model is then studied in the formalism of statistical physics. From the partition function $Z = \sum_\sigma \exp[-\mathcal{H}(\sigma)/kT]$ all macroscopic quantities can be extracted. Here $\sigma = (\sigma_i)_{i\in\Lambda}$ is a possible spin configuration, T is the absolute temperature and k the Boltzmann constant. For example, the magnetisation M is given by $M(H,T) = kT/N \cdot \partial(\ln Z)/\partial H$. The Ising model is considered ferromagnetic, if the average magnetisation $M_0(T) = \lim_{H\to 0} M$ is nonzero. Ising (1925) solved the model exactly in one dimension by explicitly counting all spin configurations and showed that it exhibits no ferromagnetism.

Only 20 years later, in 1944 Lars Onsager solved the two-dimensional Ising model on an infinite square lattice without external magnetic field. He used the method of transfer matrices to calculate the partition function, which had been introduced by Kramers and Wannier (1941). As predicted by Kramers and Wannier from an approximate calculation, the two-dimensional model exhibits a phase transition that is reflected in a singularity of the specific heat, and the emergence of a nonzero order parameter (Fig. 2.4(a)) [7]. Likewise, the three and higher dimensional Ising model undergoes a phase transition at a critical temperature T_c.

The Ising model is a particularly interesting case as its role has changed dramatically over the century that has passed since its inception (Niss, 2005, 2009, 2011). While initially, the Ising model was not very popular at all, interest in the model

[7] It is instructive to consider the way Onsager presented his solution: he began by applying the transfer-matrix method to the linear chain, and continued with a torus (2d with periodic boundary conditions), to finally arrive at a rectangular lattice (2d with fixed boundary conditions). Thus, starting from the familiar he moved on to his novel solution.

was stimulated by Onsager's solution in 1944. In the early days of the Ising model and still a few years after the 2D solution the model was predominantly studied out of mathematical curiosity, not because it was thought to yield much physical insight. This was due to the fact that it described none of the above phenomena quantitatively; only qualitative agreement between model and data was observed (Niss, 2009, p. 256). For example, motivated by the Ising model, Cyril Domb investigated the transfer-matrix method in depth, but had to *search* for applications for this method (Niss, 2009).

On the other hand, Domb also initiated the study of the Ising model as a unifying model to study cooperative phenomena *in general* and thus initiated an increasing amount of research on the critical point behaviour of systems. A critical point is present at a singularity in the specific heat of the system. The behaviour of a system at the critical point is characterised by so-called critical exponents that characterise the power-law scaling of certain parameters close to the critical point. For example, close to the critical point the magnetisation behaves according to $M(T) \propto |T - T_c|^\beta$, $T \to T_c$, where β is the critical exponent. Similar equations hold for the specific heat C, the magnetic susceptibility χ, the equation of state, and the correlation length ξ. These equations capture the *essential* behaviour at the critical point, although the specific properties of different systems that exhibit a phase transition may be radically different. The Ising model played a major role in the development, justification and acceptance of the "scaling hypothesis", which conjectures a universal relationship between the critical exponents, *regardless* of the system. In a sense the Ising model provided "experimental data" that counted as support for this hypothesis (Niss, 2011). On the other hand, it stimulated questions and suggested experiments, for example regarding the importance of fluctuations of the correlation length in the vicinity of the critical point (Fisher, 1981, Niss, 2011).

In particular, the Ising model was used to make precise statements about the analogies between disparate systems. Over the course of the years it was shown to be equivalent to a range of models from other areas, including the lattice gas, and the binary alloy. In the former interpretation of the Ising model the value of σ_i corresponds to site i being occupied or unoccupied by an atom or molecule (only single occupancy is allowed), in the latter that value corresponds to the two types of atoms in the alloy (Huang, 1987). The analysis of the model (and its comparison with data obtained on all of these systems) suggested that the crucial features for the shape of the magnetisation curves at the critical point are (i) the dimensionality of the lattice (cf. Fig. 2.4(b)), (ii) the symmetry of the order parameter, and (iii) the range of interaction (Griffiths, 1970). All other details of the interaction like the shape of the lattice become irrelevant.

The Ising model's main competitor was the Heisenberg model, which allows for arbitrary orientations of the spins but is otherwise equivalent to the Ising model. As Niss (2009) concludes:: "(...) a *major* factor in the acceptance or rejection of a physical model in the 1950s was its realism - only realistic models were seen as significant." (p. 262). The Ising model was not thought to be a realistic model, while the Heisenberg model was. However, the mathematical intractability of the

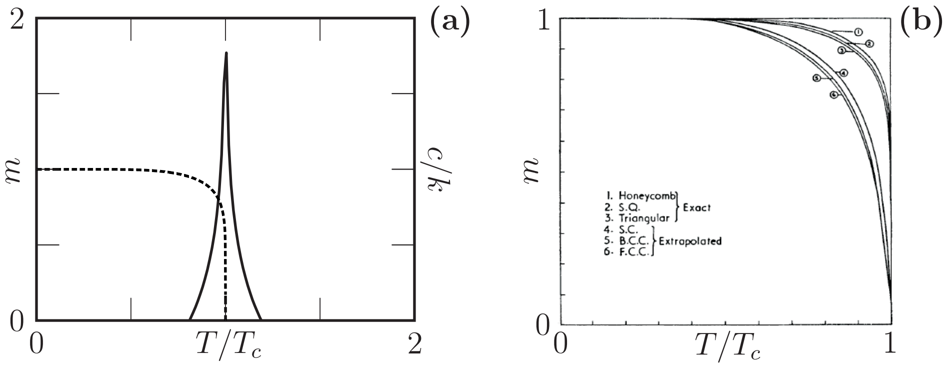


Fig. 2.4: **(a)** Plot of the specific heat c/k (solid line) and the spontaneous magnetisation per spin m (dashed line) in two dimensions on a square lattice as a function of the temperature T/T_c. Here k is the Boltzmann constant, and T_c the critical temperature. **(b)** Spontaneous magnetisation m for different lattices in two (honeycomb, square (S.Q.), triangular) and three dimensions (simple cubic (S.C.), body-centred cubic (B.C.C.), face-centred cubic (F.C.C.)). Figure adapted from (Burley, 1960).

Heisenberg model eventually led physicists to resort to studying the Ising model. With the intensified study new experiments to test the model came up (particularly for the lattice-gas model). It was found that the model reproduced the experimental features of the singularities that occur at a phase transition surprisingly well (Niss, 2009).

With his review article on the Ising model Fisher (1981) spurred yet another outbreak of scientific activity on the Ising model that continues until today. The Ising model has become the "Drosophila of statistical mechanics", *the* standard example in any pedagogical approach to statistical mechanics and is therefore treated in every textbook on the subject[8], such as (Huang, 1987). Even today new applications of the Ising model are found (for example neural networks (Hopfield, 1982, Schneidman et al., 2006)), and further modifications performed and analysed (even the infinite-dimensional Ising model). Thus both structural and methodological connections all across the sciences are drawn[9]. On the methodological front, the Ising model is even implemented in experimental tabletop setups (Britton et al., 2012a) as a quantum simulator. The hope is that much more efficient simulations can be performed in such a setup than on a computer.

What is the moral to be drawn from the studies of the Ising model? In his textbook on statistical physics, Huang characterises the Ising model as follows:

> The Ising model is a crude attempt to simulate the structure of a phys-

[8] Interestingly, often only the one-dimensional case that is treated, but due to the mathematical difficulties involved, not the two-dimensional case.

[9] For instance, the four-dimensional case is linked to the renormalisation group. Another application is the so-called spin glass, a geometrically "frustrated" system, that is, a system that cannot reach its minimising energy. This in turn is related to neural networks and optimisation theory (Wikipedia, 2014).

> ical ferromagnetic substance. Its main virtue lies in the fact that a two-dimensional Ising model yields to an exact treatment in statistical mechanics. It is the only nontrivial example of a phase transition that can be worked out with mathematical rigour. (Huang, 1987)

Thus, it seems the key to understanding the Ising model as a tool of science is to understand what insight the two virtues of exactly solvability and exhibiting a phase transition provide into critical point phenomena.

These models are toy models All three of the models I propose to study are clearly examples of a toy model according to the criteria above (Section 1.1): all three, Schelling, Volterra and Ising construct models the assumptions of which are based on properties of some real-world phenomena they are taken to represent. Moreover, all three include only a handful of assumptions. We can conclude that they are not only idealisations, but strong idealisations (1). These idealisations result in exact solvability (or in Schelling's case, can be studied in "a half-hour to spare" (Schelling, 1978, p. 147)) (2). In all three cases the rules and assumptions are changed liberally according to the theoretical interest of the modeller (3). Not in a single sentence do the three modellers even mention concrete empirical data, their only aim is to provide a model for a specific *type* of phenomenon, namely spatial segregation in Schelling's case, the relative fluctuations of animal populations for Volterra, and spontaneous symmetry breaking in Ising's case - clearly the modellers only aim to provide *qualitative insight*. This is not to say that the models cannot be (nor have not been) adapted, rendered more precise and realistic thereafter. Finally, all three are studied as models of all kinds of particular phenomena that are of the respective type, and application domains are even *searched for* (4). While the Lotka-Volterra model and Schelling's model are phenomenological models, the Ising model is a model of theory, namely of quantum mechanics or Hamiltonian mechanics.

3. HOW TOY MODELS ARE USED AS SCIENTIFIC TOOLS

How do these toy models, and toy models in general, function to generate understanding? On the route to answering this question I begin by extracting the functions that toy models can perform from the case studies. I do so in Section 3.1. Specifically, toy models can perform representative, pedagogical, exploratory and argumentative functions. They have a conceptual and a pragmatic aspect. Subsequently, in Section 3.2 I analyse in virtue of what toy models are particularly apt to be used as tools in science. Here I will argue that the characterising features of toy models, their high level of idealisation and their mathematical tractability, are the crucial properties in furnishing the representative and the exploratory functions.

3.1 Functions of toy models

In the following I isolate four ways in which toy models function in science from the case studies above. This *illustrative* task takes the level of the respective scientific communities as a basis, i.e., considers what scientists share among one another. The *philosophical* task that arises is an account of how the properties of toy models effectuate the functions. Abstracting away from these particular functions with respect to the individual scientific agent who *uses* these models will then pave the way to understand "understanding through toy models". Specifically, toy models function in the following four ways.

(i) Toy models are constructed to *represent a type of phenomenon* that is defined by a generic feature of phenomena. They are used as substitute systems to study this feature.

(ii) They are an *exploratory tool* used to devise, develop and refine theoretical descriptions and methods, get acquainted with and gain intuition about these.

(iii) They are a *pedagogical tool* used to illustrate theoretical concepts and methods in a well-understood environment and gain intuition about them.

(iv) They are used in *scientific arguments* for certain hypotheses, in particular, the relevance or irrelevance of certain causal factors in bringing about a phenomenon.

I will now use the above case studies to flesh these claims out.

(i) Representation of a type of phenomenon It is evident that all three of the above models represent types of phenomena as their targets, given the characterisations of "type" and "representation" above. Each of the models is an ideal system (in the sense that it is highly idealised) the behaviour of which is similar in the relevant respects to different particular phenomena. The generic feature these phenomena share is precisely that in certain limits they behave like the respective toy model. The model *defines* a type of phenomenon (cf. Hausman, 1992, p. 78). For example, the behaviour of the Lotka-Volterra model resembles the generic feature of cyclic variation of population sizes that obtains in many biological and ecological systems. Schelling's model accounts for the phenomenon of segregation that occurs in a wide variety of social systems that are divisible into two groups and can move spatially. Surfers and swimmers, blacks and whites, men and women all segregate in some environment. Likewise, the Ising model represents the type "phase transition". It resembles the behaviour of its individual target systems that they exhibit close to the critical point.

However, these models do not fully represent each particular target system. For instance, the assumptions (i-iii) of the Lotka-Volterra model are idealisations about real-world fish populations that render the model unrealistic. In particular, we saw that it did not account for the data on lynx-hare interaction adequately, taken in its full simplicity. Likewise, the Ising model was considered inadequate and arbitrary with respect to "the actual magnetic behavior of the material" (Van Vleck, 1945, p. 34)[1]. Schelling's model cannot even be adequately adapted to count as a representation of any particular city; its crude assumptions such as the lattice neighbourhood, or the reduction of individuals to preferences about ratios, render the model inapplicable to any *concrete* real-world situation. Thus there is a trade-off between representation of a type of phenomenon and representation of particular phenomena. While representation of a type requires that the model be highly idealised, representation of particulars requires a close fit of the model to its target.

Notwithstanding, some degree of resemblance *is* necessary for a model to become a useful tool. For example, both for Lotka's and for Ising's model to be viewed as interesting a certain threshold in the degree of empirical adequacy had to be overcome. Two aspects contributed to overcoming this threshold: on the one hand exploration of the models yielded their most significant features, on the other hand scientists shifted their focus in *what* the models were taken to represent. Only through that shift could scientists show in what sense the models are empirically adequate. For example, Onsager's 1944 result that the Ising model exhibits a phase transition in two dimensions made it an interesting system because it allowed scientists to study a phase transition in detail. This also resulted in a shift of focus. The Ising model was now not considered only a model of ferromagnetism anymore, but of cooperative phenomena in general (Niss, 2005). With this theoretically motivated shift in focus came the possibility of experimental validation, as only the resemblance of the behaviour at critical point had to be shown. It was even considered "remarkable

[1] as quoted by (Niss, 2009, p. 258).

and perhaps unexpected" (Fisher, 1964, pp. 947-948)[2] when such evidence appeared. For example, the resemblance of the singular behaviour of the Ising model with the singular behaviour of gas-liquid transitions (as determined by comparing the critical exponents) strengthened the adequacy of the Ising model as a model of critical-point phenomena in general (Niss, 2009, p. 280) (in the eyes of the community working on this topic). Likewise, while Lotka's model failed at representing chemical reaction kinetics, it became popular when applied to (i.e. interpreted and manipulated in terms of) the dynamics of biological populations. Here, the *qualitative* agreement with empirical data was considered enough, as for example the references to qualitative agreement with data by Lotka (1925) and Volterra (1926b) show. However, *quantitative* agreement with the long-term record of the lynx-hare population could not be achieved (Murray, 2002). Notwithstanding, the interest in the model did not suffer: it was still the only grasp scientists had on these population dynamics, and it exhibited interesting features. A further factor that contributed to the positive assessment of these toy models with respect to their adequacy was that the competing models performed much worse, were not tractable, or simply did not exist. All three of the models provided a theoretical perspective, where no other was available. Only in the case of the Ising model the Heisenberg model existed as an alternative. However, because it was mathematically intractable, no inferences could be drawn about it and so it could not be *shown* to be an adequate representation.

We can see that the scientific community negotiates empirical adequacy, and thus the issue of representation. Scientists agree on a designated target and, in the light of what other tools are at hand, decide whether or not a model is an adequate representation of that target. Hence, while it is only empirical data that justifies the use of a model, it is *selected* empirical data, too. Ziman (1965) notes: "The justification of a model is (...) that it provides qualitative or semiquantitative results that are in agreement with observation." (p. 1189). Therefore, the justification of a model depends on the correct alignment of that model with the supportive experimental data, or the experiments that may produce such data. Both models and experiments represent a target phenomenon (counterfactual pattern), and they should also be representatives of one another to be aligned. Thus, since representation is never perfect some slack is introduced to testing a model (Mäki, 2005, p. 310).

Showing theoretically that a toy model captures the effective degrees of freedom of the target phenomenon may also count as justification for its adequacy. This step is only possible if the model is a model of theory and if the descriptions on different length or energy scales fall into *universality classes*, i.e., the variables describing a system on one scale become irrelevant on another scale[3]. In this case, it is possible to show that for a particular phenomenon to arise, the microscopic details of that phenomenon are irrelevant and can be absorbed in a small number of phenomenological parameters (Batterman, 2002). For example, this is done formally in the derivation of thermodynamics from statistical mechanics. In the same vein, the toy

[2] as quoted by (Niss, 2009, p. 278).

[3] Indeed, phase transitions are the paradigmatic example of universality, because the details of the microstructure can be *shown* to become irrelevant on the macroscopic level.

model may be a simplified version of a more complex model that is shown to capture the essential mechanism behind that model. In this case the behaviour of the complex model counts as "empirical data"[4]. In this sense, toy models also represent the underlying theory or related more complex models.

(ii) Theoretical exploration Maybe *the* most important and most characteristic function of toy models is their utility in theoretical explorations. Exploration is especially concerned with gaining *novel* insight into phenomena that were previously not understood. In virtue of their high level of idealisation nature toy models can lead to important *general* insights into theory, methodology, and possibly the targets. Let me separate three ways in which toy models are used as exploratory tools. First, they are used in what has been called *conceptual exploration* (Hartmann, 1995, Hausman, 1992), that is, they are manipulated "without worrying about whether those models depict or apply to any aspect of reality" (Hausman, 1992, p. 79) just for the sake of learning about the theoretical consequences of one's assumptions. Second, the structure of the model may require novel solution techniques. In this case possible techniques will be explored in order to solve the model. Let us call this *methodological exploration*. And finally, a toy model (as such representing a type of phenomenon) may be applied to yet unexplained phenomena to explore its scope of applicability, *scope exploration*. It can serve as a yardstick, a first step, in tackling phenomena that exhibit a feature that is surprising in the light of current theory. In this sense it can be used as a well-understood baseline against which reality is compared.

(1) Conceptual exploration involves the activity of adding and changing the model assumptions, changing the structure of the model's components and adding new ones. Thus a modeller may learn about the behaviour of the model under changes of the assumptions. Specifically, the robustness of a certain type of behaviour is often of interest. For example, Volterra studied the behaviour of his model if species were added, their interactions varied, the initial conditions changed, and dissipation introduced, without thinking about specific target systems in which these changes might obtain. From these studies he could then make statements about *general features* of the model's behaviour that obtain in different circumstances of assumptions. An example is the "law of the conservation of the averages" (Volterra, 1931), stating that independent of the initial conditions, the averages of the species populations remain unchanged. He learnt about the implications of his assumptions and could change them accordingly, thereby refining his model. For example, Volterra also concluded that only for an even number of species can all populations remain finite. This result seemed implausible. He inferred that one might want to study the behaviour of the system if the populations were not allowed to increase indefinitely, but depended on the population size. With this additional assumption (rendering his system "dissipative") he found that the system tends toward a stationary state (Volterra, 1931, p. 41). The model not only provided the structure in

[4] For example, this is the case in the toy model that Mendoza-Arenas et al. (2013) study.

which he explored specific changes in the assumptions, but by repeated "plausibility checks" of the results also guided the direction of the exploration.

In addition to learning about the model itself, by studying a toy model we can learn about its underlying theory (provided it is a model-theoretical toy model)[5]. For example, before the Ising model had been studied, it was not clear that statistical mechanics could accomodate critical phenomena (Niss, 2009, p. 245). Toy models provide a structured and simple set of assumptions so that they are particularly attractive candidates for this use. The effects of every aspect, solution technique, model assumption, or theoretical assumptions may be studied individually. Thus the general mechanisms of a theory can be individuated and exposed in the study of toy models.

(2) Methodological exploration can involve finding a particular way to solve a model that is in some sense an elegant and illuminating rigorous solution, or a convenient way to approximate this solution. Such a method may turn out to be useful in other contexts, too. For example, for the Ising model the transfer matrix was introduced by Kramers and Wannier (1941) as an elegant and rigorous method to solve the one-dimensional model. Only with this method in the bag could Onsager (1944) then solve the two-dimensional model. Furthermore, as already highlighted in Section 2.3 this method was then extended for further applications. Hence, this is an example, where the search for a simple solution of a toy model led to the introduction of a new method.

On the other hand, Schelling's approach was an early example an entire branch of social science that is based on the method of agent-based simulations. His idea that local rules for individual agents can lead to global phenomena at the level of populations is at the core of this method. To make this method feasible and systematically insightful new methods had to be devised to make rigorous statements about this type of setting of which Schelling's model is an example. Thus a toy model poses a specific problem the solution of which may require and lead to novel techniques.

(3) Finally, scope exploration is concerned with the search for applications of a toy model, that is, phenomena that can be adequately represented and learnt about with that model. Thus, in virtue of their mutual representant connections are drawn between different kinds of phenomena that are represented by a toy model. On the one hand, a toy model can be used as a yardstick when tackling an unexplained and so far intractable phenomenon. On the other hand, the comparison of a toy model with a phenomenon may yield a new perspective on that phenomenon. In that they draw novel connections between phenomena, or propose a novel mechanism for them toy models can "transform our vision of the state of things into [their] own likeness" (Ziman, 1965, p. 1192). Scope exploration is guided by the question: "How far can we expand the range of phenomena features of which are *adequately* represented by a particular toy model?" Both the Ising model and Schelling's model are particularly instructive examples. The applications of the Ising model range

[5] Cf. Hartmann (2001).

across all disciplines ranging from biology to the study of financial markets. Likewise, Schelling's model has been applied in diverse fields, for example, lately in social network Schelling-like models (cf. Section 2.2). The frequent talk of "Ising-like" and "Schelling-like" phenomena shows that both have become a standard baseline to which phenomena in any discipline are compared. Both their scope and the relation of the respective phenomena to the behaviour of the Ising (or respectively Schelling) model is explored. In other words, the boundaries of the category[6] defined by "Ising-likeness" are explored.

By these various exploratory activities for which toy models are used, intuition can be gained about the methods involved in the solution of a model, the concepts that are part of a model, and even the underlying systems that the model represents. That is, the various activities and theoretical concepts are associated with one another. For example, by solving the Ising model with the transfer-matrix method I may gain intuition about how the transfer matrix can be applied in general. Likewise by studying the effects of new assumptions on the behaviour of Volterra's model I may learn about the patterns that the model exhibits and gain intuition about which assumption is responsible for which feature of the pattern.

(iii) Pedagogical device Toy models are a pedagogical tool. By studying a toy model an aspiring scientist can learn the application of theories and particular methods. The best example here is the Ising model in statistical physics: it is particularly apt to present students with an ideal system in which they can grasp the role of every component. In virtually every textbook on statistical mechanics entire chapters are dedicated towards the study of the Ising model (e.g. Huang, 1987). In particular, the model is used as an example for the derivation of the macroscopic thermodynamic variables from the possible microscopic configurations of a system. Students will learn how even to an ideal system like Ising's we can ascribe everyday notions like temperature. Thus they will get a grasp of the concept of temperature in statistical physics, and gain intuition about its usage and applicability. Moreover, it is one of the rare examples that allows for an exact solution in statistical mechanics (Huang, 1987, p. 341).

(iv) Argumentative device If science has no grasp on a phenomenon, the method of choice to acquire some comprehension is to devise models that contain the essence of that phenomenon. But how obtain this essence? Here is what makes model making an art. In the process of construction a modeller will take a particular perspective on her target by imagining a world in which few (ideally, the key) factors determine the target's behaviour (Morgan, 2004). She will then construct her model as an image of the target in the model world. And she will justify the model, and with that her perspective by citing empirical data that shows that the model is an adequate representation of the target. In doing so she has used the "internal dynamics" of the model (Hughes, 1997) to show how the model comes to behave as it does.

[6] I refer here to Eleanor Rosch's work on categories as elaborated by Lakoff (1987).

If her addressees, the fellow members of a scientific community comprehend her solutions and agree that these solutions adequately represent the target, she has a case for the relevance of the assumptions to account for the target. This step in the argument is what Magnani (2004) calls *model-based abduction*. It is shown that the model is sufficient (but not necessary) to produce the target phenomenon and then inferred that the model captures the real mechanisms that produce the target. The argument is the better the more susceptible it is to what he calls *manipulative abduction*. Manipulative abduction is possible when an agent can see or experience herself *how* the assumptions bring about the effects and how the patterns of the manipulated object resemble the patterns of the target. It is this step, the fact that, in virtue of its simplicity, every addressee is able to manipulate and explore the proposed model that makes toy models such excellent argumentative devices.

There are different ways in which models can function as arguments: they can be used to substantiate (i) impossibility claims, (ii) possibility claims in that particular features are isolated that are deemed important or relevant by the modeler, (iii) possibility claims in that a particular mechanism is shown to bring about a certain effect[7]. Let's see how our modellers argue in these directions.

Schelling is the most skillful and explicit modeller in using his model as an argument for his case, namely to show that "undirected individual choice can lead to segregation" (Schelling, 1978, p. 137). While the claim that individual choice can lead to segregation would seem intuitively clear, his stronger claim that these choices are *undirected* calls for a more detailed discussion. Schelling goes about very subtly, using many examples to convince the reader of the plausibility of his assumptions. From the fact that football players in a canteen do not plan ahead to circumvent a mechanism that mixes black and white players around a table he hypothesises "that players can ignore, accept, or even prefer mixed tables but become uncomfortable or self-conscious (...) when the mixture is lopsided" (*ibid.*, p. 144). With this disarming example Schelling motivates and makes plausible his central claim that individual choices regarding segregation are merely preferences about the ratio of one group to the other in a designated spatial area. Having presented the dynamics of the model he concludes:

> We can at least persuade ourselves that certain mechanisms *could* work, and that observable aggregate phenomena *could* be compatible with types of "molecular movement" that do not resemble the aggregate outcomes that they determine. (Schelling, 1978, p. 152, my emphasis)

Here we have a paradigmatic case of the use of models to substantiate possibility claims: Schelling presumes that using his model one may argue *for* a (micro) mechanism that brings about (macro) segregation. The until then unrecognised factor

[7] (Grüne-Yanoff, 2013) labels these (and more) aspects "learning" and links them to hypothesis revision. Clearly every good argument results in hypothesis revision, but it seems to me that "learning" involves more than hypothesis revision, namely understanding (see below), which is why I will stick to the label "argument".

that *may* bring about segregation is *undirected* decisions of individual agents. Furthermore, he uses his ever-so-simple model to substantiate impossibility claims, or discredit competing hypotheses. For example, he argues that no conclusions can be drawn about inidividual preferences of the individuals in a population from the aggregate behaviour of that population. This follows precisely from from the fact that segregation can arise *without* segregationist individual decisions. Thus he has shown up limits (impossibilities) to inferences about social systems.

In the same vein, Ising intended his model as an argument for the sufficiency of short-range interactions to bringing about critical behaviour. Although he himself was not successful, his successors were: using the Ising model, they showed that short-range interactions are sufficient for critical behaviour. In addition, they found that the geometry of a system, in particular its dimensionality, is crucial to its critical behaviour. As Niss (2009) reconstructs, the Ising model was also a major tool in discrediting mean-field theory. Subsequently it provided the tools to justify the relevance of fluctuations by pointing to the fact that large scale fluctuations are responsible for critical behaviour Niss (2011, pp. 646 f.). Niss concludes that qualitative agreement was deemed sufficient for models to be used as an argument. Because of its widespread use and acceptance the appeal to this model substantiated an argument. This is because every addressee had some intutition about the behaviour of the model and, if unsure, could work it out for herself. For example, Kadanoff (1966) used the Ising model derive the scaling relations between the critical exponents[8]. Even his assumptions that were *known* to be false were not considered to thwart the argument. Only the status of the Ising model as common knowledge and the individuals' intuition about it can explain the acceptance of this kind of argument.

It seems that the models were used as arguments in virtue of their simplicity. The possibility that every addressee might manipulate them and experience for herself *how* the model assumptions bring about its behaviour. We saw how Schelling set up his argument with appeal to basic intuitions about his examples. He then continued the argument using his model, while emphasising the value of active manipulation to gain intuition (as quoted above). On the other hand the Ising model was used in semi-rigorous arguments that even included incorrect assumptions. Nevertheless, these arguments were considered valid in the scientific community. With Magnani (2004) this practise can be made sense of: to some extent the import of a model to an argument that is gained from the easy manipulability of that model can overrule some incorrect assumptions made.

Relation to Weber's ideal types Max Weber (1991b) has extracted these characteristics for what he called "ideal types" (p. 73). It seems they play a role very similar to that of toy models (cf. Hausman, 1992). According to Weber, an ideal type should capture the peculiarity of a social process that renders it "culturally

[8] Again, I refer to the reconstruction in (Niss, 2011).

significant"[9]. Ideal types are never literally true, but always obtained through a process of idealisation. They provide the means for a comparison and the development of an adequate expression of the relationships we find empirically (Weber, 1991b, p. 73)[10]. Although these characteristics obtain for toy models, too, Weber's ideal types are looser structures than models are. They are not mathematical conjectures, but trace the characteristics of a societal process. Sometimes this may involve the help of empirical generalisations, but never can an ideal type be "model-theoretical". This is because Weber deems social processes too diverse and unique such that no universally valid statements on the level of a general theory can be made. Just like toy models an ideal type is only a *tool*, not suited to fully represent reality adequately, but only conjecturing "a complex of *possible* causal relations" (p. 113, my translation). Also, like toy models an ideal type functions as (i) an illustrative device and (ii) as a guide for the scientist in his research. However, scientists are mainly interested in repeatable patterns in the plurality of particular events rather than specifics of a single particular development. Still, the goal of both is to capture the *essential features* that produce a phenomenon.

Models as experiments In virtue of their representative and their exploratory function models can function as arguments. This usage has many parallels to both the view on models as surrogate systems and models as experiments. Once a supportable relation between model and target is established, that is, they are aligned adequately, it may be used as a "surrogate" for the target (Sugden, 2000). By viewing a model as a surrogate one can infer possible generalisations or hypotheses about the effect of interventions on the target. For example, as Niss (2011) retraces, the data obtained from studying the Ising model served as substitute *experimental results*, where no real experiments could be, or were performed. The studies of Schelling's model, on the other hand, were often performed with the aim to study the impact of possible policy interventions on the housing market (e.g. Fossett, 2006).

This aspect is closely related to the view presented by Mäki (2005), Morgan (2005), namely that modelling activity is very much similar to the activity of experimentation. Both models and experiments are manipulated to achieve isolations and acquire knowledge about the world. This is justified by the resemblance of the experimentation system (real or mathematical) with the target in the relevant

[9] His paradigmatic example is his reconstruction of the development of capitalism under the heading "Die protestantische Ethik und der Geist des Kapitalismus" (1905).

[10] "Für die Forschung will der idealtypische Begriff das Zurechnungsurteil schulen: er ist keine ‚Hypothese', aber er will der Hypothesenbildung die Richtung weisen. Er ist nicht eine Darstellung des Wirklichen, aber er will der Darstellung eindeutige Ausdrucksmittel verliehen. (...) Er [der Idealtypus] wird gewonnen durch einseitige *Steigerung eines* oder *einiger* Gesichtspunkte und durch Zusammenschluß einer Fülle von diffus und diskret, hier mehr, dort weniger, stellenweise gar nicht, vorhandenen *Einzel*erscheingungen, die sich jenen einseitig herausgehobenen Geischtspunkten fügen, zu einem in sich einheitlichen *Gedanken*bilde. In seiner begrifflichen Reinheit ist dieses Gedankenbild nirgends in der Wirklichkeit empirisch vorfindbar, es ist eine *Utopie*, und für die *historische* Arbeit erwächst die Aufgabe, in *jedem einzelnen Falle* festzustellen, wie nahe oder wie fern die Wirklichkeit jenem Idealbilde steht (...)." (Weber, 1991b, p. 73, emph. in the original)

respects. Models are thought experiments [11]. They may be used in arguments as a focusing device, but, being (at least partially) non-propositional, cannot be reduced to arguments. The connection to the world seems to be a similar one for real and thought experiments, too. Both represent their target in some way. The big difference is, though, that real experiments seem to convey understanding about *actual* processes and models only about *possible* ones. However, this distinction is further complicated by the importance of theoretical models for all experimentation activity (Morrison, 2009b). Isolation *always*, both in experiments and theoretically, requires idealisation. Any inferences drawn from these experiments or models must *always* be related smoothly to the individual targets (for example, in a process of *de-idealisation* (McMullin, 1985)). What eventually justifies idealisation in experiments, models and theory is the empirical success and the technological innovation resulting from it.

In conclusion, the toy models presented above are used to perform several different functions. On the one hand, toy models are used as tools of exploration and learning (ii,iii) about phenomena, systems, methods, or concepts. In these senses, toy models have a pragmatic function, namely, gain intuition. On the other hand, toy models function as representations of a type of phenomenon (i). As such toy models are theoretical baselines, against which all sorts of phenomena are compared (cf. the notion of "Ising-likeness"). These two aspects - the representative and the exploratory - are necessary for their usage as an illustration or basis of an argument. In this sense they are (thought) experiments and provide data or individual manipulatory experience that substantiate an argument.

3.2 What makes toy models tools?

In the last section I have identified the functions of toy models in the scientific enterprise. The natural question to ask now is *what* it is in virtue of which toy models are so well suited to be used in these ways. That is, which properties of toy models afford their functions? In the introduction (Section 1.1) I proposed that the characteristic features of toy models are that (1) they are highly idealised, and (2) mathematically tractable models, (3) used to gain qualitative knowledge about a type of phenomenon, which is not fixed in that the toy model's (4) application domain is left unspecified. Furthermore, often (5) in a toy model new connections between objects are conjectured. I now motivate these features (specifically (1), (2) and (4)) by claiming that they are precisely the properties of toy models that afford their functions.

I split the argument into two parts and distinguish the pragmatically and the conceptually relevant properties. The *conceptual* properties are those aspects that relate to the status, position and formulation of toy models as theoretical constructs in an abstract model world. The assumptions, components and structure of a model,

[11] Cf. (Nersessian, 1999).

its link to existing theoretical frameworks, i.e., its position within the "world of theory" belong to this aspect. Then again, the model in the model world stands in relation to a system in the real world. In particular, the question what and whether a model represents is an expression of the conceptual aspect. On the other hand toy models have *pragmatically* relevant properties that afford their exploratory and pedagogical functions. When researchers *actively* manipulate and change a model, *argue* with it, *apply* it to novel phenomena, or *explore* its dynamics they this aspect is crucial[12].

With this distinction I want to point to the importance of research and understanding as an active process, not a static thing that exists independently of us. Yet, only the structure of models affords a rigorous justification of their usage in that they can be related to the world. Only the interplay of the two properties can afford *understanding* the world. I will now flesh out these connections by arguing for the following theses.

1. The high level of idealisation and with that isolation of essential features of an object is the crucial property of toy models in furnishing their representative function. That is, only because in their construction the modeller has idealised the target as far as possible, can toy models represent a general *type* of phenomenon. Moreover, the logical structure of a model affords its usage as a substitute system.

2. Only because of their mathematical tractability afforded by the high level of idealisation can toy models perform their exploratory and pedagogical functions. In other words, if we could not manipulate toy models easily, and if we could not grasp their assumptions, we could not explore theoretical tools, argue with, or learn theories using toy models. Moreover, their open application domain is conducive to the exploratory function.

Conceptual virtues To flesh out the first claim I start by asking: What is the relation between the adequacy of a representation and its level of idealisation? Previous authors have already noticed the strain between idealisation and empirical adequacy (Batterman, 2002, Hartmann, 2001, Morrison, 2009a). The further a modeller idealises a system in a model, the larger the error between the behaviour of the model and the behaviour of the real system will be[13]. Apparently we must acknowledge that the world is complex and not as simple as our mind would like it to be. Therefore, if many systems are to be accommodated within a theoretical construct, this construct must be highly idealised; there needs to be some "slack" between model and target structure. Here is where toy models come in. They represent a *type* of phenomenon.

[12] Mäki (2005, p. 5) makes a slightly different distinction. According to him models have a semantic aspect that relates to the issues of representation and resemblance, and an epistemic aspect, that relates to the aim of indirectly acquiring information using the model.

[13] Of course, unlikely exceptions remain: for example, the Ising model describes very well the behaviour of a phase transition to first order, but worse to second order.

Specifically, we saw how idealisation amounts to the construction of a model that reproduces certain features of a phenomenon and thus isolates these features. But where to stop? Clearly we could carry the idealisation to the extremes. For example, the notion of a spin, i.e., a two-state system, is the highest possible idealisation of a system that exhibits any sort of dynamics. However, such a system does not serve as an adequate representation of *any* system exhibiting dynamics. It is the art of modelling, to isolate precisely the *relevant* features for the *generic* dynamics that we are interested in. Thus they need to be idealised as far as possible such that they can still be said to resemble the type defined by this generic dynamics. All other properties of the concrete phenomena (elements of the type) have been eliminated in the idealisation. Any *further* idealisation will yield the features specific to that type unrecognisable. The resulting model will not be identifiable with the target and its elements anymore; it will not adequately represent it. Hence, toy models are the highest possible idealisation that still retains a link to the target. The behaviour of the model *must* resemble the relevant empirical data well enough such that we can learn about the feature of the phenomenon encoded in that data. Empirical data and only empirical data can justify the use of a model through its predictions as any scientist will acknowledge[14]. On the other hand, for toy models to represent a type of phenomenon a high degree of idealisation is a *necessary* condition. Too many details obscure the essence of the phenomenon in question. If a toy model is both highly idealised and retains an empirical link to its target type it can adequately represent that target.

In the construction of a model imagination is crucial (Morgan, 2004, Weisberg, 2007). In our imagination we may reassemble known-to-be-relevant concepts in a surprising way, render known mechanisms such that we can grasp them, or import methods and perspectives from different areas. For example, Volterra did not add any novel features to the interaction of populations, he simply formulated the problem in a way that made clear, which are the factors relevant for each aspect of the population dynamics. It has always been known that species interact by feeding upon each other. His model was novel because in it he applied this fact in the formalism of ordinary differential equations to the specific question, why a single fish population had decreased during the war. He thus rendered a thitherto intractable problem tractable. The structure he assigned to the problem provided a systematic way to study, adapt and refine the model on the one hand, and communicate these manipulations and their effects on the other hand[15]. With the manipulation of his model in order to increase the fit to the world Volterra exploits precisely this feature of the model. While some aspects may be necessary for a particular model to be

[14] Bailer-Jones (2009) collects quotes from several scientists who endorse this view. For example, the chemist Colin Russell endorses the view that "[m]odels clearly must have some relationship to empirical data or they wouldn't be models." (*ibid*, p. 9)

[15] Even in the scientific literature itself the import of introducing a structured set of concepts to a problem is acknowledged. For example, after their presentation of their toy model for the dynamics of cell injuries DeGracia et al. (2012) summarise a major benefit of the model "This model introduces the language and concepts of nonlinear dynamics to the study of cell injury, the benefit of which have been conveyed by the examples" (p. 1012).

solvable or to yield the wished for results, others may be manipulated and varied arbitrarily, for example, to de-idealise the model (McMullin, 1985, Morrison, 2009a). A toy model thus provides a baseline against which a class of real phenomena can compared. It supplies the language and structure for this comparison. Thus it affords a new categorisation of phenomena and clustering of concepts that are seen to relate to these phenomena (Hausman, 1992, p. 79). A toy model articulates a "space of reasons" (Magnani, 2012, p. 163) spanned by the effective degrees of freedom of a system, and a template that structures this space. For example, with the advent of Ising's model phase transitions were come to be seen as a *geometrical* phenomenon that depends crucially on the dimensionality of a system. Ising himself had not seen the possibility of dimensionality playing such an important role and thus concluded prematurely that no phase transition would obtain in dimensions larger than one, too (Ising, 1925). The novel insight that it provided changed entirely the perspective that was taken on critical phenomena - from a phenomenon that was specific to certain systems to a general geometrical phenomenon.

This aspect is closely related to Weber's ideal types. However, the mathematical structure of (most) toy models allows more than what Weber had in mind for ideal types in the context of social science. Due to the precision of a mathematical formulation, the model can be classified among other models or theories, and assessed rigorously. For example, the Ising model was shown to be structurally (i.e. mathematically) equivalent to the model of a lattice gas, and the model of a binary alloy. The *proof* that this is the case was only possible because of its precise formulation. Likewise, Akçakaya et al. (1995) subsumed all models of population dynamics under a general form and thus showed them to be special cases of a more general structure. Thus, in virtue of its mathematical structure can the generic feature that a toy model behaviour resembles be formalised, maybe even generalised. Simplicity in the sense of mathematical simplicity, or in the sense of a small number of assumptions is the main virtue in facilitating these comparisons and making them feasible. The simple mathematical structure of a toy model warrants the possibility of exposing "general principles that could be valid for an entire class of systems"(Picht, 1969)[16]. These general principles then define a category of phenomena in a rigorous way. Whether or not this category adequately groups phenomena or systems will have to be found empirically. The necessary slack in the adequacy of the model to its individual targets is useful: it affords a common code of communication about a comprehensive type (i.e. contains diverse phenomena). Only highly idealised models can represent a comprehensive type as I argued above. On the other hand, because the application domain of a toy model is open its structure encourages and affords a systematic transfer of concepts between different areas of science.

In summary, I argued that *only because* toy models are highly idealised structures can they represent a type of phenomenon. As a result toy models provide a new perspective on their targets. They may introduce new concepts and terms to a problem and arrange them in a surprising way. Thus they create new clusters of concepts

[16] as quoted by (DeGracia et al., 2012, p. 1001)

that we associate with the generic feature of a type. The mathematical formulation of a toy model gives a precise meaning to the concepts that are associated with its target feature. On the basis of its precise structure it can then be assessed both theoretically and empirically. This structure also substantiates a common language in which a group of scientists can communicate about the targets of a toy model. These two components, the systematic assessment and a common language, make negotiations about the adequacy of a model possible.

Pragmatic virtues What makes toy models so useful and widely used is not only their high level of idealisation *as such*. Most importantly it is the mathematical tractability afforded by it (cf. Batterman, 2002) that affords toy models' fruitful usage in science. Moreover, their open application domain invites exploration and is therefore conducive to exploration with toy models (cf. Ylikoski and Aydinonat, 2013, p. 15).

Scientists can only explore the consequences of a model or use it in arguments if they have a grasp on its components that does not require months of intense study. In order to explore the inner relations of a model, i.e., the purely theoretical relationship between assumptions and behaviour of a model, it must be possible to adapt and change these assumptions or add new ones in a systematic way. Moreover, playing with the model, i.e., varying its assumptions and parameters is only feasible, if we have an effective method of doing so. That means, we must be able to adapt the model easily, and explore the consequences of that adaption reasonably quickly. The structure of a model affords the systematicity of toying. Its exact solvability and a cognitively manageable set of assumptions are necessary for toying to be feasible *and* systematically comprehensible by a scientist. Suppose we had a model that involved N components each of which interacts with every other component differently. To isolate and understand the effect of every component and each type of interaction on the resultant behaviour becomes exponentially difficult with N. Hence it is desirable to reduce the number of components to a manageable set. This is precisely the case for a highly idealised model. One might argue that in the age of fast computers, any such problem can be solved easily and even more exact, too, since the phenomenon can be modelled more "exactly". However, when using computers simplifying or even approximative algorithms are needed to limit computation time. This in turn obstructs the goal, namely to provide an exact account of how the behaviour of the toy model emerges from its assumptions. Batterman (2002) even concedes: "In this case (...) the solutions will be, in some sense, less exact." (p. 22). For toy models to be used as exploratory tools a cognitive *grasp* of the model is necessary: it must be both highly idealised, and systematically (preferably even mathematically) tractable.

Toy models should contain the essentials of a phenomenon or a complex mechanism. If used in an argument these essentials, the idea behind the model should therefore be comprehended readily. If the addressee of an argument does not grasp the assumptions of the argument and the way its purported conclusion came about, she will not comply with the argument and thus it has not achieved the goal of convincing the addressee. Arguments like these can only be made if the basis of the argument is common knowledge, or quickly comprehensible. This points to an important point in the process of exploring toy models: because they are so simple everyone with a reasonable education in the subject will be able to understand, or even participate in the exploration, and contribute ideas. From their study of a research team Nersessian and Chandrasekharan (2009) conclude that this communal process "facilitates consensus on the generation and adoption of new ideas and thus focuses the creative work of a team" (p. 180). Here we can see the importance of the dynamics of the *usage* of a model for the process of developing an adequate representation with the aim to *understand* a phenomenon. The communal process is possible because of the tractability of the model. Evidently, the more heterogeneous and the larger the relevant "team" is for which the model is proposed as a tool, the easier accessible and thus tractable the model must be. Provided this is the case, debate can then be guided by the model's structure. As mentioned above, the model defines a space of reasons, the structure assigned to it, and the language tailored towards it guide arguments and reasoning about the target type, but also its elements. A toy model is the simplest comprehensible system that exhibits the dynamics of interest and thus, like an ideal type, provides a baseline for comparison. Hence, this aspect is both conceptually *and* pragmatically relevant.

Finally, it is beneficial that the model is not tailored towards a particular phenomenon, comprising concepts that *only* represent the particulars of that phenomenon. Rather than being tailored to fit a specific system a toy model provides an ideal system. That ideal system may then be compared to real systems or other models to establish similarities and differences. It provides a perspective on these systems that is furnished by the variety of possible interpretations and thus not constrained to a particular one. In that sense a toy model in itself is an uninterpreted structure. It invites to be explored without worrying about the empirical consequences of every manipulation. That is not to say that in their exploration scientists have no interpretations of the model in mind. A particular interpretation can be a useful guide for which manipulations might be feasible or yield an interesting result. For example, when manipulating the Ising model, one can have a ferromagnetic system, or a lattice gas in mind. The result of such a manipulation, inspired by a concrete interpretation, will not depend on the interpretation nevertheless. It will only make reference to the abstract structure that all particular targets share. In other words, precisely the vagueness and freedom of choice of an interpretation, allows for the free exploration that can result in a new perspective. The multitude of possible interpretations affords links between unconnected concepts in our minds. In turn, our experiences and *intuitions* about a concrete situation furnish the room for exploration of the ideal. This can be intuition gained in another model, too, as Morrison (2009a) illustrates nicely for Maxwell's development of electrodynamics. He was guided by mechanical and fluid-dynamical considerations, both well established theories that he would have studied in depth, but his resulting theory was entirely novel. On the other hand, purely formal manipulations may be interpreted freely. For example, the mere change of parameters, such as the relation between the preferences of each group in Schelling's model can afford different interpretations. Thus the segregation effect arising for swimmers and surfers at the beach, is associated with the segregation of blacks and whites in U.S. cities. Our knowledge or experience of one can furnish an intuition about the other.

In conclusion, it is only the high degree of idealisation and the mathematical (or, in general, theoretical) tractability of a toy model that allows us to explore the relationship between its assumptions and its behaviour so freely. Likewise, arguments with models that lead to the revision of hypotheses are only feasible if the model is comprehensible for the entire community it is addressed at. Thus, if we are to understand the import of a model we can never isolate it from its usage and the community it is used by. Finally, the unspecified domain of application of a toy model allows the free manipulation and interpretation of the model. Thus it furnishes associations of previously unrelated phenomena in the exploratory quest.

What are we to take home from this discussion? In the last paragraph we were led to intuition as an important factor when exploring a model. This referred to experience with, and intuition about the targets of a toy model. Going a step further, it seems that an effect that is vital to the widespread use and purported import of toy models is the intuition that we gain about the ideal model system. Does this intuition lead to understanding? We are moving towards the realm of psychology. Indeed, it does seem that the *active* manipulation of toy models plays a crucial role in the generation of understanding. Schelling himself noted:

> I cannot too strongly urge you to get the dimes and pennies and do it yourself. I can show you an outcome or two. A comptuer can do it for you a hundred times, testing variations in neighborhood demands, overall ratios, sizes of neighborhoods, and so forth. But there is nothing like tracing it through for yourself and seeing the thing work itself out. (Schelling, 1978, p. 150)

In the same vein Michael Fisher notes:

> If one had a large enough computer to solve Schördinger"s equation and the answers came out that way, one would still have *no understanding* of why this was the case. (Fisher, 1983, p. 46)

So why is it that simulations do not yield understanding. This claim is intuitively correct. To make it more precise and understand how toy models *do* furnish understanding in contrast to simulations of the Schrödinger equation, a somewhat clear-cut account of the notions of "understanding", "intuition" and "explanation" is in order. In the following section I will try to provide such accounts and show why toy models do provide understanding, where mere numerical calculations of the Schrödinger equation for a particular system do not.

4. HOW TOY MODELS ARE USED TO OBTAIN UNDERSTANDING

To begin this section let me propose a "toy analogue" to toy modelling. Consider a girl playing with building blocks. She has a box that contains blocks of different colours, shapes and sizes. She takes some blocks and starts putting one upon the other, in all possible configurations, with the aim of rebuilding the neighbour's house. Her building may collapse a few times and be rebuilt in many different shapes. Whenever she feels like it she can go outside and compare the block house to the real house. Thus she can find out if she has done a satisfactory job, or collect ideas on how to continue. She may also use the blocks to build a church or a tower. She will find that the stability of the building will depend on similar factors. The ease of handling the toy blocks allows her to try out a whole lot of different types of buildings, different configurations of blocks within a building, different colours and different sizes. Eventually, it seems, she has understood something about house building. How so? Loosely said, she will have an intuition of what a house should look like if it is to be stable. She may be surprised if, by some trick, a house that would not have been stable in her toy world is actually built in the real world. But what do building blocks and real houses have in common. Would she not have to perfectly model real houses to understand them? Will she be able to satisfactorily explain *why* some houses collapse and others do not? It seems not.

In this section I want to flesh out the questions that arise from this scene. Before I do so, in Section 4.1, I begin by briefly reviewing some opinions on the question how minimal or toy models can generate understanding or aid in learning. My discussion of these opinions underpins what the scene above shows: to understand the epistemic import of toy models we need to take seriously the activity and intuition of the individual modeller. In Section 4.2 I lay the foundations for such an account by fleshing out the notions that lie at its heart: explanation, understanding and intuition. Specifically, I comply with an account of understanding as qualitative knowledge. Finally, in Section 4.3 I wrap things up in arguing for my main thesis: by studying a toy model one can gain understanding about the type of phenomenon it represents. To make sense of what it means to gain intuition about a model or the ability to recognise qualitatively its characteristic patterns I take a cognitive perspective. To this end I draw on a model of the knowledge structure of and reasoning processes in the mind, namely mental models. I argue that the role of toy models can only be understood through the distinction between their sharable external representation and their private internal representation (as a mental model). With this distinction at hand we can also make sense of how toy models guide scientific progress. Finally, I draw the connection to explanation and argue that toy models can contribute to possible explanations.

4.1 Understanding with models

Before I turn to giving an account of how toy models are used to obtain understanding, let me first review some ideas on the topic. In particular, the question has been asked, how we can learn from or understand with minimal models[1]. Batterman (2002) emphasises the tension that exists between a full model that is mathematically intractable and will therefore require approximations to be performed in its solution, and a simple model that captures only very few aspects of its target, but is exactly solvable. According to him, highly idealised models are a means to extract "*stable phenomenologies* from unknown and, perhaps, unknowable theories" (p. 35) that capture the universal features of a system. Their virtue lies precisely in being exactly solvable. Because of this minimal models provide understanding of the universal behaviour they display, which obtains asymptotically (i.e. in certain limits) in concrete phenomena. Once the universal behaviour is understood, we can then move on to understand the distinct features of individual phenomena. For Batterman the role of universality is crucial in justifying the use of minimal models. If universality does not obtain it is not valid to reduce the description of a phenomenon to very few interrelated parameters.

Grüne-Yanoff (2009, 2013) asks how we can *learn* from minimal models, where he understands learning as hypothesis revision. He argues that minimal are used as *epistemic surrogates* to obtain the grounds for revising one's hypotheses about the world. Minimal models can perform this function in virtue of their internal dynamics that affords inferences about the effect of a change of assumptions or parameters. Constructing models amounts to constructing credible worlds, where the "credibility" of that world is its plausibility given one's experiences. A model can be judged credible just like novels that tell a plausible story, which might in fact not be true. This means that *no* resemblance with the world is necessary for credibility to be asserted. Moreover, intuitions and experience are important when determining credibility. If the world constructed with a model is judged credible, it is deemed possible. In this sense minimal models as surrogate systems can function as arguments. They can serve as justification for the revision of, for example, impossibility hypotheses about an object. Their utility is a heuristic one, directed at developing claims about the world, which may be tested.

Finally, Ylikoski and Aydinonat (2013) argue that we *understand* a phenomenon if we are able to make correct "what-if" inferences about it[2]. By manipulating minimal models a scientist may perform a robustness analysis of a (model) phenomenon under changing assumptions and parameters. They are used to obtain modal un-

[1] These models are minimal in the sense that they are the highest possible idealisation that still contains the essential features of a phenomenon (as discussed in Section 3.2).

[2] See below for an explication of "what-if" inferences in the sense of Woodward (2003).

derstanding, i.e., understanding about possible worlds. According to Ylikoski and Aydinonat clusters of minimal models (i.e. families of models that describe different aspects of a phenomenon or are obtained from one another) modify the "menu of possible causes" (*ibid*, p. 11) for their target. They provide a "causal mechanism scheme" for a general phenomenon, a mechanism that is deemed sufficient to generate that phenomenon. This scheme can be part of a causal scenario constructed to account for a specific phenomenon. Thus minimal models furnish how-possibly explanations and give us an idea of which causal scenarios may account for a phenomenon. The import of a particular model can therefore only be made sense of in the light of competing accounts. They claim that minimal models also expand the scope of an individual's scientific understanding that does not depend on knowledge of the *actual* causes of specific phenomena. This is because knowledge of one's epistemic tools (models) is crucial to "developing preconditions for the proper understanding of real-world phenomena" (Ylikoski and Aydinonat, 2013, p. 26). Hence, again the utility of models presents itself in the *search* for an account of a phenomenon, i.e., as a heuristic device to function in the selection of a causal scenario as an adequate description.

All three of these accounts catch important aspects of the role of models in science. Namely,

(a) Models capture the universal behaviour of a type of phenomenon against which particular phenomena can be compared.

(b) The pragmatic aspect: models are valuable surrogate systems in which possible mechanisms and theoretical tools can be explored.

(c) Minimal models and the mechanisms they provide can be used as arguments for the revision of (im)possibility hypotheses.

These functions are closely related to the ones I have extracted in the last section. Grüne-Yanoff's conclusion that minimal models serve as arguments that aim at hypothesis revision seems sound, also in the light of my discussion above. As I have shown above, for us to learn something about *the world*, some link between the model and its target must exist. This link may be via a more complex model that is explored using the toy, or directly in that its behaviour qualitatively resembles the relevant dynamics. Only then can it be used qualifiedly to argue about the world. Grüne-Yanoff (2009) argues that modellers such as Schelling do not provide such an account. As a first step, credibility based on general knowledge of patterns and an identification of the model components with some real-world objects seems to be sufficient. However, it seems that to assert credibility is to assert that at least *in principle* it should be possible to show qualitative resemblance with the relevant data.

Ylikoski and Aydinonat make a stronger claim than Grüne-Yanoff, namely that minimal models provide understanding and contribute to our ability to construct a (possible) explanation. In the same vein, the scenario of the girl playing with blocks above shows that models may not always support concrete explanations, but sometimes only the ability to construct a possible explanation. While the girl has an intuition about the stability of houses and which factors may contribute, she cannot provide a full explanation for it. This is because her (phenomenological) model does not relate to an underlying theory (e.g. gravitation). However, the toy model does serve as a tool to gain understanding in the sense of being able to make correct "what-if" inferences about houses. Moreover, phenomena that may have been unrelated previously are related in that behave similarly as the model that jointly represents them. On the one hand the mechanism of a model may be used as a part of an explanation for a concrete phenomenon. On the other hand the links a model provides to other features of a phenomenon, models, or theories may be useful in the construction of a possible explanation. Finally, impossible explanations may be disqualified on the basis of a model.

Two aspects still require to be made sense of: *how* and *why* is the exploration of *toy* models (or the possible worlds they span) particularly apt in the context of *individual* understanding? How does individual understanding relate to the communal activity of science? It is commonly acknowledged that toy models convey intuition. But what does it mean to have an intuition about a phenomenon? What does it mean to understand it? When have we explained it? It seems that the *activity* of the individual scientist who models is at the core to understanding the epistemic import of toy models. On the other hand, the use of models as arguments relates to the aspect of communication within a group of scientists. In the following I want to first clarify the notions of explanation, understanding and intuition in Section 4.2, and then present an account of toy modelling from a cognitive perspective in Section 4.3.

4.2 Explanation, understanding and intuition

A central goal in science, besides description and prediction, is arguably to generate an understanding of a phenomenon[3]. This is particularly true for modelling activity as, for example, Hartmann (1998) asserts: "*gaining understanding of the processes involved by means of exploring the consequences of single features of a theory* is certainly very central to the very idea of modelling" (p. 18, emph. in the original). Traditionally, understanding is taken to come along with explanation. In that view, if I have explained something, I can understand it, and if I understand it I can explain it. Nevertheless, the two notions are clearly distinct in that understanding is *private*. It is something that an individual agent does. Explanation is *public* in that it is an act of communication. So how are the public notion of explanation and the private notion of understanding related?

The distinction between explanation and understanding has a long history in

[3] In the light of the above discussion of how modelling as one of the central activities of science proceeds, this claim is self-evident. Specifically, de Regt and Dieks (2005) argue for the claim that "a scientific theory should be intelligible" (p. 143), in the same vein as, for instance, Fisher and Ziman do.

philosophy. Wilhelm Dilthey (1894)[4] construed the difference between the two concepts on the basis of a distinction between the method of social science and that of natural science. According to Dilthey natural sciences proceed by "hypothetical explanation", while social sciences are concerned with understanding a phenomenon (Lenk, 1972). However, this attribution of explanation to the natural, and understanding to the social sciences is blurred in the modern view on scientific theorising as, for example, Michael Fisher's comments show[5]. In this section I will try and bring some light into the distinction by first reviewing the most important accounts of explanation, and then proceeding to clarify the notion of understanding. Specifically, I will provide a philosophically comprehensible characterisation of what it is to understand something that transcends the commonly used appeal to the intuition of the reader[6]. In a third step, I will try and capture the meaning of "intuition" since it seems to be central to understanding "understanding". Finally, I relate the three concepts.

Explanation There is no unique and definite notion of explanation. Roughly speaking, in science explanation seems to be concerned with (Lenk, 1972) (1) "the explication, (...) or definition of concepts or scientific terms" (p. 693), (2) reducing sentences to instances of a general principles or generalisations, and (3) explaining sentences by referring to events or facts. Therefore explanation is about making *statements* about a phenomenon that clarify and explicate it in terms of related phenomena, generalisations, or concepts. It is an act of communication, publicly accessible in virtue of its linguistic formulation. The goal of any explanation of a phenomenon is to make the addressee understand that phenomenon (Woodward, 2011). Whether an explanation is satisfactory depends on how it relates to the question asked, as well as the tacit intentions and background of the involved actors. It is context dependent with respect to the communicating agents.

What exactly characterises a good explanation is one of the most widely disputed topics in philosophy. Uncontroversially, the most influential account of explanation is Carl Hempel's deductive-nomological model (Hempel, 1966). All currently held views in one way or the other respond to his view on what an explanation is. Two major ideas prevail in the discussion of what makes an explanation convey understanding.

(1) In the unificationist account of explanation that is mainly due to Philip Kitcher (1989) to explain a phenomenon is to show how it can be derived in an "argument pattern" that fits as many phenomena as possible. That is, the explanation should show how the phenomenon fits into a broader theoretical context. In particular, in this view deductive-nomological explanations are good explanations. It

[4] as cited by (Lenk, 1972).

[5] Cf. Section 1

[6] Appeal to intuition is usual strategy by which accounts of explanation are justified or rejected. For example, the flagpole objection to Hempel's deductive-nomological account of explanation appeals to our intuition that the length of the shadow does not explain the height of the flagpole casting that shadow (cf. de Regt and Dieks, 2005).

seems that in virtue of being a representation of a type of phenomenon, toy models perform this explanatory function for the generic feature that defines the type.

(2) According to the causal mechanistic account of explanation, mainly due to Wesley Salmon (1984), to explain a phenomenon is to show how it came about in a "causal process" (Woodward, 2011, p. 34), that is, provide a mechanism. In particular, a good explanation should provide counterfactual information about what would have happened had the explanans not occurred or occurred differently. For example, the Ising model provides a mechanism for the (ideal) phenomenon of a phase transition. It shows how a phase transition comes about by specifying the components and their interactions. In this sense it explains phase transitions. Woodward (2003) provides a graph-theoretical formalisation of a counterfactual account. In particular, according to Woodward, an explanation should give answers to "what-if" questions relating to the phenomenon in question.

Alisa Bokulich (2011) provides an account of explanations based on scientific models. In her view, in a "model explanation" for a phenomenon or a feature thereof, (i) the explanans must make reference to a model, (ii) show how "the elements of the model correctly capture the pattern of counterfactual dependence of the target system" (p. 39), and finally (iii) assign a domain of application to the explanation. For example, in this view, though it is known to be literally false Bohr's model of the atom explains the spectrum of hydrogen. This is in opposition to, for example, the Ptolemaic model of the atom because what it means to "correctly capture the pattern of counterfactual dependence" is negotiated by the scientific community and thus depends on the current state of knowledge and the "what-if" questions to be answered (Bokulich, 2012) . According to Bokulich (2012), only if the model *adequately represents* its target, that is, only if it provides "genuine knowledge of the *true* underlying [...] dynamics" (p. 735, my emphasis) by a well-defined translation key to the accepted general theory is it explanatory. Therefore, what makes a representation adequate is highly context dependent. While in physics often a "translation key" (*ibid*, p. 735) can be given[7], in the less exact sciences, where there is no overarching general theory that is taken to be "true" the notion of adequacy is (and must be) much more open and vague.

Bokulich's model of explanation exhibits both unificationist and counterfactual aspects. On the one hand, the model invoked in the explanans should adequately capture the pattern of counterfactual dependence of the target phenomenon or explanandum, on the other hand what "adequate" means depends on the current status of theory and should be coherent with it. While both types of explanation supposedly convey understanding, the way in which they do so differs: a unificationist explanation yields understanding of a phenomenon in its context by integrating it in a general framework, and thus conveys global understanding; a counterfactual explanation gives an account of how a particular phenomenon came about, thus conveying local understanding (Hartmann, 2001). Rather than being mutually ex-

[7] For example, using semiclassical mechanics the components of Bohr's model of the atom can be translated into the language of quantum mechanics. In this sense Bohr's model is an adequate representation of the atom, while for the Ptolemaic model such a translation key does not exist.

clusive these two approaches seem to describe different ways in which explanations yield understanding. Yet, even vaguer than the notion of scientific explanation is the notion of understanding. Let us bring some light to it.

Understanding To clarify the concept of understanding I want to begin with a few historical remarks before moving on to a modern conception of understanding in the context of science.

The philosophical notion of understanding is traditionally closely associated with the humanities and hermeneutics, as opposed to explanation that was associated with the natural sciences by Dilthey: "We explain nature, but understand psychic life" (Dilthey, 1894, GS 5, 144, my translation)[8]. In this view, to understand something is to see it as integrated in a constructed interpretative framework and thus "relive it" (*nacherleben*). In particular, understanding is not simply seeing the structural isomorphy between the reconstructed phenomenon and the real, but the significance and connection of a phenomenon (typically of social nature) in its (historical-societal) context[9]. According to Weber (1991a) in the context of social sciences this amounts to extracting the values (*Werte*) that motivate the actions of individuals, which lead to a phenomenon. Subsequently one interpretatively reconstructs this complex of values and action in a consistent way such that inferences about it can be drawn. In particular, *active reconstruction* is therefore a crucial component of understanding[10].

However, Hempel, and following him, Abel and Popper (Apel, 2001) saw the significance of understanding in the context of the natural sciences. According to them understanding can have an important *heuristic* function in the *context of discovery*. In particular, in Abel's view, to understand a phenomenon is to apply personal experience to a perceived phenomenon. Put differently, to understand a phenomenon is to draw analogies between that phenomenon and already known phenomena. In this sense, the new phenomenon is integrated in an agent's framework of knowledge. This aspect is also called the "familiarity view" stating that explanations provide understanding by relating new phenomena to familiar ones[11]. That is also to say that understanding implies nothing about the *real* connections of facts, but about *possible* connections relative to our current state of knowledge. As such these connections can serve as heuristic guides to further improve understanding. Therefore, understanding a phenomenon seems to include the *integration* of that phenomenon within one's knowledge.

[8] as cited by (Apel, 2001, p. 923).

[9] Cf. (Weber, 1991b)

[10] This is true both from the perspective of understanding through hermeneutics, and the rational concept of understanding that was coined by Kant. Hermeneutic understanding refers to the *active* interpretation or explication of linguistic expressions. On the other hand, Kant construes understanding as that which makes experience from perception according to a rule (Apel, 2001).

[11] Carl Hempel (1966) also suggested that explanation and understanding are linked in this manner: "It is sometimes said that scientific explanations effect a reduction of a puzzling, and often unfamiliar, phenomenon to facts and principles with which we are already familiar." (p. 83).

When talking about understanding in a science context it is important to distinguish understanding of a phenomenon in the world and understanding an abstract theoretical construct. While in the second sense "understanding" is used as in "After long hours of study I have understood Newton's law of gravitation; two masses always exert a $1/r^2$-force on one another.". However, to understand a real-world phenomenon requires the application of the concepts of a mass or a distance to real-world objects. That is, one has to be able to recognise objects as instances of the theoretical model in the sense of "Finally, I understand why apples fall to the ground; they have a mass and, by Newton's law, are therefore attracted by the earth.". Scientific understanding seems to require the intellectual comprehension of a theoretical construct and the ability to recognise real-world phenomena as instances of that construct.

de Regt and Dieks (2005) realise the importance of the context to whether or not we say that we understand a phenomenon. Whether or not an individual scientific agent[12] S can comprehend a model depends on the capacities, skills and knowledge of S. In particular, comprehension of a theory or a model is not the ability to perform exact calculations - this can be done on a computer without any import to understanding - but the ability to draw *qualitatively correct* inferences about the model. This is local understanding about particular counterfactual dependencies. On the other hand, the consequences of the model should be coherent with and fit into our established theoretical knowledge. This is global understanding about how the model relates to other theoretical constructs. These two components constitute intellectual comprehension of the model. Together with experience of how and when the model applies they imply what is often called "a feeling for" the model (de Regt and Dieks, 2005, p. 156), that is, *intuition* about how the model behaves and when it is an adequate description. The ability to *apply* a model to real-world phenomena depends on the ability to compare the representation of a phenomenon to the behaviour of the model.

In accordance with de Regt and Dieks (2005) I propose the following characterisation:

> **(U)** A scientific agent S *understands* a phenomenon P, if S *comprehends* a model M of P relative to a context C_S, and M *adequately represents* P (and meets the methodological requirements) relative to a context C_C.

Whether or not M adequately represents P, and what the methodological requirements for M are is negotiated by the relevant community and depends on the current state of theory as well as pragmatic considerations, all of which are summarised in the community context C_C. These properties of M are its conceptual virtues. From the above considerations regarding comprehension I conclude that

> **(C)** S *comprehends* a model M of P if (i) M is coherent with and related to S's conceptual framework, (ii) S can draw qualitatively correct

[12] By using the term "agent" I want to point to the importance of intentions, and individual background knowledge and capacities in the context of understanding. A scientific agent is an agent that proceeds according to the methodological rules of science.

> inferences about M, (iii) recognise how the consequences of M relate to P as an instance of M.

In particular, whether M of P is comprehensible for S depends on the one hand on the skills and knowledge of S (the individual context C_S) and conditions that M must fulfil in order to be comprehensible by S. These properties of M are its pragmatic virtues such as simplicity or visualisability (de Regt and Dieks, 2005, p. 142). While intellectual comprehension of a model and the skill to apply it are often learned in parallel, they are clearly distinct capacities[13]. The intuitive usage of a model is made possible by familiarity with the concepts it employs of and the ability to draw inferences about a model[14]. In particular, the criteria (i), (ii) and (iii) of **(C)** imply what de Regt and Dieks take as "intelligibility" (parallel to "comprehensibility"), namely that a scientist "can recognise qualitatively characteristic consequences of [a theory] T" (p. 151). However, their account lacks the aspect of coherence and classification of the model with respect to S's beliefs about the world. This aspect had been made explicit in the original account of Heisenberg (1927), which they refer to. Note that the above characterisation **(U)** also makes the importance of personal experience to understanding explicit in that recognition is a skill that is *learnt.*

Intuition To close this section, let me relate the notion of understanding as qualitative knowledge and skill of recognition to what scientists themselves say about understanding. In this context "intuition" is the key word. For example, John Ziman (1965) comments:

> [Being a good theoretical physicist] is the ability to see, or experience in seeing, a given model, or a set of law, as one of a whole class of models, of more general and abstract properties. (...)
>
> The ability, or this training, must be balanced with "physical intuition" [...] is the experience of the properties of all sorts of models, a feeling for the way they will behave, and readiness to construct or adapt a model or a toy to suit the problem on hand. (p. 1192)

Here Ziman stresses the importance of skills, in particular, the ability to classify theoretical models and the ability to relate models to a particular problem (phenomenon) based on an intuition about these models. Thus Ziman closely links "physical intuition" to modelling, specifically, abilities regarding the recognition of

[13] In their characterisation of understanding de Regt and Dieks (2005) do not refer to how the model relates to the world, but absorb this relation in the "logical, methodological and empirical requirements" (p. 150). As I have argued above this is a key aspect linking a model to the world. It might argued that being "a theory T of P" and meeting "empirical requirements" (p. 150) implies what I mean by "adequate representation". However, the model-world relation seems to be crucial for the description of a phenomenon by a model and should be made explicit to reflect this importance.

[14] Cf. (de Regt and Dieks, 2005, p. 159).

characteristic features and the construction of models. For him, too, the other side of the coin is the classification and integration of a model in a broader framework.

According to the *Oxford English Dictionary* intuition is "the ability to understand something instinctively, without the need for conscious reasoning". Intuitive insight as an immediate grasp of the *entire* contextual significance is thus contrasted with discursive insight. We obtain discursive insight through reasoning processes. As opposed to intuitive insight it can only cover partial aspects of a phenomenon (Kobusch, 1976). Like understanding, intuition seem to be psychological in nature. However, the format intuitions take is unclear - they may be beliefs, dispositions or faculties and concern real objects or propositions (Pust, 2012). Let me take intuition to be a kind of *tacit knowledge* about patterns, that is, the ability to recognise such patterns (in particular counterfactual ones) that is based on experience and learning.

Explanation, understanding, and intuition seem to be related in that all three are forms of knowledge. Providing an explanation for a phenomenon should yield understanding, which in turn requires intuition. By getting accustomed with a model we can acquire the ability to recognise instances of that model and draw inferences about it. Here the construction process seems to be crucial: by reconstructing a phenomenon (e.g. by constructing a model for it) we gain the experience of its properties that yields intuition about it. In this sense there is a similarity between mastering a skill and understanding a comprehensive object (Polanyi, 1961). Just as driving a car becomes intuitive in that it does not involve conscious thinking, is a model of a phenomenon comprehended (and thereby the phenomenon itself rendered understandable) in that we can recognise characteristic features of it and intuitively draw qualitatively correct inferences about it.

While the intuitively used ability of driving a car is not sharable, we may be able to give instructions that enable a learner to gain the "correct" experiences that are relevant to mastering that skill. In the same vein, understanding of a phenomenon may be sharable, for example, by providing a scientific model that adequately represents that target phenomenon. While understanding itself may be non-propositional, propositional instructions can be given that convey how understanding can be achieved[15]. With these instructions at hand a scientific agent is then able to gain the relevant experiences in manipulating the model and gain intuition about its behaviour. I suggest that a model explanation in Bokulich's sense is such an instruction manual. It renders a phenomenon or a feature thereof understandable in that it pinpoints the relevant connections in a structured way. However, to obtain understanding in the sense of **(U)** a scientist *additionally* requires intuition about the model, which can be gained from manipulating the model.

[15] Nancy Nersessian (1999) makes a similar point: "Once the initial [thought] experimenter understands the implications of a thought experiment, she can guide others in the community to see them as well by crafting a description of the experiment into a narrative" (p. 20).

4.3 Understanding with toy models

In this last section I will bring together the analysis of the case studies in Section 3 and the explication of understanding in the previous Section 4.2. Specifically, I argue for the following three theses.

1. Through the study of a toy model M_T one can gain understanding of the type of phenomenon P that is represented by M_T.

2. Partial understanding can be gained about a particular phenomenon p that is of the type P.

3. We improve our understanding of the elements of the type of phenomenon P in a process of repeated analysis and adaptation of toy models, or the development of new models.

To show that toy models convey understanding in the sense of qualitative knowledge and skill of pattern recognition **(U)** two aspects must be shown: toy models are comprehensible, and they adequately represent their target, granted that they meet methodological requirements. Therefore, on the one hand, understanding depends on the skills/abilities of the scientist who aims to understand a phenomenon, on the other hand on properties of the model through which she does so. These properties of the model, in turn, fall into virtues that facilitate comprehensibility of the model, and conceptual properties that are relevant to assess the relation of representation. Now the distinction I drew above between the pragmatic and the conceptual virtues of toy models falls into place. In the following I will argue that the pragmatic virtues of toy models facilitate comprehensibility via the activity of exploration. I showed above how the conceptual virtues facilitate representation of a type.

The argument is based on the following premises:

(a) By exploring a toy model a scientist can develop an intuition about that model, as well as theoretical knowledge about its characteristic counterfactual patterns and its relation to already known theories or models.

(b) Intuition (in science) about a model M is the ability or skill to recognise a pattern that is characteristic for M. Therefore it is the ability to recognise instances of M.

(c) A toy model can represent a type of phenomenon.

(d) Whether or not a toy model *adequately* represents a type of phenomenon is negotiated by the scientific community and therefore established in the process of exploring that model.

(e) Toy models are scientific models, i.e., meet the methodological requirements defined by the relevant scientific community.

Exploration of toy models has the three aspects extracted in Section 3: conceptual exploration results in knowledge about the counterfactual patterns of a model, i.e., how assumptions and parameters are related to the model behaviour. Methodological exploration results in an intuition about which tools are suitable for the solution/manipulation of the model. In scope exploration scientists determine for which phenomena the model is an adequate description. What "adequate" means is determined by the theoretical context and empirical requirements[16]. The theoretical context consists in how well competing models (if any exist) describe the target, whether they are consistent with one another, and especially whether or not the model is consistent with the underlying theory (if it exists). Therefore, if a scientist has the theoretical skills to manipulate a toy model and gained experience about its behaviour and the relation to its target the toy model can be integrated into her conceptual framework. That means it can be associated with other models or theories on the one hand, but also with paradigmatic target phenomena that share a generic feature. This association can happen both on the level of the individual agent and on the level of a scientific community. The ability to recognise this feature, or the characteristic counterfactual pattern of the model results from manipulatory experience with the model. Therefore, a toy model is a comprehensible model with respect to a scientist's capacities, and it can adequately represent a type of phenomenon with respect to the theoretical context and the community.

Given the above premises and the account of understanding **(U)** it follows that through the use of a toy model we can gain understanding about the type of phenomenon it represents (1). A type of phenomenon is the set of phenomena that exhibit the generic feature that is characteristic for the type. Therefore, toy models provide insight into that feature of every individual phenomenon. Thus they yield partial understanding of these phenomena and the mechanisms deemed relevant to their study (2).

Mental models: a cognitive perspective Given that toy models can be related to other theoretical constructs and convey intuition through their exploration, they convey understanding. However, it remains unclear, what exactly it means to "develop an intuition about" a toy model, or "relate" a toy model within a conceptual framework. My characterisation of intuition relates it to understanding and explanation. However, it remains to show how the *recognition* of counterfactual patterns works, how a toy model is *integrated* into a scientist's conceptual framework, and how *toy models* function in this process recognition. While I have identified some properties of toy models that facilitate the development of intuition (ideality, tractability, utility for exploration), it still remains unclear *why* these properties furnish intuition, essentially what it means for toy models to "convey understanding". These questions are irreducibly cognitive; they regard the individual's mental processes.

[16] For example, the Ising model was initially not considered an adequate description (even after Onsager's solution); the scientific community preferred the Heisenberg model. However, when the attitude on modelling changed (Niss, 2009), the exact solvability of the Ising model became more important a criterion so that the Ising model became popular as a model of phase transitions.

An account using cognitive science suggests itself.

Two aspects are needed to give such an account: first, an account of intuition that shows how we recognise ideal features in complex phenomena, and second an account of how toy models convey such intuition. To flesh out these questions I draw on (1) the "mental model" hypothesis, and (2) the "common coding" hypothesis (cf. Magnani, 2012).

Mental models are a model for cognitive processes, in particular, reasoning and knowledge representation, but also text interpretation. In essence, the hypothesis states that we reason by simulating the object of reasoning. We simulate objects by constructing a mental models and performing manipulations on these models. What is a mental model? "Broadly construed, (...) a mental model is a structural analog of a real-world or imaginary situation, event, or process that the mind constructs in reasoning" (Nersessian, 1999, p. 11)[17]. It is a structural analog if it adequately maps the pattern of counterfactual dependencies of the target, that is, the object of reasoning and any further information relevant to the reasoning task.

The concept of a mental model is inherently vague. For example, it is unclear "what *format* a mental model takes" (Nersessian, 1999, p. 12), whether or not it is propositional. Moreover it is an open question how these models are generated in the brain and how operations on them are performed. However, in accordance with Nersessian (1999), for the purposes of this thesis, a version of the mental-models thesis that is agnostic to these questions is sufficient. We may associate mental models with representations such as images that are constructed and manipulated for certain tasks. For example, some tasks may require mentally imaging an object and simulating operations on this object[18]. The result can then be compared with a real-world object on which the operations have been performed manually. Such a comparison requires "being able to take two cognitive [mental] models and compare them, noting the ways in which they overlap and the ways in which they differ" (Lakoff, 1987, p. 71). That means that mental models can be compared to direct representations of particular phenomena, which are assumed to be of the same type as mental models. We can now understand how the *recognition* of patterns works. Furthermore, according to Lakoff (1987) our knowledge is organised in terms of interrelated ideal mental models. Given the ability to compare mental models and given that our knowledge is (at least partially) organised in terms of mental models, both the recognition of patterns in empirical phenomena (comparison between an ideal mental model and a direct representation) and the integration of a phenomenon in a context can be made sense of. Whether or not a propositional representation is of a distinct type remains open[19]. Likewise, *how* two internal concepts are compared

[17] Nersessian (1999) refers to the account of mental models presented by Philip Johnson-Laird (1983).

[18] Cf. (Bailer-Jones, 1997).

[19] This is particular interesting in the context of text interpretation. It has been argued that we interpret texts by constructing mental models, simulating a concrete interpretation of the propositional statements in the text. These allow for many different interpretations (Bailer-Jones, 1997, p. 81 ff.).

and by which processes an overlap is asserted or denied in cognitive processing is unclear. These are empirical questions about the nature of representation in the brain that exceed the scope of this thesis.

How are toy models and mental models associated? While an *identification* of mental models with toy models is clearly false, I suggest that they can be associated[20]. How so? According to my characterisation above (Section 1.1) a toy model is a set of assumptions about a concrete system, and therefore propositional in nature. In the language of cognitive science a model in this sense is an *external* one in that it may be *shared* (Bailer-Jones, 1997). It is an external tool that helps to obtain understanding, the utility of which is afforded and enhanced by its commonly accessible mathematical, linguistic, or visual representation. In this sense models are an interface on which we operate to obtain understanding. External models are contrasted with *internal* models, that is, mental models that are inherently *private* (Bailer-Jones, 1997). I suggest that toy models (as external models) provide the "manual" for the construction of a mental model that incorporates the ideal model process and examples.

(Only) in the context of movement representation or dynamics the connection between mental models and external models can be understood with the "common coding" hypothesis. According to this hypothesis there is a shared representation between the execution, perception and imagination of movements in the brain[21]. An exemplar effect cited as evidence for perception-action common coding is the "extension of the mind" to the tools we use (Chandrasekharan, 2009, p. 1070 f.)[22]. Moreover, it has been shown that children learn mathematical concepts such as fractions much better by executing manual movements using blocks. This serves as evidence for imagination-action common-coding (*ibid*, p. 1071). Thus the activity of construction is highly conducive to theoretical learning. In the same vein, the girl from the above example of playing with building blocks will learn about the stability of houses and how it depends on the way blocks are put onto one another. While the research on common coding is still at an early stage, there seems to be "significant evidence in favour of a common code" (*ibid*, p. 1072). Colloquially speaking, this effect its nothing but "learning by doing".

In the light of these hypotheses the significance of the exploratory role of toy models is made sense of. The activities of construction and manipulation afford a coupling of the external toy model to the internal mental models in the sense that parallel to exploring a toy model a mental representation (a mental model) is

[20] Cf. (Bailer-Jones, 1997)

[21] Common coding is closely linked to the hypothesis of embodied cognition (Chandrasekharan, 2009). "Cognition is embodied (...) when aspects of an agent's body beyond the brain play a significant causal or physically constitutive role in cognitive processing." (Wilson and Foglia, 2011). While common coding is a statement about the representations of processes at the neural level, embodied cognition takes a high-level view on the same effect, namely that certain cognitive processes and actions of the body are not separable.

[22] This hypothesis has been defended in different fields of philosophy (e.g. Heidegger, 1927, Polanyi, 1961).

constructed or changed that represents the toy model[23]. Chandrasekharan (2009) emphasizes precisely the importance of construction for the usage of models to obtain understanding. Specifically, constructing an external model involves mentally componentising, and imagining the interaction of the components. Because a mentally constructed toy model can be manipulated easily we can imagine many different ways in which the components interact. This is the significance of a toy model's high level of idealisation and tractability. Because it is mathematically/logically structured and limited it acts as a source of constraints to focus imagination. Only with this sort of focus can the activity become productive (Nersessian and Chandrasekharan, 2009). Exploiting the internal dynamics of the model, the modeller can then imagine the dynamics of the model. These in turn can be compared to related mental models, or direct representations of phenomena. With the aim to arrive at a coherent framework of mental models both external and internal model will be revised and adapted continuously. Thereby manipulatory experience is gained and with that the ability to identify instances of the model in the future.

We can now see how the *individual* manipulation of the toy model leads to understanding. But how is understanding improved and focused within an entire community of science? Every individual member's exploration of a model contributes to the progress of a community. I suggest that external models are the shared "interface" on which every member of a scientific community operates. Every individual can *in principle* reconstruct a model of interest to them based on these interfaces; in that sense models are distributed (Chandrasekharan, 2009, Magnani, 2012, Nersessian, 1999). On the one hand "everyone (...) has a sense that there is a common understanding of the artifact/system" (Chandrasekharan, 2009, p. 1075), on the other hand the exploratory quest is nourished by the diverse perspectives and the workload coordinated according to expertises. For this reason, it is crucial that everyone in the community can grasp and solve a model: the larger the community, the more impact happens and the more exploration is possible. Thus a model can become a "manifest model" Chandrasekharan (2009), a common interface on which scientists can operate freely and share their results by publishing new building instructions and their effects. In this vein, Chandrasekharan (2009, p. 1079) suggests that for external models to support discovery they should (1) offer a large range of ways to generate dynamics in the mentally constructed model, (2) allow a high level of control over these dynamics, and (3) involve a lot of construction and manipulation by scientists. As I have argued in Section 3.2 these criteria are met by toy models. It is their tractability that affords the high level of control and thus comprehensibility of the model. By their high level of idealisation can toy models represent a type of phenomenon. The combination of these two features leads to the possibility of obtaining individual understanding using the model. Their external representation allows for an extensive and focused study of the model across different scientific com-

[23] There is a nesting of representation: the mental model represents the external model in question, which in turn represents a type of phenomenon. The type stands for the individual phenomena that exhibit the generic feature characterizing that type. Because representation is transitive, the mental model represents a type of phenomenon.

munities. Therefore, toy models are a tool to coordinate a community's common perspective on the world and guide research. A common understanding is achieved in the sense that all individuals of a community share the same perspective.

Now my third hypothesis falls into place. In a process of repeated individual exploration and analysis[24], as well as updating of the shared model different perspectives on a phenomenon are probed. This process is particularly fruitful when toy models are used heuristically, when there is no common conception on how to describe a phenomenon. At this stage toy models are used as argumentative tools. On the other hand once the model is "manifest" it can be used to teach the common conception of a phenomenon. Used thusly a toy model is a pedagogical tool. In both cases, the cognitive perspective on toy modelling in terms of mental models accounts for the way in which models function as arguments or learning interfaces. Therefore, mental models are a good model for the way in which toy models function both on the level of a community and on the level of an individual to improve understanding in a process of repeated exploration and analysis.

In summary, toy models are the means through which an individual can construct internal *ideal* representations of types of phenomena and acquire the ability to recognise the patterns that are characteristic of these types. The importance to scientific understanding derives precisely from the significance of toy models as a *shared external* representation of a type of phenomenon that affords a common *individual mental* representation of it and thus focuses research on this type.

We can now answer the questions raised by the introductory scene: the girl playing with building blocks gains understanding of and intuition about house building, in particular, she understands on which factors the stability of houses depends. When playing with the blocks and repeatedly comparing the result to a real house she constructs a mental model of houses. Thus she has obtained an understanding of toy house building and the conditions under which toy houses collapse. By comparing this (toy) mental model with the representation of a real house she can assert differences and similarities. In this sense she gains partial understanding of why an individual *real* house is stable, because their shape (resembled by the toy houses' shape) is an essential feature relevant to stability of houses. However, there will be exceptions in reality that do not conform to her internal model. In this case she will be surprised. Now a comparison with the toy model will help her focus on the features that are different and thus guide her search for an explanation.

Does a toy model explain its target? I have argued that toy models convey understanding in both an integrationist and a counterfactual sense. However, what distinguishes understanding with toy models from other kinds of explanatory understanding is the importance of the individual *activity* of manipulation and construction. *Reference* to a toy model in itself, as Bokulich (2011) requires for a model explanation, may not be sufficient for a particular agent to understand a phenomenon. It additionally requires an active reconstruction process. Nevertheless, explanation

[24] In the same vein, Polanyi (1961) construes the activity of discovery as "an oscillation between movements of analysis [comparison] and integration [construction]' (p. 130).

that makes reference to a toy model renders a phenomenon *understandable* in that it pinpoints *possible* connections between objects. Thus toy models may contribute to possible explanations. They can do so in two ways: (a) On the one hand they provide an ideal mechanism that can be part of a possible explanation. (b) On the other hand if the explaining agent understands the explanandum in the sense of **(U)** she will be able recognise the features and properties of the phenomenon that must be accounted for in an explanation. Thus she holds the prerequisites for the construction of an explanation. The latter seems to be the way in which the building block toy model contributes to the ability of the girl to explain the stability of houses. However, *only* if a translation key[25] to (what is commonly accepted as) the *real* underlying dynamics of a system can be given (as is determined by a general theory), can a toy model become *genuinely* explanatory for the type of phenomenon it represents. However, if toy models are used as a heuristic tool, when there is no theory that is accepted to expose the "real" connections they can only contribute to possible explanations[26]. Like understanding explanation is context-dependent: to what *extent* a toy-model explanation is explanatory for a particular agent, i.e., lets him understand the phenomenon, depends on the individual background of knowledge and skills.

It is now also clear why numerical calculations that involve approximations do not yield as comprehensive an understanding as toy models although they may fully represent all features of the target. It is because the means of gaining an intuition about the modelled phenomenon are much more intractable. While the individual scientist who performs a calculation may understand how the patterns that obtain as the results of a calculation come about, it is much less feasible for the scientists in the community to retrace the construction process of a model and the mathematical treatment of that model. The simplicity of toy models affords their comprehension at the cost of reducing the number of phenomena represented by it in comparison with a full-grown theory. My argument is one of degree - toy models are a particularly *feasible* way to render the manifold phenomena in the world comprehensible by a large group of people. They convey a general idea, but do not trace a concrete causal mechanism. Thereby they convey understanding. The idea embodied in a toy model is just specific enough to associate it with a distinct type of phenomenon, as opposed to a theory that encompasses *all* phenomena of a much more general type, or a very complex model that describes *one* concrete phenomenon.

In conclusion, I have argued that we can understand a phenomenon if we can comprehend a model that adequately represents that phenomenon. In particular, comprehension of a model implies the ability to recognise phenomena that exhibit a particular pattern that is characteristic of that model. In the process of exploring a toy model a scientist can gain intuition about that model in that he can relate it to known concepts or methods, and recognise the characteristic features of the model's

[25] Cf. (Bokulich, 2012)

[26] They are a means to extract "*stable phenomenologies* from unknown and, perhaps, unknowable theories" (Batterman, 2002, p. 35), cf. Section 4.1.

mechanism. This is facilitated by the toy model's tractability and high level of idealisation. Moreover, a toy model can adequately represent a type of phenomenon. Therefore, by exploring and studying toy models we can gain understanding of that type. Because a type is the set of phenomena that exhibit a common generic feature, we obtain partial understanding about the individual phenomena that are elements of the type.

Finally, intuition and understanding can be understood from a cognitive perspective. Specifically, I suggest that in virtue of their pragmatic virtues (simplicity, tractability) toy models are "manuals" for the construction of mental models. Mental models are a model for the way we reason about objects, organise knowledge, classify novel phenomena. In particular, they are a cognitive account for how recognition of patterns and association with other concepts functions. The common coding hypothesis offers a plausible account for how a mental model is constructed in the process of constructing and manipulating an (external) toy model: there is a shared representation of action, imagination and perception of objects. This account is limited to movement representation, or more generally, dynamics. However, scientific models are usually dynamical in nature so that for most cases it holds. This is true, in particular, for the case studies I presented at the beginning. In the interplay of both a communal and an individual process, toy models focus and guide scientific progress. They provide a shared interface on which scientists with different backgrounds can contribute ideas. This is made possible, because they represent a type of phenomenon that may be of interest to many different scientists, or even communities, and because they are feasibly tractable by scientists with different educational backgrounds. In a process of both individual exploration and analysis, and shared exploration and comparison of toy models understanding of the type of phenomenon is improved and a possible path towards a genuine explanation set.

5. CONCLUSION

This thesis is about how it comes that toy models are taken to convey understanding as opposed to numerical simulations of complex models. To find an answer I have investigated the three questions, (1) which functions toy models perform in science, (2) in virtue of what they do so, and (3) how performing these functions furnishes an understanding of the phenomena they represent.

Summary Toy models are highly idealized, exactly solvable models that have no specified domain of application. They are used either to learn about more complex theoretical models or theories, to make surprising phenomena tractable, or to provide a novel perspective on a phenomenon. Moreover, toy models have been shamefully neglected in the philosophical literature on models. This is not only because toy models are used pervasively in all sciences. Even more importantly toy models as those models that capture the *essential* features of a problem are at the core of scientific activity. It is those scientists who are skillful and creative (toy) modellers that are considered *good* scientists.

In **Chapter 2** I have studied three well-known and well-studied models from different sciences. In their presentation I focused not only on the motivation and formulation of the models themselves, but also on the way they are used in science up to today. We saw how their role, the way they are studied, and their applications have changed considerably since their inception. Although all three of the models are drastic oversimplifications of their targets they are still considered paradigmatic examples of how good modelling works.

From these case studies I extracted four functions that toy models can perform in **Chapter 3**: (i) toy models represent a type of phenomenon, (ii) by studying toy models scientists can explore the dependencies between model assumptions and model behaviour and thus learn about their theoretical tools, and (iii) they are pedagogical tools to teach and learn about their underlying theory. The possibility to explore a toy model together with its link to a type of phenomenon as a target afford (iv) their fruitful application in arguments. Toy models function as thought experiments the results of which substantiate possibility or impossibility claims about their target.

I argued that toy models are particularly apt to perform these functions in virtue of (a) their conceptual and (b) their pragmatic virtues. Specifically, (a) only if a model is highly idealised can it represent a *type* of phenomenon and thus a broad range of different individual phenomena. The (mathematical) structure of a toy model provides the language and concepts in which scientists can talk about this type of phenomenon. It also focuses and guides research about these phenomena in that it is a well-understood reference frame or baseline against which they can be compared. A high level of idealisation and a (mathematical) structure are therefore the conceptual virtues of a toy model. (b) The pragmatic virtues of toy models are their mathematical tractability and the open application domain. These two features not only allow but are necessary for a free exploration of the dependencies within the model, of the techniques required for their solution, and the phenomena they might be applied to. Both pragmatic and conceptual virtues are crucial in their use as qualified arguments.

Finally, in **Chapter 4** I have suggested an account of understanding and set it in relation with the notions of understanding and intuition. Specifically, I construe understanding of a phenomenon to require both that the scientific agent comprehend a model of that phenomenon and that the model adequately represent that phenomenon. Comprehending a model implies the ability to "recognise qualitatively characteristic consequences" (de Regt and Dieks, 2005) of that model. With this characterisation of understanding at hand my main thesis falls into place: toy models provide understanding of the type of phenomenon they represent. Thus they provide partial understanding of the elements of that type. This is because through the exploration of a toy model can a scientific agent develop comprehension of that model in that she integrates it within her conceptual framework, gains knowledge about the counterfactual patterns of the model and develops an intuition about it, that is, acquires the ability to recognise these patterns.

The question remains *how* she does so. To answer it, I draw on cognitive science, specifically, the "mental model" and the "common coding" hypotheses. Mental models, structural representation of objects or processes that are constructed by the mind, are a model for how knowledge is organised in the mind, and how we reason about objects. As such it accounts for how we recognise the characteristic patterns of a toy model in other phenomena, and how we associate it with already known concepts. These are two key features of understanding. The common coding hypothesis makes sense of how we gain intuition about a model from exploring it. According to it there is a shared representation between imagination, action and perception in the brain. Thus, manipulating a model in the exploration process we create a mental model of the model and its counterfactual dependencies. This model influences our perception, too. It gives us the ability to *see* patterns as a whole. Toy models as external, sharable representations of both their targets and the internal model of every individual scientist focus and guide the communal activity of scientific progress. In a process of exploration, analysis, and sharing is the common and individual understanding of a toy model's target improved. Understanding is embodied in the ability to take different perspectives on a phenomenon - an "integrationist" and a "particularist" one.

Outlook What I have presented, illustrated and argued in this thesis is, and can only be, a sketch of how toy models function in science. Future work must be directed at every one of the aspects I have taken a look at here. In particular, this regards the last chapter on understanding. For example, the notions of explanation, understanding, and intuition are still more than vague. Therefore a more precise and comprehensible account of understanding and intuition, but also their relation is called for. Such an account cannot take a purely philosophical perspective, but needs to take into account the intrinsically cognitive aspects of these notions.

Moreover, my tentative claims about mental models and common coding rest on shaky foundations. While some progress has been made since the inception of these hypotheses, cognitive science is still at an infancy stage. Specifically, the aim must be to extend the scope of the work on mental models and common coding to the realm of complex mental activity that is relevant to science. Nowadays the studies usually concern very simple mental activities and movements. Only with such an account, and ideally, a comprehensive understanding of how the brain works can model building *activity* in science be thoroughly understood.

However, before this is in sight the obvious smaller steps should be taken. In particular, the basis of case studies that philosophical accounts (including my own) draw on is very limited. The analysis of many more toy models in a wide variety of contexts promises a much more detailed and comprehensive understanding of their role than what I could have presented here. In passing, I have drawn on a few further examples in the course of this thesis.

In another vein, the way in which I have made sense of toy modelling activity may be a useful perspective to take on the recently surging method of analog simulations[1]. In these simulations experimentally inaccessible phenomena such as black holes are simulated on tabletop setups of quantum systems. That is, the theoretical model of black holes is implemented in a quantum system over which scientists have a large degree of control. Thus scientists have the chance to experiment on the model actively. In that sense analog simulations are the experimentalist's playground that is equivalent the theorist's toy models. In both cases the connections established between different systems and the high level of control act as generators for new ideas that guide the direction of research. Hence, these "physical games" can lead both to new technological innovations and fundamental insights.

[1] The simulation of the Ising model in an ion-trap system by Britton et al. (2012b) is an example of an analog simulation.

REFERENCES

Akçakaya, H. R., R. Arditi, and L. R. Ginzburg (1995). Ratio-dependent predation: An abstraction that works. *Ecology 76*(3), 995–1004.

Apel, K.-O. (2001). Verstehen. In J. Ritter (Ed.), *Historisches Wörterbuch der Philosophie*, Volume 11: U-V, pp. 918–938. Darmstadt: Wissenschaftliche Buchgesellschaft.

Arnold, E. (2010). Tools or Toys? On Specific Challenges for Modeling and the Epistemology of Models and Computer Simulations in the Social Sciences. Models & Simulations 4 (Toronto, Ontario; May 7-9, 2010). From http://philsci-archive.pitt.edu/5424/, accessed 6.11.2013.

Aydinonat, N. (2008). *The Invisible Hand in Economics: How Economists Explain Unintended Social Consequences.* London: Routledge.

Bailer-Jones, D. M. (1997, November). *Scientific Models: A Cognitive Approach with an Application in Astrophyiscs.* Ph. D. thesis, University of Cambridge.

Bailer-Jones, D. M. (2009). *Scientific models in philosophy of science.* University of Pittsburgh Press.

Batterman, R. W. (2002). Asymptotics and the role of minimal models. *The British Journal for the Philosophy of Science 53*(1), 21–38.

Bokulich, A. (2011). How scientific models can explain. *Synthese 180*(1), 33–45.

Bokulich, A. (2012). Distinguishing Explanatory from Nonexplanatory Fictions. *Philosophy of Science 79*(5), 725–737.

Britton, J. W., B. C. Sawyer, A. C. Keith, C.-C. J. Wang, J. K. Freericks, H. Uys, M. J. Biercuk, and J. J. Bollinger (2012a). Engineered two-dimensional Ising interactions in a trapped-ion quantum simulator with hundreds of spins. *Nature Physics 484*, 489–492.

Britton, J. W., B. C. Sawyer, A. C. Keith, C.-C. J. Wang, J. K. Freericks, H. Uys, M. J. Biercuk, and J. J. Bollinger (2012b). Engineered two-dimensional Ising interactions in a trapped-ion quantum simulator with hundreds of spins. *Nature Physics 484*, 489 –492.

Burley, D. M. (1960). Some magnetic properties of the Ising model. *Philosophical Magazine 5*(57), 909–919.

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

Cartwright, N. (1989). *Nature's capacities and their measurement.* Oxford University Press.

Chandrasekharan, S. (2009). Building to discover: A common coding model. *Cognitive Science 33*(6), 1059–1086.

de Regt, H. W. and D. Dieks (2005). A contextual approach to scientific understanding. *Synthese 144*(1), 137–170.

DeGracia, D. J., Z.-F. Huang, and S. Huang (2012). A nonlinear dynamical theory of cell injury. *Journal of Cerebral Blood Flow & Metabolism 32*, 1000–1013.

Dilthey, W. (1894). *Ideen über eine beschreibende und zergliedernde Psychologie.* Verlag der Königlichen Akademie der Wissenschaften.

Epstein, J. M. and R. Axtell (1996). *Growing articficial societies: Social science from the bottom up.* Washington, D.C.: Brookings Institution Press.

Fagiolo, G., M. Valente, and N. J. Vriend (2007). Segregation in networks. *Journal of Economic Behavior & Organization 64*(3 -4), 316 – 336.

Fisher, M. E. (1964). Correlation functions and the critical region of simple fluids. *Journal of Mathematical Physics 5*, 944 962.

Fisher, M. E. (1981). Simple Ising models still thrive. *Physica 106A*, 28 – 47.

Fisher, M. E. (1983). Scaling, university and renormalization group theory. In F. Hahne (Ed.), *Critical Phenomena*, Volume 186 of *Lecture Notes in Physics*, pp. 1–139. Berlin Heidelberg: Springer.

Fossett, M. (2006). Including preference and social distance dynamics in multi-factor theories of segregation. *The Journal of Mathematical Sociology 30*(3-4), 289–298.

Frigg, R. and S. Hartmann (2012). Models in science. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2012 ed.).

Gibbard, A. and H. R. Varian (1978). Economic models. *The Journal of Philosophy 75*(11), 664–677.

Griffiths, R. B. (1970). Dependence of critical indices on a parameter. *Physical Review Letters 24*, 1479–1482.

Grüne-Yanoff, T. (2009). Learning from minimal economic models. *Erkenntnis 70*(1), 81–99.

Grüne-Yanoff, T. (2013). Appraising Models Nonrepresentationally. *Philosophy of Science 0*(ja).

Hartmann, S. (1995). Models as a tool for theory construction: some strategies of preliminary physics. In W. Herfel (Ed.), *Theories and Models in Scientific Processes*, Volume 44 of *Poznań studies in the philosophy of the sciences and the humanities*, pp. 49–67. Amsterdam: Rodopi.

Hartmann, S. (1998). Idealization in Quantum Field Theory. In N. Shanks (Ed.), *Idealization in Contemporary Physics*, pp. 99–122. Amsterdam: Rodopi.

Hartmann, S. (2001). Effective field theories, reductionism and scientific explanation. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 32*(2), 267 – 304.

Hartmann, S. and D. Bailer-Jones (1999). Modelle. In H.-J. Sandkühler (Ed.), *Enzyklopädie der Philosophie*, pp. 854–859. Hamburg: Meiner.

Hausman, D. M. (1992). *The Inexact and Separate Science of Economics.* Cambridge: Cambridge University Press.

Heidegger, M. (1927). *Sein und Zeit* (19 ed.). Tübingen: Max Niemayer Verlag.

Heisenberg, W. (1927). Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift für Physik 43*(3-4), 172–198.

Hempel, C. G. (1966). *Philosophy of natural science.* Foundations of philosophy series. Prentice-Hall.

Henry, A. D., P. Prałat, and C.-Q. Zhang (2011). Emergence of segregation in evolving social networks. *Proceedings of the National Academy of Sciences 108*(21), 8605–8610.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences 79*(8), 2554–2558.

Huang, K. (1987). *Statistical Physics* (Second ed.). New York: John Wiley & Sons.

Hughes, R. I. G. (1997). Models and representation. *Philosophy of Science 64*, S325–S336.

Hughes, R. I. G. (1999). The Ising model, computer simulation, and universal physics. In M. Morgan and M. Morrison (Eds.), *Models as Mediators.* Cambridge University Press.

Husserl, E. (1985). "encyclopaedia britannica"-artikel (1927). In K. Held (Ed.), *Die Phänomenologische Methode*, pp. 196–224. Stuttgart: Reclam.

Ising, E. (1925). Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift für Physik 31*(1), 253–258.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*, Volume 6. Harvard University Press.

Kadanoff, L. P. (1966). Scaling laws for Ising models near T_c. *Physics 2*, 263–272.

Kitcher, P. (1989). Explanatory Unification and the Causal Structure of the World. In P. Kitcher and W. Salmon (Eds.), *Scientific Explanation*, pp. 410–505. Minneapolis: University of Minneapolis Press.

Kobusch, T. (1976). Intuition. In J. Ritter (Ed.), *Historisches Wörterbuch der Philosophie*, Volume 4: I-K, pp. 524–540. Darmstadt: Wissenschaftliche Buchgesellschaft.

Kramers, H. A. and G. H. Wannier (1941). Statistics of the two-dimensional ferromagnet. part i. *Physical Review 60*, 252–262.

Lakoff, G. (1987). *Women, Fire and Dangerous Things*. Chicago: The University of Chicago Press.

Laurie, A. J. and N. K. Jaggi (2003). Role of 'vision' in neighbourhood racial segregation: A variant of the Schelling Segregation Model. *Urban Studies 40*(13), 2687–2704.

Lenk, H. (1972). Erklären, erklärung. In J. Ritter (Ed.), *Historisches Wörterbuch der Philosophie*, Volume 2: D-F, pp. 690–701. Darmstadt: Wissenschaftliche Buchgesellschaft.

Lenz, W. (1920). Beitrag zum Verständnis der magnetischen Erscheinungen in festen Körpern. *Physikalische Zeitschrift 21*, 613–615.

Lotka, A. J. (1910). Contribution to the Theory of Periodic Reactions. *The Journal of Physical Chemistry 14*(3), 271–274.

Lotka, A. J. (1925). *Elements of physical biology*. Baltimore: Williams & Wilkins.

Macy, M. W. and A. van de Rijt (2006). Ethnic preferences and residential segregation: Theoretical explorations beyond Detroit. *The Journal of Mathematical Sociology 30*(3-4), 275–288.

Magnani, L. (2004). Model-based and manipulative abduction in science. *Foundations of Science 9*(3), 219–247.

Magnani, L. (2012). Scientific Models Are Not Fictions. In L. Magnani and P. Li (Eds.), *Philosophy and Cognitive Science*, Volume 2 of *Studies in Applied Philosophy, Epistemology and Rational Ethics*, pp. 1–38. Berlin Heidelberg: Springer.

Mäki, U. (2005). Models are experiments, experiments are models. *Journal of Economic Methodology 12*(2), 303–315.

Mäki, U. (2009). MISSing the World. Models as Isolations and Credible Surrogate Systems. *Erkenntnis 70*(1), 29–43.

McMullin, E. (1985). Galilean idealization. *Studies in History and Philosophy of Science 16*(3), 247–273.

Mendoza-Arenas, J. J., T. Grujic, D. Jaksch, and S. R. Clark (2013). Dephasing enhanced transport in non-equilibrium strongly-correlated quantum systems. *arXiv*, 1302.5629.

Morgan, M. and M. Morrison (1999). Models as mediating instruments. In M. Morgan and M. Morrison (Eds.), *Models as Mediators*. Cambridge: Cambridge University Press.

Morgan, M. S. (2001). Models, stories and the economic world. *Journal of Economic Methodology 8*(3), 361–384.

Morgan, M. S. (2004). Imagination and imaging in model building. *Philosophy of Science 71*(5), 753–766.

Morgan, M. S. (2005). Experiments versus models: New phenomena, inference and surprise. *Journal of Economic Methodology 12*(2), 317–329.

Morrison, M. (2009a). Fictions, Representations, and Reality. In Mauricio Suárez (Ed.), *Fictions in Science*, pp. 139–157. New York: Routledge.

Morrison, M. (2009b). Models, measurement and computer simulation: the changing face of experimentation. *Philosophical Studies 143*(1), 33–57.

Muldoon, R., T. Smith, and M. Weisberg (2012). Segregation that no one seeks. *Philosophy of Science 79*(1), 38–62.

Murray, J. D. (2002). *Mathematical biology* (3 ed.). Springer.

Nersessian, N. J. (1999). Model-based reasoning in conceptual change. In L. Magnani, N. J. Nersessian, and P. Thagard (Eds.), *Model-based reasoning in scientific discovery*, pp. 5–22. New York: Kluwer Academic/Plenum Publishers.

Nersessian, N. J. and S. Chandrasekharan (2009). Hybrid analogies in conceptual innovation in science. *Cognitive Systems Research 10*(3), 178 – 188.

Niss, M. (2005). History of the Lenz-Ising model 1920 1950: From ferromagnetic to cooperative phenomena. *Archive for History of Exact Sciences 59*(3), 267–318.

Niss, M. (2009). History of the Lenz Ising model 1950 1965: from irrelevance to relevance. *Archive for History of Exact Sciences 63*(3), 243–287.

Niss, M. (2011). History of the Lenz Ising model 1965 1971: the role of a simple model in understanding critical phenomena. *Archive for History of Exact Sciences 65*(6), 625–658.

Onsager, L. (1944). Crystal Statistics. I. A Two-Dimensional Model with an Order-Disorder Transition. *Physical Review 65*, 117–149.

Pancs, R. and N. J. Vriend (2007). Schelling's spatial proximity model of segregation revisited. *Journal of Public Economics 91*(1 2), 1 – 24.

Picht, G. (1969). *Mut zur Utopie*. München: Piper.

Polanyi, M. (1961). Ii. knowing and being. *Mind LXX*(280), 458–470.

Pollicott, M. and H. Weiss (2001). The dynamics of Schelling-type segregation models and a nonlinear graph laplacian variational problem. *Advances in Applied Mathematics 27*(1), 17 – 40.

Pust, J. (2012). Intuition. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2012 ed.).

Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

Schelling, T. C. (1971). Dynamic models of segregation. *The Journal of Mathematical Sociology 1*(2), 143–186.

Schelling, T. C. (1978). *Micromotives and Macrobehavior*. W. W. Norton.

Schneidman, E., M. J. Berry, R. Segev, and W. Bialek (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature 440*, 1007 – 1012.

Suárez, M. (Ed.) (2009). *Fictions in science: Philosophical essays on modeling and idealization*. New York: Routledge.

Sugden, R. (2000). Credible worlds: the status of theoretical models in economics. *Journal of Economic Methodology 7*(1), 1–31.

Van Vleck, J. H. (1945). A survey of the theory of ferromagnetism. *Review of Modern Physics 17*, 27–47.

Volterra, V. (1926a). Fluctuations in the abundance of species considered mathematically. *Nature 118*(2972), 558–560.

Volterra, V. (1926b). Variazioni e fluttuazioni del numero d'inidvidui in specie animali conviventi. In *Memorie Della R. Academia Nazionale Dei Lincei*, Volume 2 of *VI*, pp. 5–112.

Volterra, V. (1931). Variations and fluctuations of the number of individuals in animal species living together. In R. N. Chapman (Ed.), *Animal Ecology*. New York & London: McGraw-Hill.

Weber, M. (1991a). Der Sinn der ,,Wertfreiheit"der soziologischen und ökonomischen Wissenschaften (1917). In M. Sukale (Ed.), *Schriften zur Wissenschaftslehre*, pp. 176–236. Reclam.

Weber, M. (1991b). Die ,,Objektivität" sozialwissenschaftlicher und sozialpolitischer Erkenntnis (1904). In M. Sukale (Ed.), *Schriften zur Wissenschaftslehre*, pp. 21–101. Ditzingen: Reclam.

Weisberg, M. (2007). Who is a modeler? *The British Journal for the Philosophy of Science 58*(2), 207–233.

Wikipedia (2014). Spin-glas. http://de.wikipedia.org/wiki/Spinglas, Accessed on 9/1/2014.

Wilson, R. A. and L. Foglia (2011). Embodied cognition. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2011 ed.).

Winsberg, E. (2013). Computer simulations in science. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2013 ed.).

Woodward, J. (2003). *Making Things Happen*. Oxford: Oxford University Press.

Woodward, J. (2011). Scientific Explanation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2011 ed.).

Ylikoski, P. and N. E. Aydinonat (2013). Understanding with Theoretical Models. Working paper, University of Helsinki and Bahçeşehir University.

Zhang, J. (2004a). A dynamic model of residential segregation. *The Journal of Mathematical Sociology 28*(3), 147–170.

Zhang, J. (2004b). Residential segregation in an all-integrationist world. *Journal of Economic Behavior & Organization 54*(4), 533 – 550.

Ziman, J. M. (1965). Mathematical models and physical toys. *Nature 206*(4990), 1187–1192.

ERKLÄRUNG

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Die Stellen der Arbeit, die anderen Quellen im Wortlaut oder dem Sinn nach entnommen wurden, sind durch Angaben der Herkunft kenntlich gemacht. Dies gilt auch für Zeichnungen, Skizzen, bildliche Darstellungen sowie für Quellen aus dem Internet.

München, den 13. Januar 2014