Astrophysics in Simulacrum
The Epistemological Role of Computer Simulation in Dark Matter Studies

Katia Stephanie Wilson

ORCID iD: 0000-0002-3989-3032
Doctor of Philosophy
December 2016

History and Philosophy of Science
School of Historical and Philosophical Studies
The University of Melbourne

Submitted in total fulfilment of the requirements of the degree of Doctor of Philosophy.
Abstract

Computer simulation is a technique that is widely used across many scientific disciplines. In areas like astrophysics, the day-to-day practice of doing science is in fact dominated by simulation. In the face of simulation’s ubiquity and usefulness, this thesis asks the question: how can the virtual teach us about the real? The first part of the thesis introduces and historicises the main philosophical issues that surround simulation in science. The historical and philosophical literature on simulation in science is chiefly concerned with the relationships between simulation, theory, and experiment, though there is little agreement on specifics. The relationships between simulation and theory, and simulation and experiment, are complex, and result in what I describe as theory crafting. This iterative and interactive way of generating knowledge involves using simulation to draw together multiple sources of evidence. In order to more clearly tease out how simulation fits in with existing ways of knowledge making, the journal Simulation is used to provide a historicised look at the main themes of simulation epistemology from 1960 to the present. The shifts within these themes helps break down simulation’s various associations with models and experiment, and also provide a partial answer to how the simulation technique became established as a legitimate tool for producing knowledge.

The second part of the thesis consists of two physical case studies drawn from modern astrophysics that were discovered with and solved using simulation. The bar instability problem, and the case of the missing satellites, provide two different perspectives of the role of simulation in solving problems, particularly in interaction with observation. In the first case study, simulation is shown to take on roles that are traditionally the purview of theory and experiment, demonstrating a highly fruitful flexibility. This flexibility, as the second case study shows, gives simulation the capacity to construct self-sufficient universes that can be the subject of observation and measurement. In such a manner simulation draws out the empirical content of theory, resulting in data that can be considered epistemologically on-par with observational or experimental data. It through the superposition of the simulated virtual worlds with our own that new knowledge emerges, the outcome of a complex negotiation process between the evidence, theory, model, and the simulation itself that is transformative of the categories involved.
Declaration

This is to certify that:

(i) the thesis comprises only my original work towards the PhD
(ii) due acknowledgement has been made in the text to all other material used; and
(iii) the thesis is fewer than 100,000 words in length, exclusive of tables, maps, bibliographies, and appendices

Signed, Katia Stephanie Wilson, 5 December 2016
Acknowledgments

I would like to thank my principal supervisor, Dr. Kristian Camilleri, for his truly invaluable advice over the course of my graduate studies. This thesis has benefitted greatly from his critical eye, searching questions, and patient re-reading of drafts.

I also extend my gratitude to my secondary supervisor, Dr. Gerhard Wiesenfeldt, whose insightful comments on later drafts helped in clarifying both my arguments and my writing.

This thesis would not have been possible without the support of my family. My parents, Rita and Charles, have provided an endless well of encouragement and strength. I am deeply grateful for the time and emotional support they have given me. I also extend my sincerest thanks to my brother Richard for his optimism and sympathetic ear, and to my sister-in-law Kathryn for many commiserating conversations with a fellow PhD candidate.
Table of Contents

Introduction

1 The Philosophy of Simulation
   1.1 Introduction
   1.2 Materiality and Experiment
   1.3 Models and Simulation
      1.3.1 The Simulation Model
      1.3.2 Semi-Analytic Models
      1.3.3 Theory Crafting
   1.4 Theory and Experiment, and Simulation
   1.5 Philosophical Novelty

2 Man and Machine: Themes in Simulation from 1963 to 2016
   2.1 Introduction
   2.2 The Physical Computer
   2.3 Models and the Modeller
   2.4 Man and Machine
   2.5 Selling Simulation
   2.6 Supersimulations and Seduction
      2.6.1 Simulating as Science
      2.6.2 Seduction
   2.7 Characterising Simulation
      2.7.1 From Traditional to Sui Generis
      2.7.2 The Value of a Historicised Perspective

3 The Simulacrum of an Experimental Science: The Bar Instability and Dark Halos
   3.1 Introduction
   3.2 Miller and the Metaphor of Experiment
   3.3 A Local Instability
   3.4 Toomre’s Criterion and The Narrative of Theory and Experiment
   3.5 The Bar Instability
   3.6 The Ostriker-Peebles Conjecture
   3.7 The Problem with Halos
   3.8 Resolving the Instability
3.9 Persuasive Pictures: Simulation and Observation 115
3.10 Reconstructing Astrophysics as an Experimental Science 123

4 The Case of the Missing Satellites: Technoscience and Astrophysics
4.1 Introduction: Technoscience in Astrophysics? 125
4.2 The Problem 128
4.3 Possible Solutions 132
4.4 ‘Future Observations’ 135
4.5 The ‘Solution’ 143
   4.5.1 Hydrodynamic Simulations 145
4.6 Technical Negotiation 148
4.7 Simulation as an Epistemic Target 154
   4.7.1 Simulation and Empirical Systems 161
   4.7.2 Fear of Seduction (Again) 165
4.8 Conclusion 168

5 Generating Knowledge with Simulation
5.1 Introduction 171
5.2 The Empirical Content of Theory 172
5.3 Visualisation and Simulated Observations 177
5.4 Returning to the Issue of Philosophical Novelty 180
   5.4.1 Styles of Reasoning 180
   5.4.2 Cultures 182
5.5 Excursus: The (De)contextualisation of Simulation 184
5.6 Conclusion 187
List of Figures

Figure 1: Number of publications in astronomy/astrophysics per year referencing the phrase ‘semi-analytic’ in the topic field. 18
Figure 2: “To provide a proper framework to review the credibility of a simulation, it is convenient to divide the simulation environment into three basic elements as depicted.” 45
Figure 3: “Simulation of Mechanical Systems adds an Important Dimension to Instruction in Engineering Design.” 50
Figure 4: Systematic moves of many particles in the game described by Miller and Prendergast. 81
Figure 5: Toomre’s analysis showed a threshold velocity dispersion is required. 88
Figure 6: An illustration of the computer model used by Hockney and Hohl showing the N by N array of cells used to calculate the gravitational potential. 90
Figure 7: “Break up of an initially uniformly rotating and balanced cold disc of 50,000 stars.” 91
Figure 8: Hohl’s simulation demonstrates the formation of an unstable bar-like structure with two arms. 96
Figure 9: “Effect of the halo on the evolution of the model galaxy.” 101
Figure 10: “Status of the wave theories. As yet, only the thickened portions of this schematic route seem reasonably secure.” 106
Figure 11: Evolution of a leading density wave in Zang’s simulation, showing the growth of spiral pattern through swing amplification. 108
Figure 12: “Superposition of all 21 Sc rotation curves.” 110
Figure 13: The total rotation curve for the galactic model is the top curve. The other curves show the contributions to the total curve from the four mass components. 112
Figure 14: Hohl compares the simulated image (left), and the atlas photograph (right) by placing them side by side. 119
Figure 15: Distribution of dark matter halos in the KKVP simulation. 131
Figure 16: Luminosity function as observed (red dot-dashed lower curve), as corrected for sky coverage (green dashed middle curve), and with all corrections included (blue solid upper curve). 138
Figure 17: The effect of different physical processes on the luminosity function of satellites in one of the simulated halos. 145
Figure 18: Distribution of dark matter (left) and stars (right) of the same region in a Local Group hydrodynamic simulation. 147
Figure 19: Left: Hubble Space Telescope (HST) Ultra Deep Field (UDF) image, an observation. Right: HST mock observation from Illustris, a simulation. 156
Figure 20: Sample galaxies from Illustris arranged by standard morphological classification. 159

For copyright reasons, figures 2 to 20 have been removed from the open access version of this thesis.
Introduction

The ubiquity of computer simulation across the entirety of modern science has in the last decade or so drawn the attention of historians and philosophers of science.\textsuperscript{1} Case studies on simulation in nuclear physics, particle physics, climate science, nanotechnology, evolutionary biology, epidemiology, economics, and social science have emerged, accompanied by a concerted effort to make philosophical sense of simulation in science. Though often mentioned in these studies as an example of a simulation-dominated discipline, computer simulations in astrophysics have received little direct attention. Yet astrophysics provides us with a unique perspective on how simulations are used to produce knowledge. Firstly, astrophysics is a discipline without experiments, as the targets of study are too far away (in time and space) or too unwieldy to submit themselves to manipulation or intervention. Secondly, direct observation of many astrophysical objects can be difficult, and the resultant data is often ambiguous or subject to uncertainties. Dark matter in particular is resistant to direct observation; it literally cannot, even in principle, be observed. Dark matter studies in astrophysics therefore provide us with a situation where experiment is not available to provide empirical evidence, and observation's capacity for the same has been severely curtailed. What empirical evidence is available is almost always ambiguous – and yet, with the use of simulations, scientists still somehow manage to progress knowledge in these areas.\textsuperscript{2}

The central motivation for this thesis is to understand how computer simulations generate new knowledge. The thesis does not seek to provide a definitive answer to this ambitious question, but it does aim to clarify some of the issues at stake and in doing so contribute to the growing body of historical and philosophical literature on simulation by providing insight into the role of simulation in science in areas where traditional experiment has no place. The focus of this thesis is therefore epistemological; broadly, I ask the question: how can the virtual teach us about the real? To actually turn this question into something answerable, this thesis undertakes to answer a narrower question that nevertheless addresses many of the same underlying issues: how does astrophysics use simulation to gain knowledge, particularly about dark matter?

In answering this question, the thesis undertakes an in-depth exploration of two problems (and their ‘solutions’) in modern astrophysics that were discovered by and solved using simulation. The overall methodology is very much that of an integrated approach to history and philosophy of science, though the thesis will tend more towards the philosophical than the historical (with the possible exception of Chapter 2). The primary material is sourced from scientific publications and analysed using concepts found in both the philosophical literature on simulation and, where

\textsuperscript{1} See Winsberg, 2009a and Grüne-Yanoff & Weirich, 2010 for reviews. For this thesis, read ‘simulation’ as ‘computer simulation’.

\textsuperscript{2} Climate science finds itself in a very similar boat methodologically speaking, but is also highly politically charged, which necessarily colours the use and interpretation of simulation and simulation results (see for example Norton & Suppe, 2001 and Lahren, 2005). It is in fact suggested in this thesis that simulation looks and behaves differently across different disciplines, especially where it intersects with other forms of knowledge generation, like experiment (see section 1.5).
pertinent, in history and philosophy of science (HPS) more generally. Though the philosophical analysis of the case studies is central, it is interwoven with and embedded into the historical story. This method of analysis – and treatment of scientific concepts – sees the historical perspective as essential for understanding the philosophical one.

The thesis goes from generality (in Chapters 1 and 2) to specificity (in Chapters 3 and 4). While the chapters can be taken as self-contained (with the natural exception of the concluding chapter), in providing answers to different facets of the motivating question, the chapters are all linked by a multitude of common strands and overlapping themes. In looking at dark matter studies in particular, this thesis focuses on simulation. The hope is that, by adopting this focus, the case study analyses will be able enhance our understanding of the epistemological activity of simulation qua simulation, rather than being distracted, for example, by the interplay between simulation and experiment.

That being said, the very nature of the motivating question is epistemological and it is therefore necessary to explore the relationships between simulation and other modes of knowledge generation, namely experiment and theory. A good deal of the HPS literature on simulation is concerned with understanding these links; Chapter 1 (‘The Philosophy of Simulation’) acts as both a literature review and an introduction into the main philosophical themes of this thesis. Recurring concepts include the relationships between simulation, experiment, and materiality; between simulation and models; and between simulation, experiment, and theory. From Chapter 1 emerges the understanding of simulation as a distinct technique that nevertheless retains important continuities with established ways of generating knowledge. While simulation is not material like experiment, it does have that crucial manipulability; and though the model is a very important part of simulation, simulation is not reducible to ‘the model’. In fact, the relationship between simulation, theory (modelling), and experiment is revealed to be far more tightly coupled than it is usually portrayed.

Chapter 2, ‘Man and Machine: Themes in Simulation from 1963 to 2016’ then situates all this philosophy historically by undertaking the analysis of a simulation journal. Aptly titled Simulation, this journal provides a look at how simulation has been represented over the span of half a century, and allows us to do so in a manner that directs the gaze fully at simulation as a technique, and simulating as a profession. In identifying several themes that have shifted the focus of simulation practitioners, this chapter functions both historically and historiographically. The understanding of the scientist as simulationist is constructed over the course of the chapter.

---

3 The term HPS is used to broadly indicate the literature that talks about simulation in science from a philosophical, historical, or sociological perspective (or some combination thereof), including work found in science and technology studies (STS), science studies, and historical epistemology.

4 On the integrated approach to HPS, see Schickore, 2011 and Mauskopf & Schmaltz, 2012. Examples of works that I see as taking a similar integrated approach are Hasok Chang’s Inventing Temperature (Chang, 2004) and Nicolas Rasmussen’s Picture Control (Rasmussen, 1997). There is also some methodology in common with Lorraine Daston and Peter Galison’s Objectivity (Daston & Galison, 2007), which practices a sort of historical epistemology that understands epistemological categories – like objectivity, and with relevance to this thesis, empirical evidence – as historicised. For historical epistemology more generally see Feest & Sturm, 2011.
suggesting a way in which to understand the role of the modern scientist in disciplines where ‘doing science’ mainly involves simulating. The themes that emerge out of Simulation’s papers point to several of the same aspects of simulation – experiment, modelling, the physical computer – that are dealt with in Chapter 1, suggesting that the philosophy of simulation would benefit from considering simulation in a historical context. Out of the aforementioned themes, Chapter 2 also sketches a plausible reason for why simulation came to be adopted as a legitimate scientific technique so quickly and so ubiquitously: through borrowing credibility from existing scientific traditions. It is thus emphasised that simulation’s apparent similarity to both experiment and theory have played important roles in justifying simulation knowledge.

The second part of the thesis contains the two detailed astrophysical case studies. The first case study – outlined in Chapter 3, ‘The Simulacrum of an Experimental Science: The Bar Instability and Dark Halos’ – begins in 1968 with one of the first computer simulations of a spiral galaxy. Simulations of the early galaxy model showed it to be unstable, and Chapter 3 tells the story of how this instability was resolved, and how dark halos came to be included in the galactic model. This case study focuses on the language of the scientific literature in order to expose how simulation took on some of the epistemological roles of first ‘experiment’ and then of ‘theory’. Through such construction of classical scientific narratives, simulation helps develop and ‘confirm’ a new three-component model of a spiral galaxy. This case study brings to the fore simulation’s flexibility in moving between experimental and theoretical roles; simulation is able to both predict and confirm. Accompanying the text that used simulation as experiment in both a metaphorical and a narrative sense, the audience is also invited to understand the images of the simulation as ‘observations’ of the ‘experimental’ system. In this manner, astrophysics is reconstructed, in the postmodern sense, as an experimental science.

The second case study – Chapter 4, ‘The Case of the Missing Satellites: Technoscience and Astrophysics’ – then jumps ahead thirty or so years to 1999, when simulations revealed a problem with the model of galaxy formation that has only been ‘solved’ in the last few years. The missing satellites problem surfaced when simulations of the Local Group of Andromeda and the Milky Way predicted that we should see several more small neighbouring galaxies than we actually do. Highly advanced simulations in the 2000s and in the early part of this decade have come up with a solution to this problem, which, as Chapter 4 relates, involves a highly complex process of negotiation between observational data and simulation results. The act of negotiation mediated by simulation deeply entangles theoretical and empirical data. This chapter borrows an idea from technoscience – namely the assumption that knowledge is embedded in the technology that produces it – to suggest that simulation has the capacity to turn theory into a meaningful target of investigation. It is through using simulation to combine the ‘empirical content’ of the theory with observational data that a solution to the missing satellites problem is found. One of the outcomes of this blending of data sets is the emergence of ‘plausibility’ as a legitimate ontological criterion.

The final chapter of the thesis, Chapter 5, ‘Generating Knowledge with Simulation’, draws together the preceding four chapters and underscores the common themes and main outcomes
of the study, including a tentative answer to the question ‘how can the virtual teach us about the real?’ It will be concluded that it is the ability of simulation to act as both theory and experiment, for simulation to be epistemically and methodologically flexible, that allows for the ‘superposition’ of real and virtual worlds. As ‘experiment’ (or ‘observation’), simulation constructs self-sufficient universes that can be the subject of observation and measurement. As ‘theory’, simulation can then compare these virtual worlds with the real one. New knowledge emerges through the comparison of the simulation worlds with our own, but the way in which this happens is far from straightforward. The act of producing knowledge with simulation requires a negotiation between the evidence, the theory, the model, and the simulation itself that is transformative of the categories involved.

The comparison of real and virtual results in a superposition: as the virtual and real worlds are overlapped, they are also merged. It is through the blending of simulation data and observational data that astrophysicists can negotiate knowledge of dark objects out of ambiguous evidence. The thesis of this thesis is therefore that simulation produces knowledge in astrophysics by establishing worlds that allow the empirical content of theory to be gathered, and then by negotiating the blending of this data with the empirical data from observation. In this process, simulation may come to constitute a new kind of evidentiary base, granting the technique more epistemic power than is perhaps currently seen in the HPS literature. For scientists in disciplines where unambiguous or plentiful empirical data is difficult to come by (like astrophysics), and which therefore do not fit in with traditional narratives of generating knowledge, simulation becomes a powerful tool with which to do science.
1 The Philosophy of Simulation

1.1 Introduction
Many definitions have been proffered that attempt to capture the core features of simulation. Take, for example, the compilation of definitions of simulation that appears in the journal *Simulation* in 1979. The definitions are mostly from simulation textbooks (both general, and specific to various disciplines), as well as some of the journal’s previous editorials. A brief review of the list reveals some common themes; of the 22 definitions offered: 19 mention in some capacity models or modelling; 13 refer to a representational capacity (of some real system); 11 mention ‘experiment’ or manipulation; 8 emphasise a dynamic or temporal process (‘over time’); and 6 make explicit reference to the use of the computer. This chapter will not attempt to provide yet another definition of simulation – a tricky task, as suggested by the fact that several of the twenty-two definitions disagree on specifics. However, many of the same themes that occur when ‘defining’ simulation are reflected in the HPS literature. What this chapter does intend to do is introduce the main philosophical issues at stake when talking about simulation in science.

I begin (in section 1.2) by looking at the connection between simulation and experiment, a link which may be drawn fairly non-controversially based on commonalities in methodology – manipulation, error-tracking, data analysis, and ‘tinkering’. However, a good portion of the HPS literature links experiment and simulation by materiality, understanding the physical computer as the material instrument that is used to enact simulation. This is used to make simulation an ‘intervention’ on a material instrument, and thus describe simulation (misguidedly) as ‘experiment’. Though simulation is an embodied technique (see section 5.5), this section will maintain that the material similarities argument is useful for neither the epistemology of simulation nor that of experiment. Simulation can, however, be considered epistemologically on par with experiment.

The majority of the simulation definitions in the 1979 article mentioned above highlight models or modelling. The next part of this chapter (section 1.3) looks at the various manners in which the HPS literature links models and simulation, from understanding simulation as ‘model applied’ to considering the model as integrated within and inseparable from the simulation. It is not the intention of this discussion to go into the large body of HPS literature on models; rather, section 1.3 attempts to highlight the complexities tangling simulations and their models. It is not a simple matter to separate the two, a point which will be illustrated with concrete examples from astrophysical simulations of both constructed models and so-called ‘semi-analytic’ models. In general, this thesis will consider the theoretical model as part of the simulation (and modelling as part of simulating).

That simulation can be both an ‘experiment’ and depend heavily on models is what initially made simulation so interesting to philosophers of science. The final part of this chapter (section 1.4)

---

5 Pritsker, 1979.
will discuss the idea that simulation has something of a ‘dual methodology’ that involves components of both theory and experiment. Though this interpretation – whereby simulation consists of both theoretical and experimental aspects – is preferred in this thesis, it remains difficult to locate simulation on the methodological map. Another complicating factor is the question of philosophical novelty – does simulation represent a new way of doing science, or is it just a compound of old ways? – which is further clouded by the differences in simulation practice across scientific disciplines, and the loaded terms with which we describe simulation (like ‘theory’ and ‘experiment’).

1.2 Materiality and Experiment

There are several aspects of simulation that ‘look like’ experiment. For example, Deborah Dowling describes how scientists ‘tinker’ with their simulation, making the practice of simulations a skilled craft. She also highlights the similarity between error tracking and troubleshooting (debugging) in traditional experiments and computer simulations. The manipulation and intervention with a ‘target’, the preoccupation with uncertainty and error-tracking, the production of data-like results, and the familiarity simulators develop with their system through ‘tinkering’ and exploratory play, have all been associated with the practice of experiment in HPS literature. The degree to which we may take this similarity is, however, disputed. Are simulations epistemologically like experiment? If so, how and to what extent? What effect does such similarity have on the validation of the knowledge simulations produce?

A key facet in the discussion around these questions is the materiality of simulation, experiment, and the epistemic target. To begin with, one might ask: is simulation material, or non-material? Answering this question is important for this thesis because astrophysicists cannot, in the ordinary sense, experiment with or materially intervene on their targets of investigation. Does simulation fill this role instead? The answer, I think, is no. Simulations are not material, and are not epistemologically the same as experiment, but they are on par with experiment when it comes to producing legitimate knowledge. Let us, however, come to this conclusion by considering what others have said on the subject.

Though the interpretation of simulation as non-material seems more intuitive (it is, after all, a virtual technique), for some scholars the fact that a simulation is instantiated on a physical, material, machine is a crucial point, especially for describing its similarity to experiment. Wendy Parker, for example, distinguishes between computer simulations (a type of representation consisting of a time-ordered sequence of states) and a computer simulation study as a mode of practice, which is the broader activity of setting up the simulation, running it, and collecting data.

---

8 In a sense, simulation is both – simulation is built using virtual code based on theoretical models, but the act of simulating is embodied and a material tool (the computer) is required. We will return to this later (see sections 2.4 and 5.5).
It is the computer simulation study that is an experiment, in which ‘experiment’ is an investigative activity involving intervention, and the system intervened on is a (material) programmed digital computer. Stephen Norton and Frederick Suppe describe simulation as the real world incarnation of some base model, physically interacted with through the computer. These authors therefore consider the simulation a material experiment, and thus claim there are “no fundamental epistemological differences between traditional experimentation and today’s computationally intense observations”. While few make such a bold statement about the relationship between experiment and simulation, Thomas Weissert suggests that if mathematics is included in the realm of nature, then simulation can be thought of as a subset of experiment: “simulation is an experiment into an unknown area of mathematical structure using a physical apparatus.” The experimental apparatus is the computational machine and the graphical display device, though by contrast to experiment, very little theory of the apparatus comes into play when using the knowledge from a simulation. Weissert explicitly wishes to derive a philosophy of simulation from the existent philosophy of experiment, though he still considers the philosophy of simulation to be separate to that of experimentation.

What the frameworks of W. Parker, Norton and Suppe, and Weissert do, is make simulation material through the use of the physical computer. Since, in these contexts, ‘experiment’ is broadly understood to be an intervention on something material, if simulation is cast as an intervention on a material system (the computer), then simulation can be understood to be the same as, or a subset of, experiment. If we can demonstrate, then, that the target system and the apparatus are connected in simulation in the same way they are connected in experiment, then simulation is basically a form of experimentation. Furthermore, it is through material similarities that we intuitively expect experiments to produce valid results – because the experiment is made of the same ‘stuff’ as its target, because we are intervening on nature itself, experiments do not need to justify the veridicality of their representation, unlike simulation. Linking simulation to experiment is therefore also an attempt to justify simulation knowledge in the same way experimental knowledge can be justified: by demonstrating that the simulation (and its model) is similar, in some sense, to its target system – what Matthew Parker calls the ‘universality’ principle.

Others have, however, pointed out a flaw in this argument. Francesco Guala writes:

The difference [between experiment and simulation] lies in the kind of relationship existing between, on the one hand, an experimental and its target system, and, on the other, a simulating and its target system. In the former case, the correspondence holds at a ‘deep’, ‘material’ level, whereas in the latter the similarity is admittedly only ‘abstract’ and ‘formal’.

---

9 W. S. Parker, 2009.
10 Norton & Suppe, 2001, p. 86.
11 Weissert, 1997, p. 112.
12 M. W. Parker, 2009.
Let us ignore, for a moment, the question of just how these correspondences can be drawn in ordinary experiments, and accept that experiment works because there is a material similarity between the experiment and the actual target of investigation. What is at stake, then, for applying this argument to simulation is not that the apparatus of simulation is material. It is the correspondence relation between the experiment and the target or phenomena under investigation that is of importance – Guala argues that simulation is fundamentally, ontologically different from experiment. While “both experiments and simulations are knowledge-producing devices ... the knowledge needed to run a good simulation is not quite the same as the one needed to run a good experiment”.\textsuperscript{14} A higher level of theoretical understanding must be in place before one can build and run a simulation, while (traditional) experiments can rely on the fact that they are made of the same ‘stuff’ as the real world. Similarly, Hughes claims that in physical simulations (where one physical system is a substitute for a different physical system) there is “an obvious correspondence between the processes at work”\textsuperscript{15} that is not present between a computer simulation and its target system. Claiming that simulation constitutes an intervention on a material system, and therefore applying the material similarities argument to simulation, is not going to work on any meaningful level because that material system – the computer – is not the epistemic target.

While the physical computer is undeniably material, the arguments that translate this materiality into simulation and make it epistemologically meaningful are unconvincing. The material similarities argument can, however, be turned on its head, and still be used to argue for simulation as experiment. Some scholars have argued for the similarity between simulation and experiment not by making simulation material, but by making experiment non-material. W. Parker writes that, “when it comes to justifying inferences about target systems, [what matters] is not materiality, but relevant similarity”.\textsuperscript{16} These similarities may be either formal or material, but, “crucially, having experimental and target systems made of the same materials do not guarantee that all of the relevant similarities obtain”.\textsuperscript{17} In fact, drawing similarities between experiment and target systems (the real world systems under investigation) is not a simple task. For physical simulations – the use of one physical system to simulate another physical system – it may well be that the very laws of the analogue differ from those of the target system.\textsuperscript{18} Susan Sterrett points out that even in the case of traditional experiments, it is no simple task to identify which similarities between systems are relevant. That task involves a significant amount of predominantly theoretical work, the result of which is that the method of comparing the two systems will have become formalised, usually into some sort of numerical dimensionless parameter.\textsuperscript{19} This of course means that simulation, as long as it demonstrates similarity with the

\textsuperscript{14} Guala, 2002, p. 70. Emphasis in original.
\textsuperscript{15} Hughes, 1999, p. 139.
\textsuperscript{16} W. S. Parker, 2009, p. 493.
\textsuperscript{17} W. S. Parker, 2009, p. 493.
\textsuperscript{18} Mattingly & Warwick, 2009.
\textsuperscript{19} Sterrett, 2009, p. 10. The relevant parameters are dimensionless because they take the form of ratios.
appropriate dimensionless parameter, is as much an experiment as any other.

Margaret Morrison is broadly in agreement with Sterrett, though her unit of choice is the model. Morrison argues that “in many cases of traditional experiment the causal connection to the target is not what is doing the epistemological work”. In the case of the recently reported detection of the Higgs boson, it can be argued that materiality is neither a sufficient nor necessary criterion for experiment, as theory and models played a crucial role in making ‘the discovery’ intelligible as such. Modelling assumptions (and simulations) were necessary not only for the construction of the experiment, but for the analysis of the data, especially in the more complicated modern experiments: “in order to have any understanding of what is happening in these experiments, one requires highly sophisticated models that are themselves sometimes hypothetical.” The upshot of Morrison’s argument is that the inference from experiment to target system looks very similar to the inference from model to target system (she makes an argument for models as ‘measuring instruments’). And the inference from simulation model, to model, to target system, is also no different: the association of the simulation particles with physical properties via the mathematised model functions in the same manner as the association of physical properties with experimental phenomena via the theoretical framework. Once again, simulation is a type of experiment.

This approach to studying simulation (in simulation-experiment systems) is a good way to get new perspectives on the epistemology of experiment. However, tying simulation so closely to experiment, while also attempting to avoid materialising simulation, has the effect of de-materialising experiment; or at least, making the materiality of experiment as incidental as the materiality (from the physical computer) of simulation. Following Morrison, the method of identifying material similarities can be the same in experiment as it is in simulation since we don’t actually rely on the precise materiality of either system. Simulation does not need to be material to associate itself with physical properties, but neither does experiment. Perhaps this is not a problem. However, understanding simulation in this manner – tied to experiment by either material or non-material similarities – necessarily requires that one interprets simulation through the epistemology of experiment, which is itself being simultaneously reworked as simulation introduces new facets to the philosophy. No doubt this is a fascinating intersection, but it also obscures the roles simulation plays as a distinct technique. In fact, one of the motivating reasons for drawing case studies from astrophysics was to develop an understanding of how simulations can generate knowledge in the absence of experiments.

Morrison, in fact, is careful to emphasise that:

I am not arguing for the experimental status of simulation tout court; instead I want to push the more modest proposal that in some cases and with specific types of simulation techniques we can be reasonably justified in claiming that simulation outputs can function

---

20 Morrison, 2015, p. 242. ‘Causal connection’ can be read as material similarity.
21 See also Massimi & Bhimji, 2015.
22 Morrison, 2015, p. 235.
as measurement. In some sense it really doesn’t matter whether we can characterise a simulation as an experiment; what matters is that their outputs can enjoy the same epistemic status when suitable conditions are met.\textsuperscript{23}

This thesis will make similar arguments (see especially Chapter 4) with regard to simulation output functioning as data, and the epistemic status of simulation. However, I think it is possible to make those claims without pushing either the experimental status of simulation (tout court or elsewise) or developing an accidentally dematerialised understanding of experiment. In making their arguments for simulation-as-experiment, authors like Morrison are replying to the common perception that experiment has epistemic primacy over simulation. They are arguing, correctly I feel, that simulation can be epistemologically on par with experiment. Yet it does not necessarily follow that in order for simulation to produce legitimate knowledge of a target system, it must somehow be experiment.

It seems obvious that some external evidence is required to make simulation anything more than theoretical, to make it an ‘experiment’. One of the major stumbling blocks in attempting to make the simulation-experiment connection is the greater epistemological power experiments are granted through what Hans-Jörg Rheinberger has referred to as the recalcitrance of an experimental system. It is because of the resistance granted by physicality that experimenters “are not engaged in platonistic exercises, in copy theory-guided asymptotic approximations to reality, or in bluntly social constructivist endeavours,”\textsuperscript{24} while simulation, on the other hand, has often been accused of just that. Ronald Giere makes this explicit in a reply to Morrison: “the epistemological payoff of a traditional experiment, because of the causal connection with the target system, is greater (or less) confidence in the fit between a model and a target system. A computer experiment, which does not go beyond the simulation system, has no such payoff.”\textsuperscript{25}

While a model is often used to correct experimental measurements, it is nevertheless the material apparatus, not the model, that is interacting causally with a target to produce the results. Much more work has to go into justifying the results of a simulation with respect to its target of investigation than that of your classical experiment. That Morrison can show a simulation to be producing ‘measurements’ is, I think, an outcome of both the markedly different nature of modern experiments to traditional ones, but more importantly of her simulation’s \textit{interaction} with the experiment. A similar outcome is seen in Chapter 4. A material interaction with the target system of some form or another is necessary for scientific knowledge about that system to be produced. Giere points out that “computer experiments not connected to actual data do not confer any additional confidence in the fit (or lack thereof) between the simulation model and the target system.”\textsuperscript{26} Experiment and observation have this connection – they produce empirical data – and simulation, by itself, does not.

\textsuperscript{23} Morrison, 2015, p. 247.
\textsuperscript{24} Rheinberger, 1997, p. 225. See also Rheinberger, 2005b.
\textsuperscript{25} Giere, 2009, p. 61.
\textsuperscript{26} Giere, 2009, p. 62.
Mary Morgan suggests that while ‘experimenting’ on mathematical models we may be surprised, only in material experiments may we be *confounded.*\(^27\) Morgan means this ontologically, but also in terms of knowing what materials go into the experiment. When your materials are “pieces of mathematics”,\(^28\) you know what goes into the model and while you may experience unexpected model behaviour, you will never encounter, as you do with experiments, a physical law that has no theoretical underpinning. In this sense material experiments have a greater epistemic power than ‘model experiments’ (which include simulations). Manfred Stöckler similarly states that simulations, “cannot supplant the semantic and epistemological role of experiments [being] arguments, not experiences”.\(^29\) These authors are not denying the importance of simulation, but highlighting the crucial, material difference that experiment has over computer models. Simulation is not material, and cannot stand as material intervention. Morrison’s Higgs boson simulations are, of course, connected to actual data; they are embedded in the experiments that the simulations help build and analyse – and this connection is crucial. In addition, the materiality of experiment – that is, the *perception* that experiment is material – is also important for understanding the epistemology of simulation, particularly where it intersects with actual data.

This does not mean, however, that because simulations require external evidence for comparison they are epistemologically inferior; no more than it implies modern experiments are epistemologically weaker because their results need to be interpreted by a range of models. What the literature seems to under-emphasise is the complexity of the negotiation between simulation and experimental (or in the case of astrophysics, observational) data. There is a back-and-forth process, during which simulation takes on a number of roles both experimental and theoretical, and can end up producing epistemically legitimate ‘measurements’ or ‘data’. Eric Winsberg agrees that while experiments can be considered to be epistemically *prior* to simulations, in that the theoretical knowledge required to develop a simulation is quite sophisticated, simulations can still produce knowledge that is on par with experimental results. Rather than relying on similarities, material or otherwise, we should instead focus on methods of justification:

> If we want to characterize the difference between an experiment and a simulation, we should not focus on what objective relationship actually exists between the object of an investigation and its target, nor even on what objective relationship is being aimed for. We should focus instead on epistemological features – on how researchers *justify* their belief that the object can stand in for the target.\(^30\)

Winsberg describes the difference in that simulationists argue for the external validity of their simulations based on the belief that they possess reliable principles for building models of the targets of their investigations, whereas experimentalists usually argue for the internal validity of their experiment by demonstrating control of their (material) object. For experiment, then,

---

\(^{27}\) Morgan, 2003; Morgan, 2005.  
\(^{28}\) Morgan, 2005, p. 324. The examples Morgan deals with are economic, rather than scientific, but are broadly applicable.  
\(^{29}\) Stöckler, 2000, p. 369.  
\(^{30}\) Winsberg, 2010, p. 63. See also Winsberg, 2009b.
sometimes the materiality of the experimental system is very important for justification; for simulation, however, it never is.

We cannot argue for the validity of simulation knowledge based on material similarity, and yet all these scholars (and the scientists!) seem to agree that simulation knowledge can be on par with experimental knowledge. There have been several attempts to legitimize this connection by making simulation, somehow, identical to or like experiment. Turning simulation material seems to miss the point, but turning experiment non-material is somewhat problematic. Winsberg's approach seems more balanced since it focuses on epistemological aspects – how simulation knowledge is justified by the researchers themselves to be about the target. Morrison is in fact right when she says it does not matter whether we can characterise simulation as experiment – what matters, and what will be important in the case studies of this thesis, is how certain simulations come to be on par with experiment, and also how certain simulations don’t.

Before moving on, it is necessary to mention the elephant in the room that looms behind a good deal of the philosophy of simulation, and that is the issue of representation. Discussions of relevant or formal similarities raise questions about how well the simulation represents the target system, as does the accuracy and complexity of the models the simulations use and produce. Many definitions of simulation refer to its mimetic or representational capacity, sometimes with sinister overtones. It is in fact rather impossible to talk about simulation (or for that matter, models and theories) without talking about representation in some capacity. The astrophysical simulations discussed in this thesis generally aim to represent some phenomenon or object, in order to learn more about it. On another level, one of the recurring themes of this thesis the fear of seduction, which is centrally concerned about the confusion between representation and reality. Sections in Chapters 3 and 4 discuss the role of the pictures generated by simulation, as visual representations. However, these discussions limit themselves to the intention of representation: it is sufficient, for our purposes (and for our selection of simulations), to say that when simulations model a real world system they are intended to count as representations of that system. Questions on a semantic or metaphysical level of whether such simulations actually do represent their targets, how faithfully they do so, or how these representations can be mapped, will not be dealt with in this thesis.

1.3 Models and Simulation
Rather than attempting to associate simulation with experiment on a deep epistemological level, several scholars have instead described simulation as something like an experiment on a theoretical model. Models are put to wide and varied use in science, as the large body of HPS literature on models can attes to, but it is not often that a theoretical model itself can be implemented dynamically and explored temporally and spatially (within the computer).

Simulation is an excellent way of making a model ‘observable’, as Dowling notes in her sociological study:

The physical instantiation of the software on an electronic machine is important, partly because it creates a process, and partly because it turns an abstract model into a physical system that can be ‘observed’.32

Here the materiality of the simulation, in the form of the physical computer, is recalled because simulation is as much an embodied activity for the scientist as calculating the results of an equation using a pen and paper – it occurs in the real world, and requires interaction with real objects: “the action of running the software on a machine is particularly crucial. In instantiating the code, a process occurs which produces an output.”33 Putting this form of materiality aside for now (see sections 2.4 and 5.5), Dowling highlights an important thing simulation allows us to do with models: make them dynamic.

Understanding simulation as a dynamic process refers to two things: first, simulating itself is a temporal act that has multiple stages (often feeding back into one another) including construction, running, and processing. The act of simulating takes time. Second, simulation is highly suited to studying systems that are themselves dynamic or change over time; a good deal of astrophysical simulations are concerned with the time evolution of various systems in order to determine what sorts of conditions or processes are occurring to make what we see today. Stephan Hartmann likewise emphasises process:

A simulation imitates one process by another process. In this definition the term ‘process’ refers solely to some object or system whose state changes in time.34

Hartmann, in fact, suggests that “scientists reserve the term ‘simulation’ exclusively for the exploration of dynamic models”.35 Paul Humphreys emphasises process similarly, as well as including the other meaning of dynamic:

System S provides a core simulation of an object or process B just in case S is a concrete computational device that produces, via a temporal process, solutions to a computational model … that correctly represents B, either dynamically or statically. If in addition the computational model used by S correctly represents the structure of the real system R, then S provides a core simulation of system R with respect to B.36

This core simulation becomes a full simulation if successive solutions are computed and arranged in an order that correctly represents the states of the system being modelled. Thus a core simulation can express a static system, and a full simulation a dynamic one, while both types of

---

35 Hartmann 1996, 84.
36 Humphreys, 2004, p. 110.
simulations are dynamic in the sense that the computer produces its solutions via a temporal process.

The perception that simulation makes theory or models dynamic is closely tied to understanding simulation as ‘experimenting’ on models. If a model provides certain laws or rules about what state a system is at a certain time \( t \), the simulation can actually show that system at every time \( t \). The properties of the model can be dynamically explored in simulation, parameters changed, multiple runs conducted of various system states, the model wound forwards or backwards in time, and generally be ‘experimented’ upon. Dowling found that this sort of dynamism is actually used to separate ‘the model’ from ‘the simulation’ in actor’s categories – simulation is dynamic, both in terms of the system its describing and the temporal process of exploring that system, while a model is static.\(^{37}\) The relationship between the simulation and the model is clearly philosophically complex and – like the relationship between simulation and experiment – variously defined. The question this portion of the chapter then asks is: (relevant to astrophysics) can simulation and model even be sensibly separated?

1.3.1 The Simulation Model

Understanding simulation as an ‘experiment’ on a theory or model should not imply the naïve view that a theory or model can be implemented into a simulation without undergoing some sort of (fairly significant) transformation. Many scholars differentiate between the theoretical model and the simulation model – the latter can be broadly understood as the theoretical model programmed into the simulation. Several early definitions of simulation echo this understanding of theoretical model translated into a computer (and made dynamic); for example: “a computer simulation is the mechanisation of a mathematical model in a manner designed for experimentation”.\(^{38}\)

Axel Gelfert distinguishes among different sorts of model-in-the-simulation, describing his own version of ‘mathematical models’ and ‘computational models’. The latter is literally a model that computes – a number cruncher – while the former is what involves theoretical elements. Simulations seem to involve both types of models; ‘simulating a model’ is "running a computer simulation of the dynamic behaviour of a model given a certain input, e.g. consisting in parameter values and boundary conditions".\(^{39}\) The computational model, “often a crucial step in the actual implementation of computer simulation”,\(^{40}\) is there to turn the mathematics into something the computer can solve. Gelfert is, rightly or wrongly,\(^{41}\) distinguishing between the theoretical model


\(^{38}\) Rahe, 1972, p. 13.

\(^{39}\) Gelfert, 2009, p. 498.

\(^{40}\) Gelfert, 2009, p. 504.

\(^{41}\) Humphreys makes the point that mathematically formulated theories often come with an interpretation that guides or justifies their employment, and so it seems naïve to settle for an account that separates interpretation from formalism (Humphreys, 1995/96).
mathematised, and the numerical techniques (the ‘computational’ model) used to solve the mathematical model in the simulation.

Winsberg creates even more divisions and describes the integration of a theoretical model into a simulation through a hierarchical relationship: first we start with a mechanical model – a bare bones characterisation of a physical system usually accompanied by a family of equations. To this we add parameters, boundary conditions and initial conditions to restrict the model to a specific class of data, and gain a dynamic model. Adding simplifying assumptions (either eliminative or creative) creates an ad hoc model, and programming this into the computer creates a computational model. This last model gives results in the form of a data set, which is then analysed and produces a final model of the phenomena. Thus the computer simulation develops “a complex chain of inferences that serve to transform theoretical structures into specific concrete knowledge of physical systems”.\textsuperscript{42} The computational model is then comprised of the basic theoretical system, parameter and boundary conditions, simplifying assumptions, and the computer code.

The process of getting the theoretical model into the simulation is certainly not a simple one. Even converting continuous mathematical functions into something discrete which can be solved by a computer creates differences. Morrison describes:

> The simulation model is the result of applying a particular kind of discretisation to the theoretical/mathematical model. ... What I call the ‘simulation system’ consists of the computer, the simulation model, and the programme [that] allows us to follow and investigate the evolution of the model physical system (i.e., the theoretical model/representation of the target system).\textsuperscript{43}

Morrison describes the discretisation process as “a bottom-up framework that begins with the phenomenon or target system and a theoretical/mathematical model of that system”\textsuperscript{44} (but it may not always be the case that there is a model of the target system; see next two sections). The process of translating or converting the model into the simulation first uses approximation methods to discretise the model and then develops an algorithm that the computer uses to solve the relevant equations.

Morrison points out that one of the key features in developing this algorithm is the computational costs involved. A recurring problem astrophysicists will face in the various examples and case studies is the compromise between resolution and scale. If a simulation is large-scale, of a size representative of the universe, it is a ‘full-box’ simulation, and will generally “resolve the entire computational domain with a single particle mass and force resolution”.\textsuperscript{45} In terms of physics, these are cosmic scale simulations; the Millennium and Millennium-II simulations fall under this

\textsuperscript{42} Winsberg, 1999b, p. 275. See also Winsberg, 1999a.
\textsuperscript{43} Morrison, 2015, p. 219.
\textsuperscript{44} Morrison, 2015, p. 219.
\textsuperscript{45} Kuhlen, Vogelsberger, & Angulo, 2012, p. 66.
heading as they aim to follow the evolution of dark matter structures over volumes, “comparable to the largest current surveys.”\textsuperscript{46} If, on the other hand, a simulation favours resolution, it is a ‘zoom-in’ simulation. A smaller volume covered means higher resolution, which typically translates to more particles of smaller mass each, though there is another compromise to be made here. Examples of these are cluster scale simulations, such as Phoenix,\textsuperscript{47} which simulate the dark matter components of galaxy clusters; and galactic scale simulations such as Via Lactea and Via Lactea II,\textsuperscript{48} and Aquarius,\textsuperscript{49} which focus on the simulation of a galaxy-sized halo, usually of approximately the size of the Milky Way. The more detail you want in your simulation, the smaller your volume needs to be, and vice versa. Making these sorts of compromises involves discarding or idealising bits of the model system; for example, many of the more complicated and smaller scale physical processes in a galactic system (like star formation and destruction, supernova feedback, etc.) are either reduced to a single parameter or determined to be negligible. The simulation will not typically have the processing power to resolve these small scale processes, or model them fully. In many respects, the computational model is an entirely different model of the target system, though it is based on theoretical understanding.

Furthermore, simulating is a profoundly non-linear process. A recent article in \textit{Simulation} describes the lifecycle of a simulation as iterative in nature, including activities of problem formation, requirements engineering, conceptual modelling, architecting, design, integration, experimentation and use, presentation, certification, storage, and reuse.\textsuperscript{50} Even after the initial construction of the simulation (which may involve the separate construction, running, testing, and re-writing of smaller parts of the simulation), parts of the simulation may be rebuilt after it is run, the simulation may be altered and rerun after it is analysed, or it may produce partial results that are analysed as part of the build. The reasons for altering the simulation model may be computational, (an error in the code), algorithmic (calculating the wrong results), or theoretical (including discarding, or altering the instantiation of a physical process). Any of these adjustments may result in a change in the ‘simulation model’, though again it is difficult to describe what this model includes. Like Gelfert, we may separate the numerical techniques used to compute the equations, or like Morrison we may see them as defining the simulation model.

I am not trying to develop a robust way of understanding models in simulations, but rather wish to demonstrate the complexities tangling simulations and their models, especially in view of the next two sections. Which bits of a simulation count as the ‘simulation model’? One might say it is instantiated in the code, but the simulation program itself is not a simple beast. Matthew Spencer, in analysing the outcomes of an ethnographic study of an applied computation group, itemises fourteen separate pieces of software to build, run, edit, and analyse simulations using \textit{Fluidity}, a fluid dynamics code.\textsuperscript{51} The compiler, various code libraries, the user interface, the mesh

\textsuperscript{46} Springel et al., 2005, p. 630. For the Millennium simulation see: Springel et al., 2005. For Millennium-II: Boylan-Kolchin, Springel, White, Jenkins, & Lemson, 2009.
\textsuperscript{47} Gao et al., 2012.
\textsuperscript{48} Via Lactea: Diemand, Kuhlen, & Madau, 2007a. Via Lactea II: Diemand et al., 2008.
\textsuperscript{49} Springel et al., 2008.
\textsuperscript{50} Balci, 2012.
\textsuperscript{51} Spencer, 2012, pp. 30-32.
generator, debugging tools, and post-processing software are usually lumped together under the heading 'computer program' in philosophical literature – but there is not one 'program' that runs the simulation (sometimes not even one computer), especially for the larger simulations. The theoretical model cannot be straightforwardly translated into 'the simulation model' in a manner than enables reference back and forth between the two.

1.3.2 Semi-Analytic Models

It may not even be the case that the underlying theoretical model – such as it is – exists independently of the simulation. A good example of this are what astrophysical simulation studies call 'semi-analytic' models. 'Semi-analytic' is an actor's category and is therefore somewhat nebulous and poorly defined; the term generally refers to models that introduce some sort of 'analytical' component into a 'numeric' model. Analytic work refers to pen and paper analyses and the more traditional theorising using established mathematical laws and relationships. Sometimes 'analytic' is also used as a shorthand for the inclusion of simplifying assumptions (based on theoretical knowledge) that make a problem simpler to work out. Numerical work can include both computer assisted calculations ('brute force' solving of unwieldy equations, or other intractable mathematics) and the results of (or pieces of) other simulations.

An influential paper from White and Frenk published in 1991 described a method for studying the formation of galaxies that adopted "the simplest models consistent with our current understanding of N-body work on dissipationless clustering, and of numerical and analytic work on gas evolution and cooling."\(^{52}\) Combining analytic and numerical work through the use of semi-analytic models has, since the early 1990s, become a widespread technique for simulating complex systems in astrophysics (see Figure 1). This is coincident with the rise of the personal computer, but while smaller computers made simulating more accessible to every researcher (see Chapter 2) they also placed computational limitations on the size and scope of the simulations that could be run. Indeed, one of the main motivating reasons for using a semi-analytic model is to circumvent computational deficiency. As a concrete example: Zentner and Bullock used a semi-analytic model to avoid simulating the formation of galaxy structures from scratch. At the time (2003), computational power placed limits on the resolution a simulation could achieve; in order to obtain statistics on small-scale structure, theoretical prescriptions about the behaviour of these objects needed to be included. The authors therefore used a semi-analytic model:

Our goal is to present and apply a semi-analytic model that suffers from no inherent resolution effects and is based on the processes that were observed to govern substructure populations in past N-body simulations. This kind of model can generate statistically significant predictions for a variety of inputs quickly and can be used to guide expectations for the next generation of N-body simulations. Conversely, this model represents in many

\(^{52}\) White & Frenk, 1991, p. 52.
ways an extrapolation of N-body results into unexplored domains, and it is imperative that our results be tested by future numerical studies.\textsuperscript{53}

The model was considered semi-analytic in that it involved theoretical assumptions or hypotheses that were based on previous simulations in order to simplify the current simulation. In covering computational deficiency, a semi-analytic model could provide statistics at the small scale Zentner and Bullock's simulation could not resolve.

![Figure 1: Number of publications in astronomy/astrophysics per year referencing the phrase 'semi-analytic' in the topic field. Data obtained from the Science Citation Index, 23/07/15.](image)

An astrophysical paper in the early 1990s considered that “the greatest strength of the semi-analytic approach is that it permits model and parameter space to be explored quickly and efficiently. It can thus be regarded as a tool for the understanding of the conditions necessary if a hierarchical model is to produce a population of galaxies that can match the great wealth of observations that are presently available.”\textsuperscript{54} At the same time, however, the ‘analytic’ component is often describing physical mechanisms that are poorly understood: “it should be stressed, however, that these predictions are subject to a great deal of uncertainty. The important physical processes that contribute to the formation of stars and galaxies are still the subject of much investigation and debate.”\textsuperscript{55} In these cases the analytic component is ‘parameterised’, whereby an effect of a physical mechanism on the system is approximated with a simple expression, without

\textsuperscript{53} Zentner & Bullock, 2003, p. 50.
rigorously modelling that mechanism in the simulation. The ‘analytic’ or theoretical part of the semi-analytic model is hence not always a ‘model’ as such. Cutting edge hydrodynamic simulations boast the complete absence of semi-analytic models (see Chapter 4) as they are able to fully simulate processes that have since been better understood, while previously “semi-analytic models have resorted to *ad hoc* prescriptions”\textsuperscript{56} because astrophysicists had lacked both the ability and knowledge to directly treat all components in their simulation.

One cannot, in fact, use a semi-analytic model *without* simulating (at least in astrophysics):

> The term ‘semi-analytic’ conveys the notion that while in this approach the physics is parameterised in terms of simple analytic models, following the dark matter merger trees over time can only be carried out numerically. Semi-analytic models are hence best viewed as simplified simulations of the galaxy formation process.\textsuperscript{57}

It may be the case that the particles in a simulation are governed only by simple gravitational laws. However, tracking hundreds of particles as they move with respect to each other is a monumental task that can only be accomplished by the computer. The tracking algorithm allows the scientist to follow (through the computer) the evolution of individual structures; “it is this ability to establish such evolutionary connections that makes this kind of modelling so powerful for interpreting observational data.”\textsuperscript{58} Semi-analytic models are examples of models that are not wholly simulation models, yet which do not exist independently of simulation. They may institute elements from other theoretical models, but they also include the informational data and knowledge generated from simulations – emergent data which cannot be straightforwardly tied to an external theoretical model.

### 1.3.3 Theory Crafting

In fact, it seems to be the case that in situations where there is no underlying coherent model, the model is co-constructed with the simulation, not only during the ‘building’ stage but also as the simulation runs. As a simulation evolves a universe and tracks each structure over the course of the evolution, it dynamically constructs a model of how structures form. This full model of structure formation cannot be found in the underlying theory, which does not robustly predict densities or distributions; it cannot be found in the simulation model, which contains only algorithms for movement and initial conditions; and it cannot be found in some analytic model. Indeed, one of the main motivations for using computer simulation in the first place was to model non-linear and complex systems that could not be done otherwise: "without direct numerical simulation, the hierarchical build-up of structure with its three-dimensional dynamics would be largely inaccessible."\textsuperscript{59}

\textsuperscript{56} Vogelsberger et al., 2014a, p. 178.
\textsuperscript{57} Springel et al., 2005, supplementary information, p. 9.
\textsuperscript{58} Springel et al., 2005, p. 631.
\textsuperscript{59} Springel et al., 2005, supplementary information, p. 1.
Not much focus in the philosophical literature is given to the question on whether ‘modelling’ is part of ‘simulating’. Perhaps this is because, as I have sought to demonstrate, a meaningful distinction between two is elusive. Practitioners also vary in their interpretation, though by and large see part of the simulationists job as modelling (see Chapter 2). For example, John McLeod, the long-time editor of the Simulation journal wrote: ‘my favourite definition of simulation is: ‘The development and use of models of the study of ideas, systems, and situations.’ Ben [Clymer] would not include ‘development.’”60 Clymer, McLeod’s colleague, felt that a simulation could not be performed until a valid mathematical model, complete with numerical values of all constants, became available. Obviously there must be some underlying theory, mathematised, before a simulation can begin, but the sophistication of that theory may be very low indeed.61 Often the numerical values of constants may be adjusted during the simulation so the results match observations more closely – so-called free parameters. We might also ask how well a set of hypotheses, assumptions, mathematical equations, or qualitative statements need to be articulated to qualify as a ‘model’. Within the context of this thesis, that is not a very fruitful question, however. Instead, the understanding of simulation this thesis will be operating with includes processes like model construction, implementation, and alteration (though without denying these processes can occur independently, and without attempting to differentiate where or when that happens). More broadly, since the focus of this thesis is knowledge generation, the process by which information is gathered, modelled, and (re)introduced into a simulation is described as a process of theory crafting. To describe what I mean by this, let us go into the construction of a specific simulation, whose aim is to model a characteristic slice of the Universe.

This is not an in-depth example, but should serve an illustrative purpose.

The Millennium Simulation (MS), so-called because of its size, was a large scale simulation that aimed to inspect the formation of dark matter structures through the evolution of the Universe. Structure formation, that is, formation of galaxies, clusters, and superclusters, is widely considered to be a hierarchical, clustering process: smaller structures form from density perturbations generated very early in the Universe, and these structures then accrete and merge to form the larger structures observed today. The Millennium simulation followed approximately $10^{10}$ particles from very early in the universe ($z=127$) to present day ($z=0$) across a volume of $500h^{-1}$ Mpc (megaparsecs).62 The results of this simulation, carried out by a consortium of European and US astrophysicists, first started to appear in 2005 (there has since been a Millennium-II).63 The whole simulation, performed on an IBM p690 parallel supercomputer, took 28 days to run and produced around 20 terabytes of data.

The authors of Millennium modelled their small slice of the Universe following standard lambda cold dark matter (ΛCDM) prescriptions. The simulation began with weak density fluctuations in an otherwise homogenous and rapidly expanding universe. These fluctuations, determined by a

---

60 McLeod in Clymer, 1966, p. 117.
62 This volume size was chosen because it is large enough to include some rare objects, but small enough that the halos of low luminosity galaxies could still be resolved reasonably well.
63 Boylan-Kolchin et al., 2009.
linear theory power spectrum, were amplified by gravity, eventually turning into the structures we see today through hierarchical growth. Dark matter was assumed to dominate – the effects of baryonic matter were considered negligible – and was modelled using a Boltzmann code based on the assumption that dark matter, being non-interacting, behaves like a collisionless fluid. This was described as “a coarse approximation whose fidelity improves as the number of particles in the simulation increases”.\textsuperscript{64} Cosmological parameters were made to be consistent with WMAP data.\textsuperscript{65} All this means is that the initial framework of the simulation was drawn from an underlying theory or initial model that described how we think structures formed and how matter behaves.

Of course, these aspects were not all there was to the simulation model. The theory needed to be translated into the computer, which for a large simulation like Millennium involved several interacting codes. The exact power spectrum used was determined by a random number seed, with initial conditions given by CMBFAST, a code based on the cosmic microwave background (CMB) anisotropies.\textsuperscript{66} The evolution of particles, modelled as under gravity in an expanding background, was probably the most computationally expensive part of the simulation, requiring a dedicated code, GADGET, to be written.\textsuperscript{67} Apart from these elements, part of the model construction itself was done ‘on the fly’ as “with a simulation of the size of the Millennium Run, any non-trivial analysis step is demanding.”\textsuperscript{68} For example, the mass power spectrum of dark matter had to be ‘measured’ in two parts: long and short range. The calculation of the long range was written into the code, but the short range could not be calculated simultaneously because of the space and time cost. Therefore, the authors of Millennium “carried out a measurement each time a simulation snapshot was generated and saved on disc.”\textsuperscript{69} As the simulation ran, it output time slices of the evolution of the particles, sixty-four in total. Each time one of these slices or snapshots was produced and saved, the short range mass power spectrum of dark matter was calculated at that time.

The studies in the previous section used semi-analytic models to account for phenomena that could not be fully simulated – so too did Millennium. Structure formation involves areas of uncertain physics, like the effect of the interstellar medium on star formation, active galactic nuclei (AGN) feedback effects, and so on. These so-called ‘subgrid’ processes “must be treated by phenomenological models whose form and parameters are adjusted by trial and error as part of the overall data-modelling process”.\textsuperscript{70} Due in part to poor theoretical understanding and in part

\begin{itemize}
\item \textsuperscript{64} Springel et al., 2005, p. 629.
\item \textsuperscript{65} WMAP stands for Wilkinson Microwave Anisotropy Probe, which provided the first full-sky map of microwave background radiation. From it, certain cosmological ‘constants’ have been determined (though continuously refined), including the age of the universe, baryon and dark matter density, and the Hubble constant. These parameters, in combination with hypotheses about the production of particles in the early stages of the Universe and the subsequent formation of structures and domination of various forms of matter and energy, forms the ΛCDM model.
\item \textsuperscript{66} Springel et al., 2005, p. 635.
\item \textsuperscript{67} Springel, Yoshida, & White, 2001.
\item \textsuperscript{68} Springel et al., 2005, supplementary information, p. 6.
\item \textsuperscript{69} Springel et al., 2005, supplementary information, p. 6.
\item \textsuperscript{70} Springel et al., 2005, p. 629.
\end{itemize}
to issues of resolution, Millennium implemented models of subgrid processes phenomenologically, that is, by their expected qualitative effect on structures. This required a certain amount of free parameters that were not wholly determined theoretically, but over the course of the simulation process. The subgrid processes were part of the overall theoretical model of structure formation, but in practice were constructed and adjusted during the simulation process.

In order to provide a practical solution to the two entwined problems of theoretical paucity and processing demands, the authors of Millennium “developed a framework that combines very large computer simulations of structure formation with post-hoc modelling of galaxy-formation physics”.\textsuperscript{71} This framework drew together the various aspects of the simulation described above: Millennium, before the simulation was run, implemented the basic prescriptions of a CDM hierarchical clustering scenario according to standard theoretical understanding. This involved the use of dedicated codes to simulate both the initial conditions and the ‘rules’ the particles behave. During the run, the ‘simulation model’, if one can make such a distinction, was augmented with several free parameters of theoretically-motivated subgrid processes and on the fly calculations of power spectra. The results of the simulation were: several time slices of the structure in the simulated universe, a power spectrum measurement, and friends-of-friends (FOF) group catalogues (these dictate which lumps of particles form ‘galaxies’).

This output, however, was not by itself enough to describe the Universe (or a piece of it). The Millennium authors went on to “test, during postprocessing, many different phenomenological treatments of gas cooling, star formation, AGN growth, feedback, chemical enrichment and so on.”\textsuperscript{72} More theoretical activity occurred in this stage, in checking various treatments of the parameterised subgrid models to see which was the most appropriate. One might also suggest that more construction of the simulation model occurred at this late stage, even though the simulation had already been run. The FOF catalogues produced during the simulation were not sensitive enough to identify substructure – very small scale structures – in the objects in the simulation. A different algorithm was required, applied after-the-fact to the catalogues produced by Millennium, to identify these small scale structures and determine some basic physical properties (angular momentum, rotation curves, velocity dispersion, etc.).\textsuperscript{73} Each larger piece of substructure was then given a mass estimate. Having determined all halos and subhalos at all output times, the masses were then tracked over time to build what is known as a ‘merger tree’: “we determined the hierarchical merging trees that describe in detail how structures build up over cosmic time. These trees are the key information needed to compute physical models for the properties of the associated galaxy population.”\textsuperscript{74}

The merger trees were arguably the final ‘model’ of structure formation the simulation produced, though it should now be apparent that it is not a simple task to disentangle the model from the

\textsuperscript{71} Springel et al., 2005, p. 629.
\textsuperscript{72} Springel et al., 2005, p. 631.
\textsuperscript{73} Springel et al., 2005, supplementary information, p. 7.
\textsuperscript{74} Springel et al., 2005, supplementary information, p. 9.
simulation, even in terms of theoretical knowledge. It also seems misguided to attempt to
distinguish between the activities of ‘modelling’ and ‘simulating’ when constructing a simulation.
Millennium was constructed using an underlying theory of the target system, models of relevant
processes, values calculated during the run, parameters determined based on cumulative
measurements, the outcomes of previous runs, and assumptions and approximations based on
computational constraints. Furthermore, the scientific knowledge that emerged from the
simulation was also the result of extensive post-processing, during which other simulations were
often used to make sense of the data. As Winsberg suggests, the transformation from theoretical
knowledge into simulation, and out again, is a complex process – in fact this complexity is one of
the ways in which simulation can help us generate new knowledge. Such complexity means,
however, that what ‘simulation’ is, is quite difficult to pin down. Winsberg’s hierarchy helps us
understand in a simplified manner the transformations that go into knowledge in order to
simulate it, but it is hard to locate the simulation, or simulation practice, among it. In practice
there seems to be no clear distinction between the ad hoc and computational models. In fact, one
could argue that the act of running the simulation is the act of model construction, in the
Millennium example.

Morrison’s ‘simulation system’ also sits uncomfortably. It is not always possible to have an
underlying theoretical model from which one can construct a simulation model – though this
necessarily depends on your definition of ‘model’. Certainly there needs to be some basic
understanding of your target system to even begin simulating, but often the purpose of the
simulation is to construct a model of the target system – to more expansively describe and
understand the physical mechanisms of the system, which must surely be part of the model of
that system. The merger trees Millennium produced that described the formation and position of
structures in the simulated Universe were not part of the initial simulation model. Johannes
Lenhard agrees, distinguishing between simulation models as type-D (discrete) and theoretical
models as type-C (continuous). In complex systems, like those modelled by simulation, we cannot
derive statements about the approximate properties of D from C, and it is therefore inappropriate
to consider simulation modelling as a realisation of theoretical models.75 Morrison no doubt does
not mean that the underlying theoretical model of phenomena, on which the initial simulation
model is based, is a ‘complete’ model, but her formulation raises unanswered questions about
how developed the initial model needs to be, and still constructs a divide between the ‘model’ and
the ‘simulation’.

I do not think such a division is useful in the context of this thesis. While there can obviously be
(theoretical, mathematical, physical) models external to and independent of simulation, it does
not seem meaningful to speak of ‘the simulation’ as distinct from ‘the model underlying the
simulation’ (I am especially unsympathetic to the view of simulation as ‘model applied’; see next
section). Certain types of models are inseparable from simulation by definition – semi-analytic
models use and are used by simulations for a variety of reasons, and do not exist outside of the
simulation context. Simulation models can transform the knowledge they encode to such an

---

extent that it does not seem unreasonable to consider them models in their own right, independent of their underlying source model (in the cases where such a source exists). It is not even certain that the final simulation model bears any straightforward connection to the original simulation model – Spencer points out that computer modelling is at least two-fold: “the code can be run, but once it is run, the transformation undergone in its execution gives the computation itself an independence from its origin. It is severed from its origin. There is no straightforward way of identifying a computation such as a simulation with the source code.”\(^{76}\)

It is obvious that simulations and models are closely associated; in fact, it is often the heavy reliance of simulation on models that motivates many scholars to suggest simulation is a predominantly theoretical activity (as opposed to an experimental one). However, in this thesis I do not attempt to make any distinction between the simulation model and the simulation, though will often refer to some underlying theory or model. ‘Modelling’ is considered part of ‘simulating’, and model construction is part of simulation construction (and analysis). In looking at computational cardiac electrophysiology, Carusi, Burrage, and Rodriguez emphasise that the simulation is intertwined with the construction of the model: “there is no clear line to be drawn between the models (in the form of equations, parameters, and meshes) and the simulations, since there is feedback between the simulations and the models, as well as feed-forward from the equations to the simulations.”\(^{77}\) The empirical content of the theory, the specifics of the model, and the experimental data are all brought together by simulation, under what Carusi et al call the model-simulation-experiment (MSE) system whereby all the components interact in an embedded and multi-scale manner.\(^{78}\) It is this complex and non-obvious process of blending existing theoretical, empirical, and simulation knowledge that enables the generation of new knowledge. The altering, adjusting, adding to, subtracting from, enhancing, blending, and fitting together of various domains of knowledge – a sort of theory crafting – is one of the ways in which simulation helps scientists make sense of (their explanations for) phenomena. The following chapters, in attempting to tease out the complexities of this process, do not do so in order to identify them with ‘modelling’ (or ‘experimenting’); rather, this thesis aims to understand how the activity and production of simulation furthers our scientific knowledge without making decisions about the structures in which that knowledge resides.

\(^{76}\) Spencer, 2012, p. 156.  
\(^{78}\) In astrophysics the experimental component is replaced by observation. This is not a trivial replacement – as Chapter 4 briefly discusses, there is something to suggest that the role of simulation in non-experimental sciences (like astrophysics and climate science) may have important differences than when there is a simulation-experiment relationship (like in particle physics, or nanotechnology).
1.4 Theory and Experiment, and Simulation

Simulation seems to bear important philosophical ties to both experiment and theory.\(^{79}\) It is this apparent dual methodology that attracted HPS scholars to simulation in the first place; as Fritz Rohrlich put it early on: “computer simulation provides ... a qualitatively new and different methodology for the physical sciences, and ... this methodology lies somewhere intermediate between traditional theoretical physical science and its empirical methods of experimentation and observation”\(^{80}\). Where simulations place on this scale of theory and experiment is a matter for debate. In important ways, as the two previous sections showed, simulations are neither purely experiments nor purely models. Are simulations then a new way of doing science, a new combination of old ways of doing science, or simply reducible to the same old philosophical issues? Why does simulation look like both theory and experiment, and how do these aspects relate to each other in practice (and in epistemology)?

It is hard to avoid understanding simulation as a blend of experimental and theoretical methodologies and epistemologies. Indeed, this thesis broadly understands simulation as an experiment on (and with) a model, but in a more robust and integrated manner than is usually found in the literature and without removing the autonomy of simulation as its own distinct technique. Several elements of my position can be found in the HPS literature, particularly if one begins with the idea of the simulation as mediator. Peter Galison was among the first to explore these ideas in depth with his analysis of simulations using the Monte Carlo methods – a set of mathematical techniques used to resolve problems of a random or probabilistic nature.\(^{81}\) Galison describes simulations as theoretical because they are scale-free, position-free, and symbolic, based, as they are, on theoretical models. At the same time, simulations are experimental because they have locality, replicability, and stability. Galison sees simulations as located methodologically between experiment and theory, but also as irreplaceable external

---

\(^{79}\) Though this chapter does not go into it, there is also a portion of literature that deals with the similarities between thought experiments (TEs) and simulations. Given the long and involved (and contradictory) bibliography of TEs in HPS, I have elected not to complicate the simulation issue here, but it is worth mentioning in footnote form. The association between TEs and simulations is suggested because one may ask the same question of both techniques: how can a TE/simulation produce new information, if all that is fed into it is old information? Similar to the rest of the simulation literature, one may find a variety of ways in which TEs are associated with simulations. Beisbart and Norton argue that essentially, Monte Carlo simulations are nothing more than elaborate arguments (Beisbart & Norton, 2012), and that since performing a thought experiment is the execution of an argument (Beisbart, 2012), simulations are elaborate TEs. They note that there is no requirement that the argument be humanly comprehensible. Di Paolo, Noble and Bullock, more moderately, suggest simulations are like emergent, computational thought experiments as they allow us to employ similar tactics from TEs but for the much more complex systems of modern science. The price we pay for this is explanatory opacity, as the behaviour of the simulation cannot be understood by simple inspection (Di Paolo, Noble, & Bullock, 2000). Moving further down this track, Chandrasekharan, Nersessian, and Subramanian suggest that computational modelling has replaced thought experiments because it affords deeper insights to problems and a more sophisticated form of the model-based reasoning that TEs use (Chandrasekharan, Nersessian, & Subramanian, 2012). No doubt there are more positions along this scale, but to adequately engage in this conversation would require a detailed treatment of thought experiments as well as simulations, which suggests a different thesis.

\(^{80}\) Rohrlich, 1990, p. 507.

\(^{81}\) Galison, 1996; Galison, 1997.
intermediaries; simulations have a dual role as marginal and necessary. The Monte Carlo simulations are an example of a ‘trading zone’ in which the techniques are partly divorced from any particular signification. Simulations form a sort of ‘pidgin’ that allows theorists and experimenters to communicate.

This idea is similar to the ‘models as mediators’ view where (non-simulation) models are understood to function as mediators between theory and data. Sergio Sismondo considers models as connecting the ideal to the material; to do this the models must “simultaneously look like theory – because they have to explain, predict, and give structure – and like practical knowledge – because they have to connect to real-world features.”82 Importantly, to be able to mediate, the models must be autonomous, that is, partially independent of both theories and the world. Morrison and Morgan are careful to emphasise that models are not between theory and data, but rather outside the framework, enabling them to include both theoretical and non-theoretical elements – they have both a partial dependence and a partial independence from theories and the world.83 Embedded in the mathematical structures of models is a good deal of information that allows us to understand not only the physical laws, but also why the laws function as successful predictive devices. For Morrison, in cases where the phenomena are inaccessible (like galactic phenomena): “the model occupies centre stage and takes on the task of representing how we assume the system is constructed. And, because direct comparison between the model and the system is not possible, the model supplants the system as the object under investigation.”84 Yet this thesis would argue that it is only possible for a ‘model’ to supplant the system under investigation through simulation. The model, by itself, is not dynamic enough to function as an epistemic target – and in this case, ‘model’ is not synonymous with ‘simulation’. The ‘models as mediators’ view does allow for this, with Morrison and Morgan describing simulation as the means by which models can function as instruments for experimenting on a theory. Simulations form a bridge from the abstract model with idealised facts (theory) to the technological context with concrete fact (experiment): “the model represents systems via simulations, simulations that are possible because of the model’s ability to function as the initial instrument of their production”.85

Rather than this view, the case studies of this thesis seek to demonstrate that the simulation technique grants models extended representative and epistemological capabilities that are a result of properties of the simulation (rather than the underlying model).86 Yet there is definitely something to the idea that simulations help empirical data connect with what Winsberg calls the empirical content of a theory – the solutions to the original equations of the theoretically

82 Sismondo, 1999, p. 258.
86 ‘Simulation technique’ is used as shorthand for the gamut of activities that fall under the heading of ‘simulating’, including the non-linear and iterative processes of building, running, and processing results from a simulation. The phrase also refers to an understanding of simulation as a non-specific technology that can be applied to many different problems in many different areas of science.
motivated model. Winsberg understands the empirical content of a theory in a limited fashion, whereby the simulation reveals the non-obvious results of the model, but with the understanding that this information was implicit in the theory from the start. He urges, rightly I feel, that is possible for simulation to produce knowledge outside of the scope of the theory on which it is based. The data that a simulation produces is not just the working out of the mathematical equations of the underlying theory. As Gelfert puts it: "simulation is essentially about the generation of new ‘data-like’ material – that is, of simulated data that were not antecedently available, neither via empirical observation nor via theoretical deviation.”

This simulated data is not experimental, because it is not empirical, but it was produced with something akin to experimental methodology. The data is also not purely theoretical, because in most complex simulations the results cannot be straightforwardly associated with solutions to equations or other derivations. Galison, for example, suggests that (Monte Carlo) simulations function as experiments because they are stochastic, and are thus more accurate representations of the world since nature is, at its root, statistical. These sorts of statistics cannot be wholly derived from an underlying theory. Similarly, Rohrlich describes simulations as instigating the development of a new syntax that is phenomenological rather than fundamental, being derived from approximate mathematics rather than rigid deterministic logic. It is this phenomenological component, Rohrlich contends, that allows simulations to be experiments on models. Part of the simulated data is the empirical content of the underlying theory, but it also includes measurements of stochastic processes, visualisations and morphologies of phenomena, and best fit values for free parameters, all activities commonly found in astrophysical simulations. In many respects, the results of a simulation can be understood as empirical data in its own right, observations and measurements of the particular representation being simulated (see Chapter 4).

How simulations are able to produce such data, and use it to generate new knowledge, is the central concern of this thesis. Part of the answer to this question certainly lies in how simulation is used by scientists as a technique to construct incredibly complex systems from a variety of simpler theoretical building blocks. Dowling’s framework is excellent for understanding this process. In her study, she understands the computer model as temporarily ‘black-boxed’, in which the interior workings and the algorithms of the code are momentarily put aside. This is possible because of the complexity of the systems typically modelled using simulations, which makes it difficult for one user to understand the system as a whole. Black-boxing enables the user to interact with the simulation as an opaque, unpredictable entity, which gives the simulation system some of the confounding attributes that are found in a traditional material experiment.

---

90 Rohrlich, 1990.
Crucially, however, this black-boxing is only temporary, as to analyse and justify the knowledge gained the simulation technology must be seen as a transparent calculating machine.

For Dowling, "the usefulness of computer simulation thus depends on the construction and maintenance of this [theoretical and experimental] methodological ambiguity".92 This crucial point will also become apparent in the case studies in this thesis. In Chapters 3 and 4, the flexibility with which simulation can switch between being ‘theory’ and being ‘experiment’ is useful for both argumentative and practical purposes, and also functions fruitfully on an epistemological level. Martina Merz notes that because of such methodological ambiguity, a simulation ‘object’ – her example are the event generators in particle physics – "keeps oscillating within a space that is delimited by its physics content (‘epistemic object’) and its black-box features (‘technical object’)".93 Simulations allow the oscillation, or negotiation, or mediation, between empirical data and theoretical data, and from this interplay emerges new knowledge about the physical mechanisms, behaviour, and structure of the phenomena under study.

1.5 Philosophical Novelty

Given that simulations do not fit comfortably into our existing categories, some scholars have proposed that simulation constitutes a new way of doing science that cannot be reduced to the epistemologies of either theory (or modelling) or experiment. Humphreys is one example, pushing for a computational empiricism, in which "our slogans will be mathematics, not logic; computation, not representation; machines, not mentation".94 Winsberg suggests that simulation in fact requires its own entirely new epistemology, which should be "the study of the means by which we sanction belief in the outcome of simulation studies despite their motley methodological structure".95 It is both a mathematico-logical and an empirical epistemology, involving descriptions of how we justify simulation knowledge. It also is necessarily a motely philosophy as ultimately theoretical knowledge is just one of the ‘ingredients’ simulations use to produce results.96 Many of Winsberg’s desired aspects of a simulation epistemology will appear in the following chapters.

Not everyone is convinced that simulations constitute new ways of doing science, however. Roman Frigg and Julian Reiss argue that the issues raised by the computer simulation literature are, metaphysically, epistemically, semantically, and methodologically, the ‘same old stew’.97 Humphreys has issued a direct rejoinder, arguing that we do need a new epistemology for simulation that describes how simulation models relate to phenomena, and that simulation does represent a sui generis activity that is between theory and experiment.98 Winsberg, likewise and

---

93 Merz, 1999, p. 314. The terms Merz places in brackets are taken from Rheinberger, 1997.
94 Humphreys, 2004, p. 53.
95 Winsberg, 1999a, p. 262. See also Winsberg, 1999b and Winsberg, 2010.
97 Frigg & Reiss, 2009.
98 Humphreys, 2009.
unsurprisingly, also holds that simulation needs its own epistemology. Yet Frigg and Reiss have a point: in looking similar to both theory and experiment, surely some of what simulation does can be explained in terms of established ways of learning about the world? Just because computers are new does not mean what scientists do with them constitutes new philosophy. Norton and Suppe go ahead and state that “simulation modelling is just another form of experimentation, and simulation results are nothing other than models of data. Thus, the epistemological challenges facing simulation models are, at bottom, identical with those of any other experimental situation involving instrumentation.” Less radically, Weissert suggests that the epistemology of simulation is at least derivable from the epistemology of experiment, though he mains that there are differences between the two. On the other side of the fence, the models as mediators view reduces the epistemology of simulation to that of modelling, which has a longer history in philosophy of science and a far more impressive bibliography. Morrison, in her later work, still sees through the lens of models when she goes on to use simulation to recast our understanding of experiments, especially those larger and more complex experiments of modern science. With regards to philosophical novelty then, the question we have to answer is which parts of simulation can be identified with old methods, and whether those methods are continuous in this new medium. The fact that this question cannot be answered yet means works like this thesis are necessary.

Furthermore, simulations are worth attention in philosophy of science if only because they are modern science – particularly for disciplines like astrophysics and climate science, but also in nanotechnology, particle physics, evolutionary biology (see section 2.6.1). Simulations play a central role in the day to day work of doing science simply because computers are the only tool that can deal with the mathematics involved in today's studies of complex systems, and simulations are one of the few techniques that can make physical sense of this mathematics. Humphreys reminds us that “one of the primary features that drives scientific progress is the development tractable mathematics”. It is important to bear in mind that the ‘in practice’ is just as important as the ‘in principle’ when it comes to scientific progress – often new techniques further our knowledge more than new theories. Without computer simulation, it would not be possible to handle the large amounts of data, or understand the nonlinear relationships in a meaningful way; Humphreys agrees, making the bold statement that simulations “have introduced a distinctively new, even revolutionary, set of methods into science.” Whether or not one agrees with this last declaration, philosophers ought to pay at least as much attention to simulations as scientists themselves do.

99 Winsberg, 2009a.
100 Norton & Suppe, 2001, p. 92.
102 Morrison, 2015.
104 Humphreys, 2004, p. 57.
I can proffer at least two suggestions for why the state of the simulation literature is so disparate on the subject of theory versus experiment, and old versus new. The first is that there seems to be some indication among other simulation case studies that the use of simulation varies, sometimes radically, across disciplines. Günter Küppers and Lenhard see that in climate simulations, “scientific theory is becoming less important and partly replaced by practical ad hoc strategies in knowledge production”, while Evelyn Fox Keller notices a similar lack of underlying theory in biological simulations. On the other hand, Morrison, in focusing on high energy particle physics, assumes the presence of an underlying theoretical model of the phenomena to which the simulation can refer. Ann Johnson and Lenhard (having written on nanotechnology and climate science respectively) make an argument that simulation is leading to a new culture of prediction, since “computational models’ chief attraction is that they produce a predictive answer, not a mimetic model of the causal mechanism of a real-world phenomenon.” While prediction is also an important part of astrophysical simulations, it seems to be a very real goal of modern astrophysicists to create more and more realistic simulations that can stand in for the real world phenomena (see Chapter 4). Keller, in constructing a semi-historical taxonomy of the use of computers in science, distinguishes between ‘computer experiments’ where simulations follow the dynamics of systems of idealised particles; and simulations that are used to instantiate and construct models for which there is no or little theoretical underpinning. The former type is often referred to as ‘heuristic’, aiding in model formation and subservient to ‘real’ experiments, while the latter type is where simulations have their triumph because, Keller claims, what is simulated is the phenomena itself, rather than a set of equations or particles. Yet this chapter, and the case studies, show that in astrophysics the N-body simulations both follow systems of idealised particles and construct models for phenomena that are poorly understood. Keller’s focus is on the complex systems of biology; similarly, the tangle of interactions in social simulations leads Kleimt to distinguish between ‘thick’ simulations for use in ‘practical’ science, and the far more speculative ‘thin’ simulations used in social science. In astrophysics, however, one of the ‘hard’ sciences, it does not seem to be a case of either/or when it comes to constructing simulations from unambiguous equations yet for systems whose mechanisms are poorly understood.

This is not to say that these studies are irrelevant to simulations in astrophysics, or simulations in general – Küppers and Lenhard, Keller, and Morrison would no doubt all agree that simulations are used to make distinct models function well together (Winsberg develops a similar idea with the ‘handshaking algorithms’ used in quantum mechanics and surrounding disciplines). Johnson and Lenhard make some excellent and relevant points about the commonplace nature of

105 Alternatively, Franck Varenne suggests that the diversity can be derived from a conceptual analysis of the individual symbols at stake in computer simulations (Varenne, 2013).
110 Kleimt, 1996.
111 Winsberg, 2010, chapter 5.
simulation and how it has affected the way scientists work, emphasising the epistemic opacity of simulations and their exploratory mode of working. Simulation, it should be emphasised, is a multi-faceted and incredibly useful technique and as such finds itself in many different contexts and adapted to many different roles – Humphreys calls simulations “a set of techniques, rather than a single tool”.\footnote{Humphreys, 1991, p. 501.} We cannot assume that simulations that attempt to model artificial intelligence, or track the evolution of a galaxy, or duplicate the construction of nanoparticles, all produce knowledge in the same manner. Certain disciplines may emphasise ‘experiment’ more than ‘modelling’, or ‘construction’ more than ‘representation’. If an epistemology of simulation is even possible it would need to include the wide variety of roles and uses simulation in which participates.

A possible second reason for the wide variety of interpretations about simulation and its place in existing philosophy of science is that simulation can only be spoken of in terms of two categories – ‘theory’ and/or ‘experiment’. Though widely bandied about when talking about simulation (including in my own work), these two labels are not only rather ill-defined but can also be limiting if one only considers simulation in relation to established ways of doing science. Morrison’s work reveals interesting dialectics between models and experiment, and so often ends up talking only between those two categories. Even Winsberg’s study of climate simulation, included because “no book on computer simulation in the sciences would be complete without some discussion of its application to the study of our planet’s climate and its future”,\footnote{Winsberg, 2010, p. 93.} ends up talking far more to climate models than it does to computer simulation. In talking of simulation in terms of experiment and theory, we are attempting to find its epistemological place in science among other methods of learning about the world – but doing so leads some scholars to suggest that simulation can be sufficiently explained only in terms of those already existing categories. Some lean more heavily on theory (or modelling), others on experiment, others on a combination of the two. On the other hand, the inadequacy of the existing labels to account for the activity of simulations does seem to indicate that there is something new going on.

It would be foolish to reject outright the categories of ‘theory’ and ‘experiment’, in part because they are useful for understanding much of simulation activity and in part because they play an important role in scientific discourse (see Chapter 3). However, rather than wed simulation to theory, experiment, or some sort of theory-experiment hybrid, this thesis attempts to maintain the focus on simulation (however nebulously defined) and how this technique assists in producing knowledge. Necessarily this also involves discussions on how simulation abuts against theorising or experimenting – Chapter 2 explores how simulation becomes established by borrowing from both traditions, Chapter 3 discusses the how simulation is portrayed as ‘theory’ or ‘experiment’ and the effect of this on the legitimisation of simulation knowledge, and Chapter 4 in part looks at the interplay between simulation and (the lack of) experiment. Despite the difficulties of loaded language, it is possible to talk about these intersections while still maintaining the focus on how simulation produces knowledge. In doing so, the question of
novelty will be answered by describing simulation as a technique that adopts existing ways of
doing science, but in a transformative manner that reconfigures many of the categories of
traditional science and how they relate to one another.
2 Man and Machine: Themes in Simulation from 1963 to 2016

2.1 Introduction

The previous chapter highlighted the polymorphous nature of simulation, a technique that does not comfortably fit in any existing scientific tradition and which seems to pose some tricky epistemological puzzles. From the modern perspective, simulations "have lost their poor image and low ranking in the cognitive and epistemological hierarchy as a last resort, and have obtained a kind of autonomous status."\footnote{Küppers, Lenhard, & Shinn, 2006, p. 10.} How and why, then, has simulation been adopted so ubiquitously, and so quickly, across virtually all of science?

In providing a partial answer to this question, this chapter seeks to historicise simulation. I therefore begin with the historicist assumption that the objects or concepts of knowledge do not have an ahistoric nature that can be sufficiently defined by conceptual analysis. That is, to understand something – like simulation – we must understand how it came to be, a sentiment echoed through the varied spectrum of approaches found in studies that integrate both history and philosophy of science. A historical perspective is necessary for a well-rounded understanding, and helps inform current philosophical debates and rethink epistemic assumptions. Obviously it is beyond the scope of this chapter to provide a thorough historical analysis of simulations and its origins. Instead, the goal of this chapter is to bridge the philosophy of the previous chapter and the astrophysical case studies of Chapters 3 and 4 by describing what the connections between simulations, materiality, models, and experiment looked like fifty, forty, or thirty years ago through the identification of several 'themes'. A partial answer to the question of simulation's rapid and ubiquitous acceptance will be provided by showing how simulation was linked to the existing and well-established traditions of theory and experiment. Thus this chapter also underscores the importance of understanding how the technologies involved – both material and social – affected the justification of simulation as a technique for knowledge generation. Following these various threads up to present day also serves the secondary, historiographical, goal of this chapter by providing a reflection on how historicising simulation helps us understand its epistemology.

To achieve its aims, this chapter traces a number of ‘themes’ that emerged over the course of fifty years in an academic and peer-reviewed simulation journal, Simulation, published by the Society for Modeling and Simulation (originally Simulation Councils, Inc. or SCI).\footnote{The journal is still in publication and produced issues throughout the writing of this thesis. The main material for this chapter was retrieved in January 2014. Newer issues of Simulation have little effect on the outcomes of this chapter, as few articles after the 1980s are used.} Simulation is “the most well-read and respected journal in the field,”\footnote{Shinn, 2006, p. 191.} and provides a unique window into the evolving epistemological concerns and problems of the uptake of simulation as a technique in the later part of the 20th century. The journal provides several introspective articles about the nature
and use of simulation, a rare occurrence outside of historical- or philosophical-themed publications, which is mostly thanks to the efforts of the journal’s first editor John McLeod. McLeod, described as one of the “pillars of simulation in the United States since World War II”,117 wrote several editorial pieces urging readers and contributors to participate in discussions about simulation beyond the practical, and the journal’s first few decades are excellent for gaining an understanding of the concerns of the time.

From Simulation’s conception in 1963 to its ongoing publication today, it is possible to identify several themes that characterise different aspects of the simulation technique. It is the identification and description of these themes that will prove instructive, as they illustrate how particular aspects of simulation came to the fore during different times. The themes overlap each other, and do not rule each other out; they not meant to be restrictive or even particularly meaningful as categories, but rather to provide a means of clarification for how our understanding of simulation has changed. Modern simulation bears the manifestations of themes that dominated in previous times, as well as developing its own association. Co-evolving with the technique and reflected in the themes is the complementary development of the practitioner of simulation – the simulationist. As a new technique, simulation created space for a new expert, a role that naturally changed in tandem with simulation. This chapter also aims to develop how the understanding of the user of simulation shifted.

The chapter begins with the 1960s and the earliest theme of the physical computer. Though simulation is more commonly understood as virtual, the 1960s debates of analogue versus digital computers in Simulation gives simulation a physicality by focusing on the computer. This physicality connects simulation to the more material practice of experiment, setting up an analogy between the two that remains today. The second theme of modelling becomes apparent during the 1970s, when the focus shifts to both the model behind the simulation, and the modeller. It had become obvious to simulation practitioners that a good simulation had to incorporate a properly validated model, and this was only possible if the simulationist was also the creator of the model, as well as its implementer. The connection between the simulation and the simulator is expanded upon in the third theme, which is the relationship between man and machine in simulation practice. This theme is important from the 1960s to the 1980s, when a desideratum for a good simulation was the establishment of a rapport between the computer-simulation system, and the simulationist. Such rapport enabled greater understanding of the model or target system, and constructed an analogous embodiment relation between the simulation and its user. It was through this relationship that simulation looked the most like experiment. However, this theme also had its dark side, with worries of seduction and the confusion between representation and real also colouring the period.

Through the discussion of these themes, it will also be argued that simulation has become acceptable scientific practice through a ‘borrowing’ of credibility from the established scientific

traditions of experiment and theory.\footnote{The terms ‘theory’ and ‘experiment’ are used as shorthand for the collections of practices in science that are commonly identified with such labels. It is only of importance that one agrees such traditions are long-standing and involve methods that are epistemically respectable as science, though one recognises the philosophical difficulties inherent in using blanket terms.} In establishing continuity with both theoretical methods – namely modelling – and experimental practice, simulation was set up as a technique that followed on from existing scientific methods; rather than producing a ‘revolution’, simulation was adopted nondisruptively. The ‘Selling Simulation’ section follows on from the model and man-and-machine themes, expanding in more detail how simulation became established nondisruptively into existing practice via both non-explicit methods – using existing traditions to establish credibility – and through a more deliberate ‘selling’ performed by practitioners of simulation to generate acceptance for and use of their technique. In doing this, the practitioners of simulation affected our philosophical understanding of it, showing that the social history of simulation is also relevant for its epistemology.

This chapter then moves on to modern-day simulations. Having only attracted significant attention in the last decade or so, there is very little HPS literature on simulation prior to 2000. Philosophers are then arriving at a technique that has a hidden historicity, so the themes identified in this chapter, for example, are interpreted only in light of how they manifest in modern simulation. From this perspective, simulation seems to be a confusing blend of experiment and theory, when in fact these aspects can be understood at least in part as the results of the evolution and self-validation of a new, \textit{sui generis}, technique. The penultimate section of the chapter goes beyond \textit{Simulation} in order to discuss simulations run on supercomputers – supersimulations – and simulating as a daily scientific activity. It is suggested that there is a type of scientist that is nothing but a simulationist. Because of this increase in access and modelling capability, the modern period is characterised in terms of power. Finally, the last section of the chapter draws together the many threads suggested by \textit{Simulation}: how the conception of simulation has changed (using themes), how simulation has become nondisruptively established as scientific activity through borrowing from existing traditions, and what effect a blended historico-philosophical approach has on understanding the technique.

2.2 The Physical Computer

As \textit{Simulation} was established in 1963, none of the simulation themes can extend earlier than the 1960s. This limitation is not entirely arbitrary – there is some evidence to suggest that before the 1960s simulation was not particularly useful as a scientific technique. In 1965, simulation was described as “new, as a new-born baby is new”,\footnote{McLeod, 1965a, p. 265.} and a later retrospective recalled that “computer simulation was not a useful tool in the 1950s”.\footnote{Reitman, 1988, pp. 1-2.} One \textit{Simulation} author felt that even in 1977 the simulation field was extremely new – that what had come before was its conception
period and what came after was the "walking and talking stages".\textsuperscript{121} In astrophysics, papers using the word 'simulation' only began to appear mid to late 1960.\textsuperscript{122} It seems, then, that for describing changing understandings of simulation as a scientific technique, the 1960s are a good place to start.

Simulation in the early 1960s was a different beast to that of today, and this is precisely because of the early, physically larger computers.\textsuperscript{123} Simulation was not conducted, as we know it, in the office or laboratory on a personal computer, but in a simulation \textit{facility}. This was a separate building or set of rooms that could house the large amounts of equipment that made up a computer that could simulate. Naturally, access to these facilities was not readily available; early articles in \textit{Simulation} show that their authors were conscious of a hierarchy of use and funding. At the top, as one author described, were management who controlled the funding. The next tier down housed the simulation supervisor, whose job (apart from supervising the simulation) was to provide the information to sell simulation to management (this was of central concern to the authors of \textit{Simulation}, see section 2.5). Then came the simulation engineers, who provided technical capability when it came to building simulations and getting results, and who were the 'simulation experts' of the facility. After them were the simulation technicians, whose job it was to keep the equipment (the computers) working. Finally, (perhaps more outside of the hierarchy than at the bottom), came the problem originators, customers, or clients, who provided the reason for running the simulations (and the income).\textsuperscript{124}

Running an early simulation involved a business relationship between those who had the problem, and those with the technical skill and computers to solve it. The client wanted something to be simulated, and so approached a simulation facility whose employees would use the computers to run the simulation and then send the client the results. The degree of participation the client had in the actual simulation varied, especially at this early time. It was the simulation engineers (sometimes simply called 'users') who actually simulated, as these were the people that built and interacted with the programs the computers ran in a simulation facility, yet it is important to remember that they were by far not the only professional involved in the construction of a simulation – there were also those who ran the overall facility, the technicians who kept the equipment running, and the client who provided the problem. Early simulation was distinctly geographic, with a specific simulation being performed at a specific site. This meant that some of the earliest concerns of simulation users were boringly pragmatic – an article entitled 'The Art of Simulation Management', suggested that the biggest hazards to look out for were fire, heat, pests, lightning, improper grounding of wires, and outdated components.\textsuperscript{125}

\textsuperscript{121} Pritsker, 1977, p. 133.
\textsuperscript{122} From a search of the SAO/NASA ADS database (http://adsabs.harvard.edu) for papers containing 'simulation' (or synonym), February 2014.
\textsuperscript{123} Though there is some similarity in practice between early large computers and today's large supercomputers; see section 2.6.
\textsuperscript{124} Clymer, 1968.
\textsuperscript{125} Clymer, 1964a, pp. 4-5.
The early period of *Simulation* emphasised simulation’s physicality – the places in which simulation was performed, and the computers on which simulations were run. This theme can most strongly be seen in *Simulation*’s 1960s and early 1970s articles that debated the relative merits of analogue (or, more accurately, hybrid) and digital computers. There was a difference in the process of computation for analogue and digital computers, and for early versions of these machines this difference was significant enough that it had an effect on the types of problems for which each type of computer would be used. Analogue computers use continually changeable aspects of physical phenomena – like electrical, mechanical, and hydraulic processes – to model the problem being solved. Analogue is characterised by continuous processes; broadly speaking, using one physical phenomenon to simulate a different physical phenomenon (sometimes referred to as a laboratory simulation) is technically performing an analogue simulation. Digital computers, on the other hand, represent changing qualities symbolically, and are characterised by discrete processes. They are programmed using logic. Hybrid computers are a mix of both analogue and digital, though how this mix comes about varies – in practice, ‘hybrid’ was used for any device or facility that made use of both analogue and digital components in whatever configuration.\textsuperscript{126}

Initially, analogue computers ruled the scene. Developed earlier on, analogue was naturally quicker, cheaper, and better understood than digital. One author, as part of the checking process of a simulation he ran, compared his problem setup on analogue with that on a digital computer. The situation was not good: “the digital computer had to be programmed in machine language; so often the analogue campaign was over before the digital solution could be reconciled with analogue results”.\textsuperscript{127} Speed was obviously valued, as the less time the simulation took, the better. There were a few elements to this – computation time dealt with how long the computer took to perform the calculations, while turnaround time was the time taken from submission of a process to the return of the complete output. Digital was, early on, slower than analogue in both these aspects. Analogue was also more accessible, which apart from the obvious implication of physical access to the computer, generally indicated ease of use; programming in machine language for digital took too much time, and was also an extra and complicated skill to learn.

In the transition period – when digital was roughly on par with analogue – the main area of contention was cost, of both time and money. Which system you preferred here depended on what you wanted to do with it, the most important aspect of this being problem size: the size of the problem on an analogue system was proportional to hardware; the bigger the problem, the larger computer you would need, literally. In digital systems, problem size was connected to time, as bigger problems took longer to process. Thus the cost was dependent on the size of the problem the facility typically solved – large analogue computers cost more than digital

\textsuperscript{126} Broadly, general analogue components, plus converters, plus general digital components, makes a hybrid computer. However, analogue components, plus built-in converters, plus digital logic also results in a hybrid. John McLeod distinguished between these two usages of hybrid: for him, the former should be called a ‘combined simulation’, and the latter a ‘hybrid simulation’ (Clymer, 1964a, p. 11, first footnote).

\textsuperscript{127} Clymer, 1964a, p. 8. A brief note on spelling: to maintain consistency, throughout this thesis American English spellings are corrected to British English spellings in all quotes.
computers, but the increase of time cost might mean that digital wasn’t worth it. It also depended on the type of problem being modelled, as analogue systems were better at integration, and digital better at multiplication and function generation. Sometimes, however, it depended purely on available equipment; according to McLeod: “a digital computer is often used for simulation not because it is the best tool for the job, but simply because it is there, having been acquired for some other purpose”.  

Yet despite having the early advantage, analogue faded quickly. By the mid-1960s, articles from Simulation indicate that whereas digital had once been useless, it was now increasingly useful. By early 1970, the argument was between digital and hybrid, and pure analogue was rare. One facility published a paper in Simulation that described the factors that influenced them in choosing to go exclusively digital, and it was clear that digital had usurped all analogue’s initial advantages: digital was easier to learn, faster to debug, had clearer documentation, was more precise, and was cheaper.  

Digital was also becoming superior in speed. Initially digital simulation was quicker over shorter runs and so was preferred for the sorts of problems that required fewer runs and when the time to get the first answer was at a premium. By 1970, thanks to technological improvement, digital and analogue were about on par in terms of overall speed of computation. Digital was, however, better than analogue on turnaround time. With a hybrid setup, an engineer had to sign up for a block of time using the computer laboratory and then try and solve their problem in that time frame. The digital computer could take card inputs and return results in much less time overall. Documentation was also more readily available for the digital computer, and had complete and correct information. The hybrid system, on the other hand, had at least two different sections of documentation (for the analogue and digital components), and often the analogue section (normally a wiring diagram) was out of date (this also made debugging harder). The final nail in analogue’s coffin was that while the cost of digital components was reducing, the cost of analogue components had remained the same. “The cost advantage which analogue computers had held for a decade is about gone”, said Clymer in 1964; for digital “the cost was lower, and the dependability was higher”. Eventually digital had lower maintenance and expansion costs, better accuracy and precision, and better ease of use.

The theme of the physical computer, of a material presence to simulation, can also be seen in the understanding of the simulation practitioner. As the machine changed from analogue to digital, so too did the user or simulationist experience a shift in role. During the early stages, many of those used to analogue felt a natural resistance to learn the different approach required for digital computers. One way around this were the ‘simulators’ which ‘simulated’ analogue and hybrid computers on a digital computer. This meant that despite using a digital computer, “the analogue

---

128 McLeod, 1972a, p. 3.
129 Svreck & Sandholm, 1971. Here precision refers to the lack of variation between multiple calculations. Accuracy, by comparison, generally refers to the closeness of a calculated quantity to its measured value.
130 Clymer, 1964a, p. 8.
131 Clymer, 1964a, p. 12.
engineer has no difficulty learning to use them. This avoids problems arising from poor communication with digital programmers. It also avoids the difficulties of trying to push analogue engineers into digital training courses against their will.\textsuperscript{134} Ironically enough, it was the digital computer that made the digital computation more accessible to analogue engineers. Of chief interest here, however, is the difference in nomenclature – the analogue users were engineers, while the digital users were programmers.

As digital computers evolved, accessibility became less of a problem. The basics of the digital programming language could be learnt in about a day, versus the week it took to learn how to operate the hybrid machine. Problems also did not have to be scaled to be run on digital like they did on analogue, which made programming the model easier. The development of more efficient and intuitive programming languages gave digital the advantage of accessibility, and soon all simulationists were ‘programmers’. It is important to notice that both the ‘engineer’ and ‘programmer’ descriptions of the simulationist emphasise the physically embodied process of \textit{simulating}. By this I mean that simulating involved the physical material computer and interactions with it, including everyday interactions like typing and visualisation, and also some of the context-related issues like physical access to computers, their geographic locations and space, and the limitations and progression of computational technology. Coupled with the focus on the physical computer, the articles of \textit{Simulation} definitely return the impression that simulating involved shifting around components, the use of punch cards, typing, visual and haptic interaction with a console, and processing reams of printed data.

Highlighting this theme shows that early simulation was centred around the use of the physical computer.\textsuperscript{135} What an early simulationist wanted is illustrated (through hyperbole) by a two-page long advertisement in a 1966 issue of \textit{Simulation}.\textsuperscript{136} The ‘article’ sang the praises of the EAI (Electronic Associates, Inc.) 8400 digital computer and its simulation software, and is ideal as its very purpose was to sell simulation using the right keywords: “this, we believe, is exactly what the man interested in simulation and scientific computation is looking for in a digital computer system”\textsuperscript{,137} The 8400 was the epitome in usability, and operated at a high speed (with an emphasis on a low execution time) with a minimum of cost. It also allowed for the use of both assembly code and FORTRAN (one of the first programming languages) for designing in real time. The latter was important because analogue might be preferred for how it allowed problems to be “worked out right at the analogue console”,\textsuperscript{138} meaning greater accessibility (see also section 2.4). As the analogue and digital debates have shown, the readers and authors of \textit{Simulation} clearly valued speed, cost effectiveness, and accessibility, criteria that were the result of improvement in either the physical computer itself, or its effective operation. Monetary cost, precision, and speed

\textsuperscript{134} Clymer, 1964a, p. 8
\textsuperscript{135} Early simulation is naturally closely tied to the development of early computing. For example, Park describes how, between 1945 and 1965, the electronic digital computer was responsible for both the substantial growth of quantum chemistry as a distinct discipline, and also the emergence of specialised groups within it – a split between the computors and the experimentalists (Park, 2003).
\textsuperscript{136} Electronic Associates Inc., 1966.
\textsuperscript{138} Svrcek & Sandholm, 1971, p. 245.
of computation were properties of the computer equipment, and were improved through the development of electronic technology. Accuracy and accessibility were properties of the correct operation of the equipment, on the most basic level – the improvements in digital programming languages greatly assisted here, but of additional importance was the skill of the expert using the computer. The role of the simulationist, during this period, also revolved around the physical computer. Whether calling oneself an engineer or programmer, the early simulationist needed to know chiefly about the computer – how to operate it effectively and accurately, how to choose the right tool for the job, how to arrange analogue components or program digital ones.

This early emphasis on the physical computer, and how it is central to understanding simulation in the 1960s, is in many ways in direct contradiction with much of the current literature, which emphasises the simulated object as inescapably virtual and symbolic. Yet, as Chapter 1 discussed, the material similarities argument is beguiling precisely because simulation is enacted on a physical computer, and can therefore perhaps be like experiment in that simulating involves intervening with an instrument. For example, one might look at simulation through the lens of Davis Baird’s approach to understanding the role of scientific instruments in producing knowledge. Instruments, despite the enormous variation in that category, all have what Baird calls a ‘thing-y-ness’: “things occupy space, have mass, are made of impure materials, and are subject to dust, vibration, and heating and cooling. Things are built in ‘real time’ and must produce their work in real time.” To have a ‘thing-y-ness’ means having a physical presence and interactions.

Simulation’s ‘thing-y-ness’ was certainly apparent to its early practitioners. Simulation is not generally understood as sited, but for a time in the 1960s there were distinct geographical places (the simulation facilities) where simulation was performed. The users of these sites were concerned about heating problems, pest control, and outdated components. They wanted equipment that was faster, cheaper, and easy to use, that responded immediately to your input and ran the output before your very eyes. In addition, it was the improvement of the components that made up the physical computer that constituted progress in simulation during the early period – advancing through better memory and processing capabilities, increased availability, and a reduction in the cost of components and maintenance. Baird’s instruments, too, tied progress to such physicality; his case study of spectrochemical instruments between the 1930s and 1950s has several similarities to early simulation. For the chemists, a reduction in size of the instrument made it much more portable – “transportation is a serious obstacle to dispersion and progress”, as Baird’s actors struggled with the same issues of accessibility. Improvements in the materials and knowledge of how to manipulate them were key (Baird emphasises that these improvements were not directed or prompted by theoretical work), the reduction in time taken

---

139 Baird, 2004. Baird, like other philosophers discussed in this chapter (as we shall see) emphasises the importance of recognising the role of technology (or instrument) in history of science.
was crucial in implementing the instruments in any real capacity, and safety and ergonomics, usually dismissed, also played a role.

We see these connections between simulation (more precisely, the computer) and experimental instruments because simulating is an embodied, physical practice. Early simulators were very well aware of this, and centrally concerned with it as their instrument – the computer – was young and limiting, yet also leaking potential and about to undergo exponential growth. The material similarities argument still does not stand, because intervening on the material computer to explore a virtual object is not the same as intervening on the material object itself, as it seems we have characterised (caricaturised?) experiment to do. Yet this more historical perspective reminds us that simulation, or at the very least, simulating is physical, and at one point was consciously understood to be so. This is important to point out because during later periods, when simulation moves on to different themes, the physical computer is no longer mentioned when talking about simulations. In modern simulation the physical computer is almost invisible in the final results; understandably – today, given the ubiquity of computers capable of simulation, it is certainly less important what the specifics of your simulation instrument are. Thus the peculiarity of emphasising the physicality of the computer in modern philosophy on the subject. So where does this physicality go? Why does modern simulation (and the knowledge it produces) feel thoroughly virtual? Is it because simulation is no longer specifically sited, so determinately geographic and physically institutionalised? This may be part of the answer to the question, but echoes of physicality still appear in the publications on modern supersimulations which, like early large computers, are housed in special facilities. However, as shall become apparent in later themes in *Simulation*, as the focus changes from the physical computer to the model, interplays emerge between the user and their physical computer that acts to elide the early physicality of simulation.

### 2.3 Models and the Modeller

By the early 1970s, simulation techniques were spreading to more and more fields of investigation; from its humble beginnings in the aerospace industry, simulation was applied in the engineering and physical sciences, but also the biological and medical sciences, economics, and the social sciences.\(^\text{142}\) The reasons for this were manifest; McLeod, even in 1972, was already summarising the most common reasons given for using simulations by historical and modern actors alike: simulation provided a better understanding of system requirements when the system did not lend itself to experimentation (systems that were too difficult, expensive, or dangerous to manipulate); it was the only practical way to study the dynamics of complex systems; it was the only way to study a system that does not exist; and it was useful when there were difficulties in getting a measurement from the real system.\(^\text{143}\) All these reasons for using simulations boil down to one essential value: accessibility. This is not accessibility in the previous

\(^{142}\) McLeod, 1972b, pp. 5-6.

\(^{143}\) McLeod, 1972b, p. 5.
use of the term – ease of use – but an epistemic accessibility, where simulations provide access to
the target system, which would otherwise have been impossible, inadvisable, or unfeasible. This
epistemic accessibility is indicative of a changing focus in the 1970s towards the simuland – the
real world system being modelled – or rather, the model itself.

As a segue, consider an article published in 1979 on the ten biggest causes of failure in simulation
and analysis. There were generic causes of failure, of little interest here, that resulted from bad
project management like poor communication, premature coding, or an overly optimistic
schedule. There were causes of failure that demonstrated a continuity with the theme of the early
period of the physical computer, such as using the wrong computer language, and having obsolete
or non-existent documentation. Importantly, however, the main cause of failure was using an
unverified model. This went in hand with the failure to define an achievable goal: the authors
specified that ‘to model X’ was not a goal, but a means. The remaining causes of failure listed by
the Simulation article were malfunction on behalf of the human agent – having an incomplete set
of essential skills, that is, a team that lacked skills in leadership, modelling, programming, and
knowledge of the simuland, would produce a simulation that failed. An inappropriate level of
detail was a failure on the behalf of the judgement of those developing the simulation – too much
detail and the desired results were obscured, too little and the results were inconclusive.
Inadequate end user participation – not involving the problem originators in the development of
the simulation – would produce a simulation that did not have adequate knowledge of the
simuland, or that failed to achieve its goal. All these latter failures link back to the creation of the
model, suggesting that the process of modelling had become part of simulating.

Increasingly among the articles of Simulation during the 1970s it became apparent that to
produce a good simulation, the simulationist must also have close knowledge of his simuland. An
expert simulation technician, initially trained in physics, was asked to do simulations of economic,
social and ecological systems. In each case he asked of his clients the exact details of the time
evolution of the system, supporting theories, and real data for checking the simulation, but
"needless to say, after my requests they all decided that they wanted someone else to do the
simulation". His point was not that the systems in the so-called ‘soft sciences’ were less well
defined than those in the ‘hard sciences’, but rather that the simulator cannot be ignorant of his
field and expect to make a good simulation. This is a change in focus from the earlier period where
the simulator-engineer was more concerned with computer and its programming than the
contents of the simulation.

To give an example, a university-level course, taught in the late 1970s, presented its students with
large, complex projects. It was then up to the student to break down the system into manageable
parts, and develop a simulation that incorporated a model that was appropriately simplified and
implemented. The reason for teaching simulation in this manner was so that the student would
be forced to interact with the system they were modelling. The course specifically aimed to

develop practitioners of simulations: “the traditional approach to teaching simulation (lectures followed by a series of short assignments due in a couple of weeks) does not produce simulationists; it produces appreciative audiences for simulation”. Contrariwise, this course aimed to teach students how to model and simulate large, realistic systems, after which “the student has become a simulationist”. Most importantly, however, the emphasis was not merely on knowing your simuland, but on modelling: “to fully appreciate the scope and power of computer simulations, one must become a practitioner. ... One does not become a practitioner by watching someone else construct a model, or by coding a fully developed model, or by modelling trivial systems”. It was only by actually constructing the model oneself, that made one a true simulationist.

There are two ways of looking at this: the simulation engineer became a modeller, or the modeller gained simulation abilities. In practice, on an individual level, it was most likely a mixture of both; a sociological or historical study examining the career trajectories of early simulators would be useful here but is beyond the scope of this thesis. However it happens, this theme shows the shift towards a more epistemological understanding of simulation; from the focus around the enactment of the simulation (programming or engineering) to a focus around the theoretical component, in the hands of the modeller. From a historiographical perspective this shift is interesting because it shows us that such a central and epistemologically vital component of simulation as the model was not dominant from simulation's beginnings, but only found its clearest expression a decade later. This is not to say that prior to the 1970s there was no modelling in simulation, but that the understanding of simulation definitely has historicity.

The shift to modelling is also interesting from a philosophical perspective because the awareness of the model emerged as simulation practitioners started to realise not only that the best simulations were constructed by those with good knowledge of the simuland, but that a simulation model cannot necessarily be validated based purely on the integrity of the theory from which it derived. There was greater concern during the 1970s in Simulation with verification and validation of the simulation, one article determining that the ultimate credibility of any simulation approach lay in the methods’ ability to pass verification and validation tests. Demonstrating continuity with the theme of the previous period, verification dealt with the correctness of the algorithms themselves, and the correct programming of the computer. Validation, however, concerned the correctness of the model, and was in fact far more discussed than verification; the latter was generally considered, rightly or no, to be the tedious but unproblematic task of rote debugging.

Validation had become of paramount importance. Reflecting on the early 1950s, in 1973 McLeod noted that checks of the accuracies of a simulation "was to verify proper programming and proper
operation of the computer. The validity of the model being implemented was apparently assumed!".\textsuperscript{150} In early 1970, this assumption seemed naïve, simulation having expanded beyond the realm of computing to involve the incorporation of theoretical models. The simulation needed to correctly implement a model that was a valid representation of the simuland before it could even begin to deliver ‘correct answers’. The emphasis was on the necessity of looking at the accuracy of the model to ensure that the simulation was good, and not confusing the accuracy of the simulation computations with the accuracy of its results pertaining to the real world. Necessarily, these models needed to be checked as part of validating the simulation, and McLeod urged simulation users to rigorously check the model in multiple ways, emphasising that all these checks must be performed – not just the seemingly most appropriate – for confidence in a simulation model to be well-founded. Suggestions for checks included: using intuition (does the model ‘feel’ reasonable in response to perturbations?), curve-fitting (comparing the output of the simulation to a similar output of experiments on the simuland, as best one can), and using a ‘check solution’ (comparing the output of the simulation to that of another, similar, simulation).\textsuperscript{151}

Though I have argued that simulation activity should not be reduced to modelling practice, there was undeniably a very important part of the construction and operation of the simulation from the 1970s onwards that was tied to the model. A report published in 1979 by the SCS Technical Committee on Model Credibility aimed to define a standard set of terminology, "since the cornerstone for establishing the credibility of a computer simulation is effective communication between the builder of the simulation model and its potential user".\textsuperscript{152} They based their terminology around the construction of a diagram of the ‘simulation environment’ (Figure 2).

Apart from ‘reality’, the basic elements of the simulation environment were both models. The conceptual model, defined as: “verbal description, equations, governing relationships, or ‘natural laws’ that purport to describe reality”,\textsuperscript{153} can be broadly understood as a theoretical model in philosophical parlance. The computerised model was close to the ‘simulation model’ described in Chapter 1: “an operational computer program which implements a conceptual model.”\textsuperscript{154} It is important here to realise the distinction between ‘simulation’ and ‘simulation environment’; the actual computer simulation was in the environment as a process that linked the computerised model to reality. It was described as the “exercise of a tested and certified computerised model to gain insight about reality.”\textsuperscript{155} Though it is perhaps somewhat restricting to limit the definition of

\textsuperscript{150} McLeod, 1973a, p. 10.
\textsuperscript{151} McLeod, 1973a, p. 11.
\textsuperscript{152} SCS Technical Committee, 1979, p. 103. Note the differentiation between builder and user, a boundary which was on its way to being effaced but was not quite there yet.
\textsuperscript{153} SCS Technical Committee, 1979, p. 103 Emphasis removed. The definition is intentionally imprecise in an attempt to incorporate both simulations of measurable physical systems and those of social and biological systems where data may be ill-defined.
\textsuperscript{154} SCS Technical Committee, 1979, p. 103 Emphasis removed.
\textsuperscript{155} SCS Technical Committee, 1979, p. 104. Emphasis in original.
simulation to that involving a thoroughly certified model, it is interesting to see how, in this conception of the process of simulation (which is probably intended to be understood as a guiding generalisation rather than a statement of fact), simulation was, though separate in name, inextricably involved through interaction with model generation and validation.

Indeed, apart from 'model verification', which is a distinctly simulation activity (concerning appropriate programming, for example), both 'model qualification' ("determination of adequacy of the conceptual model to provide an acceptable level of agreement for the domain of intended application"\textsuperscript{157}) and 'model validation' ("substantiation that a computerised model, within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model"\textsuperscript{158}) are activities that occur, \textit{mutatis mutandis}, with standard, non-simulation, models. Indeed, it is because simulation looks so much like modelling that many researchers have a tendency to either consider simulation as model-applied, or reducible to modelling. Sismondo identifies several commonalities between models and simulations – so

\textsuperscript{156} This being defined as the acceptance by the user that the certification documentation (which is given its own definition) is adequate evidence that the model can be effectively utilised in a specific application. Again, the definitions are intentionally imprecise, which allows for the definitions to require, for example, that the conceptual model has demonstrated agreement with reality, without attempting to determine how this demonstration could come about.

\textsuperscript{157} SCS Technical Committee, 1979, p. 103. Emphasis removed.

\textsuperscript{158} SCS Technical Committee, 1979, p. 104. Emphasis removed.
many, that he uses ‘model’ and ‘simulation’ more or less synonymously – in their positioning between theories and material objects.\textsuperscript{159} For Morrison and Morgan, as has already been noted, the simulation allows mapping model predictions to empirical data, hence simulation is a tool for applying the model, and hence the activity of simulation is, at base, the activity of modelling.\textsuperscript{160} Hughes is less reductionary; his DDI (denotation, demonstration, and interpretation) account of modelling incorporated simulation by allowing for nested stages of modelling, some of which are wholly performed on a computer.\textsuperscript{161} For some philosophers, simulation is best characterised in terms of a model, where simulation understood as a new, or different, or adapted type of modelling activity.

I have presented my reasons for why I consider this to be a misrepresentation of simulation in the previous chapter, but it is undeniable that simulation is in large part modelling. While one may deny that simulation has anything in common with (traditional) experiment, it would be bizarre to see simulation as separable from modelling. It is less bizarre in the context of this chapter only because it is a demonstration of the historical shift of focus of simulation from material to model. The result of this shift in focus is that simulation 'borrows' credibility from the culture of modelling. By looking like a long-established scientific technique, simulation finds its place in the theoretical tradition.\textsuperscript{162} For some philosophers, the logical outcome is that simulation must therefore be an extension of the modelling technique, and it is thus interpreted and analysed in that context. The dissonance of this approach with those researchers arguing for simulation as \textit{sui generis} in a meaningful way, or those arguing for simulation as experiment, is strong motivation for conducting a detailed analysis of simulation methodology.

An example of how the simulation technique demonstrated a continuity with the long embedded scientific practice of modelling is through validation. The idea behind McLeod's and others' urge to validate the simulation was that using modelling techniques to validate the simulation model would perform the same justificatory role for simulation knowledge as it did for regular theoretical model knowledge. Necessarily, the validation techniques were adapted to be applied on a simulation, but were otherwise nothing but the same methods used to check analytic or mathematical models (\textit{sans} 'simulation'). Consider W. Parker's taxonomy of analysing errors in simulation: there may be error in the simulation program design (for example, the sample used was too small, or the numerical methods used to analyse it were inadequate), in substantive modelling (the initial conditions or equations were inappropriate), in data processing (errors producing processing or interpreting raw data before it is used to produce results), in the solution

\begin{itemize}
\item \textsuperscript{159} Sismondo, 1999.
\item \textsuperscript{160} Morrison & Morgan, 1999.
\item \textsuperscript{161} Hughes, 1999, pp. 130-131.
\item \textsuperscript{162} In this thesis, 'model' is understood as 'theoretical model', which involves the construction of a theoretical description of a real system. This is an intentionally vague description as I wish only use the model's connection to the theoretical tradition, though whether one wishes consider the model outside of theory (as a mediator, perhaps), or as the basis of theory (as with the semantic view of theory) is beyond the scope of this thesis. Placing modelling in the domain of theory is only meant as a convenient shorthand that emphasises the how simulation borrows from two classic traditions of science.
\end{itemize}
algorithm (using inappropriate solution methods), in the numbers (a numerical but not analytic solution error due to approximations or computational limitations), in the programming (an errant comma, or faulty logic that causes an infinite loop), or in the hardware.163 Some of these sources of error can be placed unproblematically under the heading of verification — numerical, programming, and hardware errors — a form of justification that comes from the computer sciences. The remaining entries on Parker’s list, however, are model techniques adapted to the simulation process. Designing the program, choosing parameters, and choosing the correct algorithm can be easily mapped to various stages of constructing and implementing a theoretical model. Analysing errors in these stages have similar analogies.

Both constructing the simulation program and validating it using techniques like intuition, curve-fitting, and ‘check solutions’, bear strong resonances with the methods of modelling and through this simulation can borrow credibility by looking like an existing and well established theoretical technique. Importantly, however, these activities are also distinctly of simulation — for example, determining the size of the sample can have both computational and theoretical considerations, as can the choice of numerical method: a particular method may make computation simpler and more efficient, but over-simplify the problem or reduce the desired complexity. Similarly, data processing and the choice of solution algorithms have considerations from both a computational and theoretical standpoint. Despite looking very much like modelling, in construction, validation, and analysis, simulation is not only modelling. Yet the simulationist must also be a modeller, as well as knowing how to translate the model into the computer (and out of it), blending their knowledge of the computer (coding, etc.) with their knowledge of the model.

2.4 Man and Machine

It is now time to turn our attention to the relationship between the user and the simulation. In Simulation between the 1960s to about the 1980s there was a strong emphasis on the rapport between ‘man and machine,’164 between the researcher and the simulation-computer system with which they worked. During this period, part of being a good simulationist meant having good rapport with your simulation. McLeod considered the reason that analogue computers dominated during the early period was because “digital computers of the time were too slow for ‘real-time’ simulation, and also because analogue computers offered — in fact required — so much better

---

163 Parker, 2008. Weissert describes similar instances of what he calls ‘strategies for belief and pursuit’: comparison (with properties of a true solution, like conservation of energy), calibration (using the apparatus to reproduce known results), verification (here meaning comparison with other results obtained by similar models), retrodiction (retroactively affirming the results using an independent theory of phenomena produced by the simulation), and consistency (different choices of parameters gives the same result (Weissert, 1997, pp. 122-125). The key difference with Weissert’s strategies are that they are adapted from an epistemology of experiment, another instance of simulation’s hybrid nature causing contradictory but equally valid conclusions among researchers.

164 Generously, ‘man’ can be taken to indicate ‘humankind’. On the other hand, and unsurprisingly given the time period, all the Simulation articles during this period use masculine pronouns, reflective of the gendered division of labour that was evident in many post-war labs (Galison, 1997, pp. 374-377).
man/machine interaction”. Another author criticised digital for “poor user-machine rapport”, and described it as presenting difficulties in tying hardware or man in the loop. This ‘loop’ was generally taken to refer to the process of (early) simulation that included, but was not limited to, the simulation facility, real-time hardware, users, programmed machines, and so on. The humans were an important part of this loop:

A simulation facility and its staff comprise a ‘man-machine system’ in the fullest sense of that expression. The machine part is the facility itself, including all computers, peripheral hardware, and equipment used to support or protect the simulation hardware and computers. The ‘man’ part is much more valuable ...; it consists of the staff and its customers.\(^{167}\)

As McLeod put it, we are part of the ‘real-time hardware’.\(^{168}\) The customers were valuable because the provided the income, but the staff were much more embedded in the system of simulation. Already understood as a process in time, simulation was also understood as a system that incorporated both the physical computer and the researchers themselves.

This metaphor – of man and machine in the loop – would have been inherited from the broader background of early 1950s and 1960s computing. N. Katherine Hayles’ work on cybernetics from the mid-1940s to the late 1990s deals with a burgeoning area of computing that involved both analogue (in the sense of a physical system) and virtual simulations.\(^{169}\) With particular relevance to this section, Hayles describes how the image of the man-in-the-middle was a feature of early cybernetics, with researchers attempting to understand how the relationship between humans and machines had changed post-World War II. In an echo of McLeod’s words, Hayles notes that “the man is significantly placed in the middle of the circuit, where his output and his input are already spliced into an existing loop.”\(^{170}\) Galison describes a similar tension around the incorporation of the human-machine system into particle physics in the 1960s. Various types of ‘reading machines’ were employed in order to read the photographs that emerged from the bubble chambers. Some physicists sought to create fully automated reading machines – “where human judgment was there FORTRAN would be”\(^{171}\) – that eliminated the man from the machine, leading to concerns that what it meant to be a physicist was changing. Ultimately, however, it proved impossible to entirely remove the human from the equation; the most effective approach was a combined one, where “the thorough integration of ‘scanning girl,’ physicist, and electronic computer was the system.”\(^{172}\)

For a simulationist, the relationship with their simulation was likewise one of increasing (and sometimes problematic) integration. In Simulation, researchers were described as being

\(^{165}\) McLeod, 1972a, pp. 3-4.
\(^{166}\) Clymer, 1964a, p. 11.
\(^{167}\) Clymer, 1964a, p. 3.
\(^{168}\) Clymer, 1964a, p. 13, first footnote.
\(^{169}\) Hayles, 1999.
incorporated into the system by the rapport they maintained with their machines, but what precisely this meant is somewhat vague. From context, user-machine or man-machine rapport indicated part accessibility and part interaction and response – users felt a rapport with the machine when they found it easy or intuitive to use, and when they were able to see the response to their input. There was a great deal of difference between sending a set of differential equations off to be calculated and receiving the response, and participating in real-time (this was important) alteration of your model and watching the effects.

As an example, let us return to the 1966 advertisement for the EAI 8400 mentioned in a previous section. The advert emphasised that "the prime focus for the 8400 system is man: the design engineer, the experimentalist, the simulation engineer." The ‘abstract’ of the advertisement deserves to be quoted in full:

Simulation, experimental design studies, engineering model building – these are computer-aided creative processes. They place great demands on the designer and the computer that assists him. The EAI 8400 and its simulation software is the first computer a designer can really consider a partner. He can talk to it, ask questions about his simulation model, demand the highest performance. He can require that all his own mistakes be found and forgiven, with error-free behaviour on the part of the computer. The creative design engineer – the man who has devised a mathematical model of a new design and wants to experiment with it – needs to be close to the machine at run-time. He needs to modify the program and data during the run. He can’t afford to be hampered by operating details such as octal conversions and symbol searches. He needs the simplicity – and the sophistication – of the 8400.174

The user was placed in the position of considering the computer-simulation a ‘partner’, suggesting a relationship that established a dialogue – the designer could ‘talk’ to their simulation, asking questions in a metaphorical sense that would help them refine their model and learn about their system, providing “real interactive control and conversation for the designer”.175 Immediate response and accessibility was emphasised with the ability to modify their program in real-time, allowing the designer to be ‘close to the machine’ and for them to experiment with it; this immediacy of reaction to a stimulus is very closely tied to the methodology of an experiment, which allowed the word to be used (I believe) metaphorically. Somewhat optimistically, the computer would also take care of any errors by finding and fixing them, as well as processing all the tedious details of operation without any extra input required. The user was the ‘creative’ force, and their partner the computer a diligent worker that was somehow both simple and sophisticated.

Having rapport was crucial to a ‘good’ simulation because it was through simulation that the user gained understanding of the processes and structure of the simuland. A basic example of this was

175 Electronic Associates Inc., 1966, p. i
the use of simulation in the education of engineering or physics students at university level. Students in these fields were confronted with the problem of a ‘hardware gap’, having little opportunity to see ‘pen and paper’ models or systems in action. In one 1964 course, simulation was used as a way of closing this gap, and forming “a bridge linking analysis to laboratory observation”.¹⁷⁶ This quote was accompanied by a figure flowchart (Figure 3), demonstrating a flow from analysis (depicted as equations and diagrams on a piece of paper), to simulation (a group of four men clustered around a computer monitor), to experiment (a laboratory setup which includes a computerised measuring device), back to simulation (this time shown as a computer unit at an unmanned desk, and accompanied with a computer generated graph), and returning to analysis to create a loop. The student used theory, a laboratory experiment, and simulation to solve the same problem, but simulation was depicted twice – once going from analysis to experiment, and again going from experiment back to analysis.

This image has been removed by the author of this thesis for copyright reasons.

Figure 3: “Simulation of Mechanical Systems adds an Important Dimension to Instruction in Engineering Design” (Brickman & Park, 1964, p. 15). The image is probably scanned in from a hand-out distributed to the students concerned.

Here again is the idea of man and simulation within an enclosed loop, this time through the process of knowledge generation or learning. It is also interesting to note, in light of the preceding and succeeding chapters, that both analysis (theory) and experiment are part of this loop.

¹⁷⁶ Brickman & Park, 1964, p. 14. With phrasing like this, it’s tempting to see this role of simulation as an example of the ‘simulation as mediator’ idea discussed in Chapter 1, but this would be mistaking the intent of the author. Rather, simulation is aiding in the students in understanding the connection between what they see in the laboratory and the theoretical interpretation.
Returning to the image: post-analysis, one had to program the problem or system into the computer. Doing so required breaking down the model in different ways, revealing its processes and how these related to the equations or theory. In the charming persiflage of John McLeod, “computers make people think”, as gathering data and formulating a problem in sufficient detail to allow you to program often gave a clearer picture and understanding of the problem, even before any computers got involved. It was this process that has the most human element; by contrast, post-experiment was the purview of the computer and its greater data processing capabilities, yet even these are an aid to understanding because they can reveal patterns the pen and paper method obscured. Even if the results were as expected, solving the problem using the computer “increases a student’s proficiency in data manipulation and interpretation, and encourages him to be creative in his reasoning”, Stephan Hartmann mentions the role of simulations as a pedagogical tool in his list of the five functions of simulation, though it receives the least attention. It was through ‘playing’ with the model and seeing the results visualised that helped students increase their understanding and develop intuitions about the model. Hartmann sees this as a cheaper and faster alternative to performing experiments, and cautions that this is only appropriate when the underlying model is trusted, which is certainly the sense in which the 1966 students were using simulation.

Simulations, for students (and indeed for physicists themselves), provided “a deeper understanding of the physical principles [and] a meaningful link between theory and experiment in problem solving” through the ability for dynamic manipulation (cf. Chapter 1), and for this understanding to emerge, one had to have rapport with the system. This sense of dynamic manipulation was due to real-time response and, in large part, to the graphical capabilities of the computer; as one simulator put it: “the way to a man’s understanding is through his eyes”. The addition of auxiliaries which gave the computer graphical display and plotting abilities enabled “a digital user to enjoy some of the rapport with his machine which has been so central to the analogue man/machine concept”. Developing ways of integrating the user into the process made the simulation feel more manipulable, and set up the simulation as a true man-machine system, not merely the purview of the machine. The graphics did not have to be fancy – one author found that simple alphanumeric ‘moving pictures’ that updated in real time according to the process of the simulation were a powerful way of helping his clients understand the process being simulated better. In the early 1970s, graphical displays were advanced enough to change in real time according to the simulation: “rapid computation and the display of solutions at rates comparable to a motion picture provide means to exploit the pattern-recognition properties of

177 McLeod, 1965b, p. 173.
179 Hartmann, 1996, p. 91.
180 Brickman & Park, 1964, p. 16.
181 Rahe 1972, 16.
182 Clymer, 1964a, p. 12.
the eye and brain”; important for instruction, but also for recognising patterns in system structure, for example.

Embedding human agency in the simulation somehow aimed to bring the scientist closer to the system, perhaps as a way of dealing with the extra layers of abstraction. One Simulation author reflected that: “simulation gets into one’s blood. To simulate a process, one must learn a great deal about it. Then during the simulation one learns even more about the process.” Having rapport with a simulation meant a sense of integratedness with the simulation setup, and through this familiarity gaining greater understanding of the system under study. Nicolas Rasmussen, in a history of the electron microscope, describes a similar rapport between the microscopists and their more traditional instruments. His scientists developed a familiarity and comfort with their microscopes that meant that “the instrument became decreasingly refractive to their wishes, and increasingly predictable in performance.” Not only did familiarity with their instruments increase their skill with the microscope, but it also enhanced the microscopists understanding of their target systems: “notwithstanding – indeed, because of – all the skill and sophistication that scientists employing electron microscopes quickly developed, for them the microscope faded into the background and became a vehicle that projected them ... bodily into the microscopic terrain they were engaged in exploring.”

There are similarities – and also instructive differences – between the sense of familiarity that is developed between the experimenter and their instrument, and the user and their simulation. Rasmussen’s microscopists developed a strong familiarity with particular machines, and things like better placement of controls enhanced this rapport. For the simulating engineering students, it was the increasingly improved graphical capabilities that allowed for real-time observation of the model – a visual factor and an immediacy of response factor – compelling even if expressed in an abstract manner (such as using simple alphanumeric diagrams). As computers, graphics, and simulation techniques improved this sense of familiarity increased (see sections 3.9 and 4.5.1). Of course, the user of a simulation does not engage in the same sort of embodiment relationship that the microscopists would have experienced with their instruments. Yet the relationship is analogous in the sense that the simulationist is positioned with respect to the simulation in the same manner that the experimenter is positioned with respect to the scientific instrument. In both cases the instrument (simulation) is used to explore a system (though whether one allows that it is the simuland that is being accessed rather than the model is a matter for debate). The experimenter interacts with the instrument, using it to manipulate, measure, or observe a physical system, just as the simulationist does with his simulation.

185 See also Timothy Webmoor on what he terms ‘codework’ in modern science, “an emergent and distinct form of practice in scientific research involving visualisation” (Webmoor, 2015, p. 24).
190 Rasmussen uses Don Ihde’s framework; following Ihde, the simulation and user come closer to a hermeneutic relationship than an embodiment one (Ihde, 1991, p. 75).
is facilitated by the rapport the scientist develops with their instrument, and here is where the instructive difference between the two examples comes in. For the microscopists, rapport makes the microscope transparent, an extension of the experimenter rather than an external tool. The simulationists, on the other hand, can only develop a limited rapport in this sense, because simulation is playing the role of both the measuring instrument and the system under experiment.

Simulation’s capacity to act as the system under experiment – in fact, to simulate – is dependent on a degree of opacity. In Dowling’s work, simulations, through the use of random number generators (to model stochastic or probabilistic systems) and the high complexity of the system modelled, are portrayed as ‘black boxes’. This means that simulations can function as opaque, unpredictable entities that receive input and deliver output without any direct reference to an internal structure, much like a real system responds to experimental stimuli. It is through this opacity – the capacity to surprise – that simulations can be conceived of as producing new knowledge via a (simulated) ‘experimental’ process. In this characterisation, the simulation becomes an experimental system (in the sense of experimenting on a model) with which the researcher interacts. Thus, “the technology is not so much experienced-through as experienced-with.” Black boxing allows simulation to develop an ‘otherness’, but also to maintain a referential component – to maintain credibility and justify results, Dowling describes how it is important to be able to ‘open’ the black box and show your work, as it were, portraying simulation “less as an opaque, interactive entity and more as a transparent calculating machine.”

There is then a duality: simulation can be described as both transparent and opaque; a black box, but one that can be opened. Dowling deals with this contradiction by describing simulation’s opacity as changing depending on its role and the stage in the simulation process. However, it can also be understood by interpreting simulation through the tradition of experimentation. On one hand the simulation functions as the experimental system – experimented on – providing knowledge about the modelled system through revealing patterns or relations. To do this, simulation must be opaque. On the other hand, the simulation functions as the instrument – experimented with – allowing the user to manipulate the simulated system to see how it works and which bits do what. Unpacking a model in this way requires the simulation to be transparent. If the previous theme was about how simulation maintained links with the tradition of modelling, this theme is about simulation doing the same with the tradition of experiment.

There is a lot to be said for portraying simulation as methodologically like experiment, that is, the practice of simulation has commonalities with the practice of experiment. Galison’s Monte Carlo workers felt that “the daily practice of error tracking bound the Monte Carlo practitioner to the experimenter”, the search for syntax or logic errors to obtain the correct expectation values in the computer program akin to an experimenter’s attempt to remove systematic errors from their

---

192 Ihde, 1991, p. 75.
193 Dowling, 1999, p. 266.
apparatus. Similarly, both simulators and ‘bench experimenters’ are concerned with increasing precision in their work. Systematic variation of parameters is another commonality, identified by Dowling and it is particularly intriguing to note that in Dowling’s sociological work, “most respondents characterised simulation as similar to ‘theory’ in terms of its relationship to ‘reality,’ and characterised simulation as similar to ‘experiment’ in terms of the types of activity it entails.” The same is being done here: it is in the practice of the simulating that we find simulation borrowing from the experimental tradition. In the ‘man and machine’ theme this has manifested in two ways, as demonstrated in the following passage from Simulation, which attempts to explain simulation’s wide appeal:

Simulation’s success is due not only to its advantages in speed of response and flexibility but also to feedback from and acceptance by users that comes from good graphic displays and good recording facilities. The crux of the matter is the directness and immediacy with which the user/researcher can relate to what he is simulating and the way in which it lets him carry out a convenient program of investigations and experiments. The learning through doing and the understanding derived from the results are the real strengths of simulation.

Firstly, there is the sense of rapport that one feels with a simulation one knows well and that responds immediately and in an accessible manner; moreover, it is the interaction with the simulation that is emphasised, the development of a tacit knowledge that, as Ian Hacking recognises with respect to the microscope, actually enables effective observation (through effective and appropriate use of the instrument) – a sort of seeing through doing. Secondly, this interaction fosters understanding – of the model, the simulation, and hopefully the simuland – learning through doing. As an experimental object, simulation has the capacity to ‘black box’, to perform as a (semi-)unpredictable entity, while at the same time functioning as the measuring instrument for that entity. Then the black box may be opened, and the mechanisms of the ‘experiment’ can be examined in a theoretical setting. Understood in this context, it can therefore be seen how simulationists have both rapport with their simulations and understanding from them. Furthermore, it is this relationship between man and machine (scientist and instrument) that borrows from experimental tradition. The practice of doing simulation looks similar to the seeing and learning through doing that characterises the practice of experiment.

195 Galison, 1997, p. 738. Galison presents his commonalities as fundamentally epistemic (Galison, 1997, p. 738), rather than methodological, and includes with them stability and replicability (though with caveats to the difficulty of, say, reproducing localised results). However, with regards to the point made above, in the sense that error tracking, particularly, is also enacted, the practice has a methodological as well as an epistemic component. Baldly, when it comes to error tracking, being a simulationist is ‘like’ being an experimenter.
This positioning can help explain why simulation-as-experiment is such an alluring idea. The understanding of a phenomena gained from an experiment has always had epistemic primacy to that gained from simulation, but it is undoubtedly clear that simulation does grant understanding of its target systems, and in a very persuasive way. Though understanding is a philosophically tricky concept, it can be characterised, especially in simulation, as something that involves a subject – a scientist. If explanation is the relationship between an explanans (a theory or model) and an explanandum (a phenomenon or simuland), understanding is the work of the scientist in making sense of the connection between the two. As such, “gaining understanding through explanation is not an automatic process, but rather a cognitive achievement in its own right”, earned by the scientist. Lenhard characterises understanding for experiment as usually having some connection to materiality: the experimenter understands the physical system through the manipulation of it. For theory, on the other hand, understanding is connected to intelligibility: “the long-standing epistemological promise of theory and theoretical models is that they can be known and understood perfectly because they are completely constructed.” Lenhard goes on to argue that the complexity of simulation undermines this thesis by providing understanding even when theoretical terms are imprecise. Importantly, in simulation “an action-oriented understanding is constructed that makes it possible to exercise control and make predictions”; simulation generates abilities in the scientist that make understanding possible. Lenhard argues that simulation allows for a new mode of understanding that makes epistemic opacity and understanding partially compatible, replacing the intelligibility of theory with the quasi-empirical access of simulation. ‘Quasi-empirical’, because in this mode, which “employs exploration, visualisation, and feedback”, provides orientation by sounding out the (virtual) territory.

Simulation, then, provides a mode of understanding that looks a lot less like understanding gained from theory than it ought to, especially given simulation's modelling bias. In fact this mode of understanding – exploration, visual outcomes, and feedback from the system – has very much in common with understanding gained through experiment. Like in the previous section where simulation could present itself as a technique continuous with the theoretical-modelling tradition, in the relationship with its user simulation borrows credibility from the experimental tradition by analogous practice. In addition to the similarities in methodology – error tracking as a form of validation (to complement model validation, perhaps), increasing precision as a goal, systematic variation to test the response of a (programmed) system – simulation allows its user to experience a sense of deep familiarity with their ‘instrument’. Such rapport fosters the understanding of both the simulation and the target system. This, of course, places a limit on the experiment-simulation analogy, because the 'target system' is part of the simulation. The difficulty arises through the very dual nature that allows simulationists to look like

---

200 de Regt, Leonelli, & Eigner, 2009, p. 3. See also de Regt, 2009. This is, of course, only one view of understanding. What is important for this argument is only the fact that understanding involves a subject, or 'someone to understand'.
201 de Regt et al., 2009, p. 7. Emphasis in original.
experimenters, in that simulation functions as both 'experiment with' and 'experiment on'. In associating experiment and simulation, it is therefore manipulability and not materiality that is emphasised.

Yet it is also possible to this relationship too far, and experience too much familiarity with your simulation. The rapport with simulation was seductive, some Simulation authors feeling that the activity of the simulationist seemed to move unwarrantably “from manipulating symbols to understand[ing] relationships”,205 since they were still, on the practical level, manipulating symbols. It was this pedantism – that, really, simulationists are just manipulating digital, abstract, and unreal symbols – that worried simulation practitioners. The preoccupation with the validity of the model was actually symptomatic of this worry: don’t be fooled by how quickly and easily simulation gives you results, because if your model is faulty, they may not accurate to the actual world. There are, too, mentions of modeller’s bias, and McLeod warns that simulation, while powerful, “is only a tool that will extend the talents (or otherwise), and the biases, of those using it”.206 Hamming, too, warned of, “the dangers of self-delusion, of finding (or even manufacturing) patterns in what are essentially random results”207 – simulation aided in pattern-recognition, but as Robert Batterman points out, “humans are all too ready to see patterns in just about anything.”208 Here one can see an instance of how the evolution of simulation and its hybrid nature resulted in confusion. The positioning of the simulationist as experimenter indicated that simulation could uncover information about the world, yet at the same time simulation (as model) was nothing but a representation of that world; as one Simulation author cautions: “a model, no matter how complex, is but a representation of reality and should never be confused with it”.209

Turkle’s sociological work also demonstrates the worries over the seductive power of simulations (see section 2.6.2). During the 1980s at the Massachusetts Institute of Technology, architects, chemists, and physicists all felt the potential for simulation to be both a brilliant new technology, and a destructive one. In architecture, the computer aided students in constructing designs on a dynamic and efficient level, where previously they would have had to personally draw their own designs. Architecture software provided more flexibility and a capacity for tinkering, but could also be too easy, tempting students to use pre-set designs and fostering too much of a detachment from the real site.210 In chemistry, teachers were worried that the software was too black-boxed, doing the thinking for the students. On the other hand, “students could see what many of the faculty could not: computational speed and accuracy were quantitative changes that had a dramatic qualitative effect ... the science on the screen began to feel more compelling than any representations that had come before.”211 Physics students described feeling closer to the data

205 Rahe, 1972, p. 16.
206 McLeod, 1973a, p. 11. Emphasis in original.
and theory through simulation, making messy and unpredictable nature look understandable.\textsuperscript{212} Yet once again faculty was worried that the black box of simulation – its opacity – was blurring the line between representation and real. Even the students were aware that in using a computer one lost a bit of the closeness with the actual data that one would otherwise have with pen, paper, and ruler.\textsuperscript{213} The way around this was to make the tool (simulation) more transparent; students needed to have knowledge of the laboratory software at a level that only came from programming it themselves. One faculty member even insisted on complete transparency: “students needed to understand how computers worked, down to the physics of the processor and the graphics screen.”\textsuperscript{214}

To be a good simulationist, one must be close to one’s simulation, close enough that one can interact with the simulation and learn from it in (apparently) the same way that experimenters learn from their instruments. One cannot, however, be too close. Too much familiarity without experience means that students, in particular, may fall into the trap of letting the machine doing the thinking for them. And if the black box is not opened often enough, we are led to concerns about mistaking representation for reality. Turkle’s simulationists (and our own) were struggling with the illusory and contradictory desire identified described by Don Ihde: “I want the transformation that the technology allows, but I want it in such a way that I am basically unaware of its presence. ... Such a desire both secretly rejects what technologies are and overlooks the transformational effects.”\textsuperscript{215} Indeed, despite these warnings, the transformation of science by simulation occurred mostly without controversy. It has been suggested, through the complicated interplay of simulation, model, and ‘experiment’ of the past few sections, that simulation became accepted as a technique through borrowing from established scientific traditions. How, then, did the disparate and piecemeal borrowing of credibility described so far manage to establish simulation in the face of such concerns?

### 2.5 Selling Simulation

By the 1980s, simulation was a widely known and applied technique. This section will describe various strategies that the practitioners themselves used to promote the simulation technique, as well as analysing the more implicit ways in which simulation became established. Through ‘selling simulation’, simulation practitioners linked simulation both explicitly and non-explicitly to existing traditions in science. Not only did this help simulation become established as a legitimate scientific technique, but it also served to link the epistemology of simulation to that of both the traditions of experiment and theory.

During the 1960s and 1970s, \textit{Simulation} authors explicitly discussed the need to ‘sell’ simulation; many authors saw it as the responsibility of simulation practitioners to generate interest and

\textsuperscript{212} Turkle, 2009, p. 28.
\textsuperscript{213} Turkle, 2009, pp. 30-32.
\textsuperscript{214} Turkle, 2009, p. 37.
\textsuperscript{215} Ihde, 1990, p. 75.
awareness of simulation. ‘The Art of Simulation Management’ encouraged simulation users to learn the language of both analogue and digital – it seemed that often the customer spoke ‘digital’, and analogue staff should be able to communicate with them without sounding like ‘foreigners’. Facts must also be distilled and communicated to managers, in such a way that they emphasised simulation’s “utilisation, accessibility, efficiency, and ... modest backlog”.216 It was important, for funding reasons, to promote simulation, even if the skilled simulationist might consider management “a bunch of dimwits who can understand only the simplest and most superficial explanations and displays”.217 Marketing their product seemed to be part of the role of the early simulationists, and some considered it essential. In a report on a joint meeting of the American associations of the Midwestern and Eastern Simulation Councils, ‘much time’ was devoted to the topic of selling simulation, as “the practice of the art of persuasion is crucial to the continued practice of simulation”.218 Techniques of persuasion including displays, movies, offered courses, mock-ups, and other forms of publicity were suggested.219 There was also the inclusion of a new section in Simulation – ‘Simulation Today’, which was intended to introduce students and outside professionals to the non-technical applications of simulation.

There were also attempts to formalise the simulation ‘profession’. The title of McLeod’s piece for one the ‘Simulation Today’ sections in Simulation – ‘Simulation: From Art to Science for Society’220 – is illustrative of this. While the ‘art of simulation’ was often spoken of affectionately, the phrase implied a technique that required skill, but which was still the purview of the lowly technician. The ‘science of simulation’, however, was a much loftier moniker, implying a well-developed technique in the hands of experts. This would also, of course, have the effect of making the knowledge produced with simulation more ‘scientific’. McLeod’s article suggested that simulationists apply scientific principles to encourage standardisation and repeatability, and to present simulation reports using a format that greatly resembled a traditional laboratory or experimental report. Though not stated explicitly, this was clearly an attempt to borrow credibility from the experimental tradition. Treating the use of the technique as a ‘science’ (rhetorically) by using similar principles and reporting was part of the effort to professionalise simulation in the eyes of external viewers. The association behind the publication of Simulation was renamed, in 1972, from ‘Simulation Councils, Inc.’ (SCI) to ‘The Society for Computer Simulation’ (SCS), to demonstrate that the organisation “is a society for professional simulationists working with computers rather than ... informal groups of regional practitioners”.221 Simulationists were aware of the expert nature of their craft, and has been seen many times before, sought credibility as professionals, not practitioners. The association has since been renamed a third time, and is today called ‘The Society for Modelling & Simulation International’, reflecting the close associations between modelling and simulation, as well as

217 Clymer, 1964a, p. 13 Clymer notes that this outlook is sometimes the biggest obstacle to successful persuasion.
218 Clymer, 1964b, p. v.
220 McLeod, 1973b.
221 McKenna, 1972, p. vii.
emphasising the international spread of the technique – no longer are simulationists mere ‘regional practitioners’.

The international flavour – it is always more impressive when an organisation is ‘international’ – was introduced in the early 1970s, through the sudden upsurge of interest in ‘world simulations’. A world simulation tried to model the state of the world with respect to large social problems of public concern like population growth, the depletion of natural resources, the distribution of wealth, and pollution. In 1969 the World Simulation Organisation, “which originated as a special Task Force of SCI”, was formed in an attempt to facilitate communication between various groups attempting to create a world simulation; McLeod was excited about this concrete evidence of the international scope of simulation work. In 1972 world simulations were a particularly hot-button topic following the publication of a report by a group called the ‘Club of Rome’ entitled *The Limits to Growth*. The Club of Rome was a group of intellectuals from academia and industry who aimed to study and discuss global economics and their political import. Their report detailed a computer model of economic and population growth versus natural resources, and it seems to have had a very gloomy outlook. When McLeod introduced the book to the Board of Directors behind *Simulation*, there was much excitement, “with the result that all members of both the Board of Directors and the Executive Committee asked Headquarters to order them their own copies!”.

*The Limits to Growth* (and similar world simulations) excited more than just simulationists – the book attracted attention from economists, scientists, and political figures. This is entirely unsurprising – human interest always sells – and so it is also unsurprising that McLeod used such rhetoric to his advantage. He stated that man – through overbreeding, pollution, and depletion of natural resources – had the power to destroy himself, and as such must find a way to control his resources before nature did it for him: “action must be taken, but understanding it a requisite for prudent action. That’s where simulation comes in.” The attention garnered by the Club of Rome publication did more for selling simulation and making it relevant than any series of seminars, at least to the general public.

Though McLeod was pleased about the attention world simulations brought to the technique of simulation, he was also worried that the inadequacy of the Club of Rome model would in fact harm simulation; he was “afraid that the publicity attendant on the publication of shocking conclusions drawn from simulation experiments run with an admittedly inadequate model will

---

222 McKenna, 1972, p. viii.
223 McLeod, 1972a, p. 4. The fate of this organisation is unclear, but it appears that the project was abandoned soon after its conception, the cause of death being underfunding (McLeod, 1972b, p. 7).
224 The Club of Rome is still active today (see http://www.clubofrome.org), and published updates to *The Limits to Growth* as recently as 2004.
225 McKenna, 1972, p. viii.
226 Bardi, 2011, p. 11.
227 McLeod, 1972b, p. 7.
'set simulation back 20 years’”. Perhaps these events are part of what triggered the growing preoccupation with the modelling aspect of simulation in the 1970s. By 1980, he admitted that when it came to policy and decision makers, in many cases simulation had in fact been oversold. It would be necessary to remind people of what simulation should and should not be expected to do, lest the technique be blamed for the model’s failure. In the enthusiasm of giving simulation the potential to help us help ourselves, McLeod was overreaching, though one author rather astutely notes that the future simulation user, “recognizes in fact that frequently the results of the simulation will not actually be adopted in practice but will serve as ammunition in political battles fought over more general but extremely vital issues”. In fact, McLeod need not have worried. The Limits to Growth was the subject of much debate and publication, but criticisms of the book and its conclusions centred mainly around the conceptual model, not the use of simulation in deploying the model. Essentially selling simulation involved advertising simulation to both customers and management as a profitable and worthwhile product. These metaphors are not used accidentally; they reflect the tone of the discussion in Simulation – one article used economic language (liberally sprinkled with other metaphors) to urge readers to plan a long campaign of persuasion that would convince all levels of management and all prospective customers, ensuring “you will slowly but surely penetrate your entire market”. Simulation was, in fact, a ‘new market technology’, to employ the term Thomas Nickles has borrowed from business analysts. Nickles’ main contention is that science is more prone to dramatic disruption from changes in practice than in representation; scientists are remarkably adaptable in the face of theory change, an observation at odds with the idea of the scientific ‘revolution’. Adapting business analysis ideas of sustaining and disrupting technologies, Nickles states that “major technical innovation (whether theoretical or experimental) is neither sufficient nor necessary for disruption of practices”. Disrupting technologies often come from areas external to the science they impact upon, and are classified into two kinds. New market disruptions are technologies that appeal to a non-existing market, and so compete against ‘nonconsumption’, rather than similar technologies. This is precisely what the early sellers of simulation were doing – attempting to convince a customer base that they had a need for simulation, that it was a viable and effective way of solving their problems. The other type of disruption are low-market or low-end disruptions, that appeal to customers at the low value end of the market, who are ‘overserved’; business examples are discount airlines and cheaply priced department stores. Nickles’ example of a low-end technology are computers, which were initially used for tedious calculations, but with the increase in speed and power have moved ‘up market’, “transforming some fields and creating others. [E.g.] High-

---

228 McLeod, 1972b, p. 7. It is not clear from the context of the paragraph in which this quote appears whether McLeod is paraphrasing, summarising, emphasising or quoting someone else (uncited) when he uses quotation marks for the phrase ‘set simulation back 20 years’.


230 Karplus, 1972, p. 23. One of course thinks of today’s climate change debates.

231 Bardi, 2011, pp. 11-12.


speed computers are essential to modelling across many fields, and, increasingly, computer simulations are a useful substitute for nature in certain kinds of experimental work.”

If simulation can be classified as disruptive, why is the inception of the technology not given greater focus and analysis in the literature? Nickles identifies the invisibility of these disruptions as the fault of general philosophy of science – the focus on major results and theory shifts disguises the fact that “theory disruptions are not always in phase with disruptions of practice”. One is reminded here of Galison’s argument that theory, experiment, and instrumentation are intercalated, the rejection of the assumption of co-periodisation that is inherited from the Kuhnian/Lakatosian tradition. The focus on models that begins for simulationists in the 1970s may well have distracted the philosopher preoccupied by theoretical change. Naturally, the building and application of models is an important part of the scientific process, and it has been emphasised so far in this thesis as an important part of simulation too, but so are the experimental and social aspects. Nickles’ scientists are “risk-taking pragmatic opportunists” that will pursue a theory or technique if they consider the problem-solving potential to be high, sometimes valuing heuristic over epistemic appraisal:

A creative scientist who regards a theoretical model as an active research site treats it as the theoretical analogue of an experimental system ... as a research tool and as itself an object of research, a system to be tinkered with in order to see what interesting results it might yield.

Simulation, in one of its roles, allows precisely this sort of creative research and tinkering to be performed on theoretical models. Coupled with its heuristic potential – simulation is always trumpeted as something that expands our epistemic horizons, making the intractable tractable – it is no wonder simulation has gained such widespread use. It is possible that this aspect of simulation’s acceptance has been obscured by the traditional philosophical bias towards theory as “an aesthetic object to be admired as filling in part of our big picture of the universe”, rather than as a dynamic and continually changing aspect of research. Simulation reveals such dynamics, and so at first blush seems like an odd combination of theory and experiment.

This explanation based on the traditional philosophical image of science has merit, but it may be that the relative invisibility of the adoption of the simulation technique has a historical explanation. Like Galison and Nickles, Rasmussen also emphasises the continuity of practice in his history of the electron microscope and its use in biology from 1940 to 1960: “very little of established knowledge or of scientific ways of life will need to be abandoned immediately on introduction of a novel instrument.”

---

236 For example see Galison, 1997, p. 799, figure 9.5.
240 Nickles, 2008, p. 357
forces that govern the acceptance of a novel technique (or theoretical claim, though Rasmussen's focus is on instruments and their accompanying techniques). A technique is more likely to be accepted if it doesn’t challenge the existing theories and habits of the discipline concerned: “the domestication of the electron microscope (and by extension, instruments in general) was achieved when a set of practices was developed that fitted and integrated, in minimally challenging and disruptive ways, the new imaging machine and its pictures with the existing knowledge and ways of life prevailing”.  

This seems contradictory to Nickles: on one hand simulation is a disruptive technology, but to be accepted as a technique, according to Rasmussen, simulation must be minimally disruptive. The two perspectives are, however, complementary rather than contradictory: retrospectively (or perhaps diachronically), simulation is a disruptive technology – the introduction of simulation has radically changed the practice of certain areas of science. Contemporaneously (or perhaps synchronically), however, simulation had to become an embedded technique and in order to do this needed to incorporate nondisruptively into existing techniques and habits. Simulation looks experimental through the way that it sets up the relationship between the practitioner and their simulation – the ubiquity of experiment metaphors in simulation literature is indicative of this (see Chapter 3). Simulation also has strong ties to modelling, and validating a simulation looks very much like validating a theoretical model. Through linking itself to these two previously existing and well-established methods of scientific practice, simulation has developed a continuity of practice that is nondisruptive by definition.

The work of unconscious adoption done by borrowing credibility from existing traditions was complemented by more deliberate work done by the simulation practitioners, as "the task of minimising cognitive dissonance is paralleled by a task on the level of social engineering, that is, the task of nondisruptively introducing new work practices and the social relations that go with them to established laboratories". In a section coincidently entitled ‘Selling the Electron’, Rasmussen outlines this social engineering for the electron microscope. A National Research Council (NRC) was formed that had control over one of the only electron microscopes available to biologists during the war, and it consisted of ‘big names’, members calculated to have clout in order to generate credibility for the use of the electron microscope in experimentation. This committee exercised an almost militant degree of control over the microscope – ensuring that only the ‘right sort’ of researcher had access, and restricting the sorts of images and research that could get published. Though this was a very contingent situation – “the disciplining work was certainly facilitated by the heightened wartime influence of elite advisory bodies in general” – there are several parallels with Simulation’s efforts on behalf of the simulation technique.

---

243 Rasmussen, 1997, p. 29
244 Rasmussen, 1997, pp. 46-69.
Biologists using electron microscopes were called on “to make the publicity drive a form of ‘missionary work’ by showing their best micrographs at every scientific congress they might attend”. This explicit selling sought to legitimise the pictures produced by electron microscopes by demonstrating that image production and interpretation had been subject to rigorous standardisation, so that “embarrassing retractions would be prevented”. McLeod likewise suggested multiple ways in which simulationists should check their models, having had similar concerns during the debate around The Limits of Growth that one poorly implemented model might reflect badly on the simulation technique. In order to ‘sell’ simulation, it had to be demonstrated that these epistemological concerns had been allayed, and this was done by validating the new technique by borrowing the credibility of an existing technique.

For simulation, pushing for the appropriate and thorough validation of models was a way of integrating the simulation technique into theoretical practice, and thus establishing its reliability by linking it inextricably with the centuries old technique of modelling. This link with the theoretical tradition, viewed by many as a technological continuation of modelling practice, is arguably the strongest instance of borrowing. Simulation is, however, peculiar: it doesn’t just borrow credibility from the theoretical tradition, but from the experimental as well. The previous section made an argument for understanding how the simulationist could be positioned as an experimenter with respect to his instrument, borrowing from the experimental tradition through analogous practice. McLeod’s efforts at a standardisation of reporting, briefly mentioned earlier in the context of attempts at professionalisation, are another, somewhat more conscious, instance borrowing experiment’s credibility. McLeod urged his readers to develop a way of documenting a model and its simulation that was standardised enough to allow the reporter’s peers to evaluate their work, “others to repeat the experiment”, and others to build on the work instead of having to start over, developing individualised models that were not conducive to adaptation elsewhere. The suggested outline for such a report contained six sections (each with several sub (and sometimes sub-sub) sections): project information, model development information, description of model, simulation(s), discussion, and literature (references and bibliography). This is essentially the standard layout for a (laboratory) experiment report – aim, method, results, and discussion – and indeed, the goals of McLeod’s simulation report are much the same as the reasoning behind writing lab reports. The casual reference of the process of simulation as an ‘experiment’ further cements this, situating simulation within experimental practice just as electron microscopy was.

It is therefore possible to understand the adoption of the simulation technique, in a general sense, as a nondisruptive inculcation into scientific practice through its borrowing from the traditional areas of scientific activity, namely experiment and theory. Within Rasmussen’s framework, one can see how the conservative forces of theoretical modelling and experimental standardisation (and analogous practice) link simulation to existing epistemological practice, and therefore

---

247 Rasmussen, 1997, p. 56.
248 McLeod, 1973b, p. 77.
smooth its introduction into daily scientific practice. The process of embedding simulation was both explicit – when it came to ‘selling’ simulation – and non-explicit. Clymer, in 1964, told the readers of *Simulation* that the art of simulation not only included the art of public relations and sales principles, but also the art of seduction. This latter art was described as:

The problem is to transform opposition into cooperation, by imperceptible degrees. Fears and suspicions must be allayed, defences must be overcome, a succession of new objectives must be sold, reluctance must be replaced by enthusiasm, all the senses and areas of sensitivity must be played upon in sequence and in parallel.

As practitioners consciously worked on persuading both management and members of the scientific community, so did simulation unconsciously establish credibility as a technique ‘by imperceptible degrees’. Instead of weakening the potential legitimacy of the technique, finding one foot in each camp actually helped strengthen simulation’s cause.

2.6 Supersimulations and Seduction

2.6.1 Simulating as Science

The introspective pieces from *Simulation* dry up after the late 1980s, something which is at least in part due to McLeod’s retirement in 1974 (though he continued to contribute to the journal until late 1980). However, there is reason to believe that by the end of the 1980s, simulation was sufficiently established as a technique that the authors of the journal felt less of a conscious need to sell simulation. By this point, simulation had become an accepted scientific technique whose epistemic value had been established. It comes as a surprise, then, when one finds a certain amount of rationalisation, for want of a better word, for using simulation as a technique among papers published concerning simulations run on supercomputers, or ‘supersimulations’, if one will forgive the neologism. To return, briefly, to the disciplinary focus of this thesis, consider the following excerpts from recent astrophysical papers describing large simulations:

Numerical simulations of three-dimensional self-gravitating fluids have become an indispensable tool in cosmology. They are now routinely used to study the non-linear gravitational clustering of dark matter, the formation of clusters of galaxies, the interactions of isolated galaxies, and the evolution of the intergalactic gas. Without numerical techniques the immense progress made in these fields would have been nearly

249 Küppers, Lenhard, and Shinn come to a similar conclusion, though through a very different sociocognitive framework: “Computer simulation can thus amply penetrate a host of application domains without jeopardizing them, and this constitutes one of the pillars of its universal success” (Küppers et al., 2006, p. 19). See Küppers et al., 2006, pp. 17-21.

250 Clymer, 1964b, p. v.

251 The exception to this is Taylor et al., 2015, which describes itself as “one of several concurrent activities aimed at reinvigorating the debate on grand challenges in M&S [modelling and simulating]” (Taylor et al., 2015, p. 648).
impossible, since analytic calculations are often restricted to idealised problems of high symmetry, or to approximate treatments of inherently nonlinear problems.\textsuperscript{252}

Without numerical simulations, the $\Lambda$ cold dark matter (CDM) model may arguably not have developed into the leading theoretical paradigm for structure formation which it is today. This is because direct simulation is often the only available tool to compute accurate theoretical predictions in the highly non-linear regime of gravitational dynamics and hydrodynamics. This is particularly true for the hierarchical structure formation process with its inherently complex geometry and three-dimensional (3D) dynamics.\textsuperscript{253}

While the initial, linear growth of density perturbations can be calculated analytically, the collapse of fluctuations and the subsequent hierarchical build-up of structure is a highly nonlinear process that is accessible only through direct numerical simulation.\textsuperscript{254}

These simulations are essential to investigate highly non-linear scales such as the halos of individual galaxies and galaxy groupings, where simple analytical approximations fail.\textsuperscript{255}

It is clear that purely analytical approaches have arrived at the limits of their reach. Fuelled by continuing advances in numerical methods and computational capabilities, the future of structure formation and galaxy formation theory is going to be led by numerical simulations.\textsuperscript{256}

In order to understand how galaxies form and evolve in their cosmological context, we must understand the properties of dark matter haloes over a wide range of physical scales and across virtually all of cosmic history. Numerical simulations provide one of the best methods for approaching this problem and have proven invaluable for studying the growth of cosmological structure and, in particular, of dark matter halos. Increasing computational power and improved algorithms have led to a steady and rapid increase in the ability of N-body simulations to resolve the detailed internal structure of dark matter halos over substantial cosmological volumes.\textsuperscript{257}

Two things are of interest here: first, there seems to be a regression to the earlier theme of the physical computer, though in a modern guise. Almost all the excerpts emphasise the non-linearity of the mathematics involved in cosmological simulations, and that these calculations would be intractable without the use of computers. There are repeated references to the increased computational power available, and the improvement in algorithms or numerical references. Were the authors not referring to state-of-the-art supercomputers, these excerpts could be dated in the 1960s, instead of the 2000s – like previously, the focus is on the physical instantiation of the technique and the power of the physical computer, and on the development of efficient ways

\textsuperscript{252} Springel et al., 2001, p. 79.
\textsuperscript{253} Springel, 2005, p. 1105.
\textsuperscript{254} Springel et al., 2005, p. 629.
\textsuperscript{255} Gao et al., 2012, p. 2170.
\textsuperscript{256} Kuhlen et al., 2012, p. 51.
\textsuperscript{257} Boylan-Kolchin et al., 2009, p. 1150.
of using the technique (what was once more specifically new programming languages, here it is the more vague 'numerical method' and 'improved algorithms'). The site is even similar because the supercomputers that are powerful enough to run complex simulations are, like the early computers, rare and belong only to certain institutions. The same restrictions of physical accessibility apply to these computers as they did to the earlier versions in simulation facilities. Even the operation of the (super)simulationist has similarities, as these large simulations involve teams of scientists with different specialities.\textsuperscript{258}

The second point of interest is that the authors of these supersimulations all seem to think it necessary to put a paragraph into the introduction of their papers that serves no other purpose than to describe the fundamental indispensability of simulation to their research. This might seem quite bizarre, because one expects this sort of justification or reassurance of a new technique where there might be question of its appropriateness, but simulation seemed to have popularised itself long since. In fact, both this point and the previous one can be attributed to the same cause, which is the introduction of a new aspect of an existing technique. Supercomputers represented a new phase in computing technology, broadly speaking, because of the sheer scale they provided. They literally made possible the construction of models for systems that were enormously complex, and thus mathematically intractable, just as the first computers made possible the modelling of systems that involved calculations too tedious or large for unaugmented human skills. Hence the reversion to the earlier theme of physical computer in the extracts above, and hence the emphasis on answering 'why simulation', which seems to serve no other rhetorical purpose.

Supersimulations are thus fitted into the continuing evolution of the simulation technique as another tool at the disposal of the simulationist. In Nickles’ terminology, supersimulations are a sustaining technology, “a successful technology that attempts to keep its established customer base happy by introducing improvements to popular product lines”.\textsuperscript{259} Supersimulations are also a good way to see how the themes of physicality and simulation-as-model combine. The modern simulation is emphasised as an instrument that has the capacity to assist scientists in exploring unknown (virtual) territories. It is modelling improved, introducing the ability to dynamically manipulate theoretical ideas in an unprecedented manner. With the increase in computational power, simulation has a greater role to play in generating knowledge. There is continuity with previous themes, but the overwhelming sense of improvement on previous themes. One may say that the theme of this period is that of power – in computing, and in epistemic access.

It is not the existence of supersimulations that makes power the theme of the modern period, though their enormous capacity does give the technique greater reach. In a more general context, every institution that generates research using simulations is in possession of a computing set-up that is on a much smaller scale, and yet provides sufficient power to run effective simulations. With the increase in availability and the decrease in cost, reasonably powerful computers able to run sufficiently complicated simulations have become available for most researchers. Even a

\textsuperscript{258} For an ethnographic study of a modern-day applied computation group, see Spencer, 2012.
\textsuperscript{259} Nickles, 2008, p. 370.
standard personal computer can cope reasonably well. This ‘scale down’ means that the power of the simulation is now accessible to all scientists on a day-to-day basis, increasing the epistemic access of more researchers. The advent of the Internet has also made computational information more readily communicable. A recent (2015) article in *Simulation* on the ‘grand challenges’ to modelling and simulation urges even *more* access: “for simulation to fulfil its promise as a third branch of science, together with deductive and empirical methods, it must become much more widely deployed, implemented, understood and used.” The authors of the article suggest that simulation become an aspect of ‘e-Science’, with models and values available as part of a distributed system rather than presented in the limited and inaccessible manner of a research paper. ‘Big Simulation’ involving large amounts of data or continuous simulations, co-ordinated modelling with multiple authors, and ‘cloud-based’ simulation are all areas where these *Simulation* authors see growth for simulation in future.

With the development of personal computing came an alteration in the way that simulationists operated – as the physical size of the computer shrunk, so the role of the simulationist expanded. Initially, a simulation facility required several experts doing different jobs to be kept running properly. The problem originators, who received and analysed the results of the simulation, were separate from those who programmed and ran the simulation. Later, there was an increased emphasis in the role of the latter for having appropriate knowledge of the simuland, and though this knowledge was generally communicated from the client, the simulationist needed to understand enough to be able to construct the model himself. Once the technology had evolved sufficiently, however, for the first time the client was in a position to construct the simulation themselves, leading to a shift in the meaning of simulationist, a conflation of, in the terms of the early simulation period, the problem originator and the simulation engineer (or programmer).

With the arrival of the PC, as Johnson and Lenhard emphasise:

> Computational research became better integrated to both theory and experiment and evolved into an indispensable dimension of the scientific enterprise, as opposed to the kind of computation that was common in the 1950s through 1980s, which was a more independent branch of investigation. Now researchers work on computational models, whether they consider themselves experimentalists or theorists. … With the mature desktop computer, model exploration becomes a key practice in scientific research, and not one relegated to computational scientists, but rather one that is commonplace among most scientists and engineers.\(^{261}\)

In a report of a 1989 conference for which the theme was ‘Grand Challenges to Computational Science’, the author noted that “few of the participants viewed themselves as computer scientists. The majority were specialists in various scientific and engineering disciplines who use supercomputers.”\(^ {262}\) These disciplines included biology, medicine, weather modelling, quantum

\(^{260}\) Taylor et al., 2015, p. 649.


\(^{262}\) Leven, 1989, p. 1456.
chemistry, materials design, relativistic physics, and cosmology and astrophysics. Though the conference topics were heavily computational, most of the participants would call themselves scientists first, though their practice of science had more in common with those in computational disciplines. One attendee announced plans to teach the sophisticated skills required for supercomputing to chemists, biologists, physicians, and engineers, as practitioners in these disciplines suddenly had a need for such skills. For the conference attendees there were now three branches of science: experimentation, theory, and numerical simulation, the latter including tools such as “visualisation, pattern recognition, friendly human interfaces, and expert systems.” The branches, naturally, all complemented each other, and, more importantly, were on equal levels – simulation is, almost obviously, part of modern science.

Like Galison and Nickles, I think that the continuity of a scientific technique has a lot to tell us. The previous sections have seen the simulationist began as engineer, evolved into programmer, became a modeller, worked like an experimenter, and finally assumed the mantle of problem originator. Through this chapter’s themes, ‘the simulationist’ has become the modern scientist. From Turkle’s sociological perspective, “in the 1980s, scientists were comforted by the idea that building simulations was the province of computer scientists and that they were in a very different field. Not too much time would pass before this division became elusive. More and more, science was done on the computer and a migration of computer scientists to the natural sciences would soon make it hard to say where one left off and the other began.” As the use of simulation spread, knowing how to simulate became as important to scientists as knowing calculus. Today, doctoral students are taught how to use simulation appropriately for their subject areas as an integral part of their training. There still exists a differentiation between being a ‘user’ or a ‘developer’ of simulation code – users approach codes as black boxed, while developers write, edit, or otherwise try to understand the code – but this (often unclear) boundary occurs within a scientific discipline. Unlike the early simulation of section 2.2, it is the problem originator that is doing the simulating, regardless of whether ‘simulating’ refers to writing the program from the ground up (by oneself or in a ‘code collective’), editing someone else’s code to suit one’s problem, or interacting only with a general user interface to adjust parameters. Even in areas of modern science that have approximately equal doses of experiment, theory, and simulating, many of the

---

263 The report notes the attempts to draw a distinction between “computational science” and “computer science.” Take the following sequence (of a simulation, one assumes): physical system ↔ mathematical model ↔ algorithm ↔ numerical implementation. Computational science has more to do with the physical system and mathematical model, and computer science deals with the algorithm and numerical implementation (Leven, 1989, p. 1457). Perhaps most conference attendees felt they were not computer scientists because their focus was not the latter, more pragmatic, activity.

264 Leven, 1989, p. 1456.

265 Leven, 1989, p. 1456.

266 Turkle, 2009, p. 42. Of course this process begins earlier – astrophysics began simulating spiral galaxies in the late 1960s (see Chapter 3), and Park writes that “in the years between 1955 and 1965 … many quantum chemists came to be their own programmers” (Park, 2003, p. 247). At the same time the process of turning an entire discipline into scientists and simulators would have taken several years.

267 Sundberg, 2010a, pp. 275-277.

268 Sundberg, 2010b. See also Sundberg, 2010a.
practicing scientists have some sort of simulation skill – as early as 1964, Galison relates that for a typical particle physics experiment, “programming was occupying 10%-20% of a spark chamber group’s work”.\footnote{Galison, 1997, p. 491.} With the increased availability of simulation-capable computers, and workshops and classes aimed at teaching scientists computational skills, Keller observes the same integration of programmer and biologist (and mathematician):

In these new settings, mathematicians and computer scientists do not simply collaborate with experimental biologists, they become practicing biologists. Conversely, and thanks in part to the rise in computer literacy and in part to the development of user-friendly computer programs that bring the techniques of mathematical analysis within the grasp of those with little or no conventional training in the subject, biologists need no longer simply hand over their questions and data to others; now, they can build (either by themselves or as active participants) their own mathematical/theoretical models.\footnote{Keller, 2002, p. 258.}

Furthermore, one might suggest that there is now a new breed of scientist that is \textit{nothing but} the simulationist. The emphasis on the absolute necessity of simulations in the excerpts at the start of this section is an example of this: galaxy formation studies (in this case closely tied to dark matter studies) are an area where simulations are described as ‘indispensable’, ‘essential’, and ‘invaluable’, without which the leading theoretical paradigm might not have been developed, and which will lead the future of the discipline. The day to day practice of these scientists is programming, running, and analysing simulations; simulating is their main way of doing science.

In astrophysics, “computer simulation is currently viewed as the sole acceptable path for exploring a complex universe.”\footnote{Küppers et al., 2006, p. 4.} There are other examples: Johnson claims that “there was no nanotechnology before digital computers”,\footnote{Johnson, 2006a, p. 41.} arguing that science develops not only through the interactions between theory and experiment, but also those of instruments, simulations, and manufacturing. The latter implies the making of products, an element of design that means nanotechnology, in particular, straddles the gap between science and engineering,\footnote{Johnson, 2006a, p. 49.} perhaps a modern form of the early tension between analogue engineer and digital scientist. Increasingly, the social and biological sciences make heavy use of simulation – in evolutionary biology, “a whole field of biology deals only with computer simulations: so-called Artificial Life”,\footnote{Huneman, 2014, p. 61.} an area that attempts to model the components of a real evolutionary growth. We also have climate science, where “computer models are absolutely essential in the efforts of atmospheric scientists to represent the earth’s climate and its possible evolution.”\footnote{Norton & Suppe, 2001, p. 67.} These disciplines have in common objects of study that are unmanipulable and inaccessible; their only tool of analysis is computer simulations, forming the role of both the instrument and the experimental system. Having
borrowed enough credibility from well-embedded scientific traditions to establish whole disciplines and sub-disciplines where the main scientific activity is simulating, 'simulationist' must fit somewhere into the role of the modern scientist for many areas of science.  

2.6.2 Seduction

The acceptance of simulation into scientific practice is not, however, without its problems. Turkle finds a continuing tension in simulation in architecture into the millennium. Some designers object to what they see to be a lack of control over their designs with the ubiquitous use of modelling programs. On the other hand, architects close to their programs feel they have more control over their designs, one subject feeling that she was “developing something she experiences as a body knowledge of [the model's] contours”. This is similar to the rapport experienced by the engineers of Simulation, but to a magnified degree. Modern simulation has such power that “it can sometimes edge out the real”. Turkle's architects describing how they can lose reference outside of the simulation. Though this thought is alarming, simulation's ubiquity might erase such worries: “the well-worn or best-loved virtual starts to take on some of the qualities of the real. It feels familiar, comfortable; it is able to assuage anxieties about being cut off from nature.” Intrinslic belief in simulations may even be a side effect of the conscious need to sell simulation; climate modellers, in one sociological study, described their models as 'truth machines' and downplayed uncertainty in communications for 'outside' of the science community, but through such overselling may lose their critical distance from the model. Again, the fear of too much closeness crops up, this time through emotional attachment and time invested in a model, to the extent that “modeller's careers and identities become intertwined with, and partly dependent on, the quality of their models, to the point that they sometimes may be tempted to deny it when their models diverge from reality.”

In 1977, Simulation published a Silver Edition containing reflections from practitioners of the past twenty or so years. One article – described as “a philosophical overview of simulation” – had a stern message: “keep everpresent in your mind the difference between model and reality.” There were similar cautions in when discussing the man-and-machine theme, and the possibility of having too much rapport with one's simulation. Hayles' actors in cybernetics struggled with objectivity in the face of reflexivity, of losing the observer among the system: “the fear is that

---

276 One may even say that simulation has reached the point of self-authentication, in common with Hacking’s styles of reasoning. Simulation as a style of reasoning has been suggested by Schweber and Wächter, where it is associated with the “complex systems computer modelling and simulation” (Schweber & Wächter, 2000, p. 585) Hacking-type revolution that we are witnessing now (see section 5.4.1).

277 Turkle, 2009, p. 50.


280 Lahsen, 2005.


282 Bekey, 1977, p. 161. For reference, the author was a professor of electrical engineering and computer science with a background in industry.

283 Bekey, 1977, p. 163.
under these conditions, reliable quantification becomes elusive or impossible and science slips into subjectivity, which to many [scientists] meant that it was not real science at all." Half a century later, scientists continue to worry that the warnings will not be heeded, and that they themselves will be seduced, for "with the computer on hand to deliver the real, simulation can seem world enough." Answering the question posed in the title of her essay ('what does simulation want?'), Turkle makes the statement: "we have seen what simulation seems to want – through our immersion, to propose itself as proxy for the real." The danger arises when this proxy supplants the real. One of Turkle’s architects relates a story where a building was designed in a program to have a limestone texture on the façade. An order was placed for ‘simulated’ limestone (limestone surfacing etched on fiberglass), and cut-out dormers, as these were the cheaper options. On screen, this looked acceptable, "but when built, the hotel provoked an outcry. Critics described it as worthy only of Disneyland. The physical building looked like a simulation, out of place surrounded by ‘real’ buildings."

Turkle here draws on Jean Baudrillard’s critical analysis. Baudrillard sees Disneyland – a place full of simulated buildings and people – as “a deterrence machine set up in order to rejuvenate the fiction of the real in the opposite camp”. It is intended to hide the fact that the entirety of America is as simulated as Disneyland, a coded dissimilarity, like the hotel façade making the buildings beside it look real by contrast. Baudrillard’s understanding of ‘simulacra’ (in its modern form) deals less with simulations, per se, and more with the relationship between reality and representation, which is precisely what both concerned Turkle’s scientists and Simulation’s authors from the 1970s and 1980s. In the world of computers and virtuality, says Baudrillard, we have lost our previous distinction between representation and its target – no longer is there a difference between the sign and the real as there was in the modernist era; it has since been erased by the endless reproduction of the production era in which the original was lost. ‘Simulation’ (as simulacrum) reigns over our modern society – the system tries to recapture the real by deliberately introducing, “coded similarities and dissimilarities”, yet these produce “only a simulated real, an effect of reality”. This engendered real in turn precedes and determines the sign; it is the precession of simulacra, where the ‘map’ (simulation) precedes the ‘territory’ (reality). This order is characterised as hyperreality.

Writing in the 1980s, Baudrillard shared (with more vehemence) the scientific concerns that use of the computer would, through either laziness or complexity, prevent students (and even their

284 Hayles, 1999, p. 69. Hayles’ form of the fear of seduction is ‘cybernetic anxiety’, wherein the human/machine distinction is not only erased but both sides of the boundary are reconfigured (Hayles, 1999, chapter 4).
285 Turkle, 2009, p. 70.
286 Turkle, 2009, p. 80.
287 Turkle, 2009, p. 52.
289 To avoid confusion, I will refer to ‘simulation’ when I mean computer simulation, and ‘simulacra’ or ‘simulacrum’ when describing the postmodern understanding of a substitute real.
teachers) from knowing what was ‘really’ happening. Baudrillard’s work characterises the fear that one occasionally sees in both historical and modern simulation writing – a postmodern worry that in the fervent search to understand all parts of the world, especially those we cannot access, we have somehow lost touch with the very reality we seek. This fear will continue to crop up in other chapters of the thesis, but it is important to note that despite Baudrillard’s pessimism and the scientists’ worries, and though the technique might have seduced its way into ubiquity, simulation’s representations do not take priority over the real. In the architect’s anecdote, computer simulation introduced the Disneyland building – a coded dissimilarity – into the real world. Through it, the other buildings looked real, and while in architecture this may be a mistake, in other matters of science such a contrast might be welcome. Building metaphors aside, Turkle’s actual conclusion is a positive one: “perhaps we could say, with no irony, it is what simulation really wants – not to replace the real but to reveal it.”

In the astrophysical case studies, it is precisely by constructing a false reality that the real one is understood (see especially Chapter 4).

Simulation, with its complicated relationship to the real, has a tendency to confuse traditional understandings of science. For a society that conducts an increasingly large proportion of its interaction on virtual mediums, the use of computers in science has become quite invisible for the modern scientist. When it comes to simulation, virtual modelling has become less and less problematic since the 1980s. There is a balancing act between the increasingly virtual methods of modern science – especially those that depend on simulation – and maintaining enough of a connection to reality. One might venture to suggest that part of these worries stem from the sui generis parts of simulation, where it looks like foreign science. Other traditions do not have quite the same representation concerns that simulation does (though of course, they have their own) – experiment is too close to reality to seduce, and theory too far from it. Simulation, however, is in precisely the wrong place, a reverse-Goldilocks of a technique. Yet, as this chapter has so far demonstrated, simulation is not entirely foreign science, but a blend of experiment and theory that while having its own way of learning, is also still derived from existing ways of learning.

2.7 Characterising Simulation

This chapter attempts to follow several strands of argument that emerge from taking a historicised look at simulation. In general, it aims to understand how simulation is characterised and how this colours our perception of the technique first through using the lens of the journal Simulation, and then in the last section from the perspective of modern simulation. Simulation has sometimes been implicitly treated as an ahistorical technique, but even such a narrow perspective as a single journal’s history shows that the face of simulation is not a constant thing. Looking at the tone of the articles in Simulation, roughly four distinct themes have been identified, each representative of an aspect of simulation that has been at one time or another the focus of its practitioners. The first theme, that of physicality, shows that during the 1960s simulation was understood mostly in terms of the computer on which it was instantiated. In the 1970s, the focus

293 Turkle, 2009, p. 80.
changed to the model of or in the simulation, and particularly the importance of validating the model in order to validate the simulation. Throughout these periods, and extending into the 1980s, the theme of rapport between man and machine emphasises the relationship between the human component and the simulation-computer system, though this is coloured by fears of seduction. Finally, modern simulation from the 1990s to present day is characterised by power, in which simulation manifests expansions of all of the previous themes.

It was, of course, the practitioners that shaped the early understanding of simulation. As the focus changed from the physical instrument of simulation – the computer – to the importance of the model, it became clear that a simulationist needed to be a modeller, as well as a programmer. The degree of rapport the user had with their simulation characterised their interaction as an experimenter-instrument relationship. As computational technology improved, and simulations became more powerful, in certain disciplines the roles of the simulationist and the scientist collapse into one. These epistemic themes are inextricably intertwined with the promotional discourse surrounding simulation, through which the champions of the new technology legitimised simulation, and thus simulation knowledge. Concluding this chapter requires drawing attention to two further elements running through the themes: first, that simulation's seemingly hybrid nature is connected to simulation becoming an established and accepted scientific technique, by the linking of simulation to experiment and theory both explicitly and non-explicitly. Second, that these outcomes affect how simulation is characterised in the literature, perhaps clarifying the reason behind the many and incredibly varied interpretations of what simulation is, and where it falls in science.

2.7.1 From Traditional to Sui Generis

If by 1986, “simulation has become an accepted methodology for study and problem analysis in most branches of both the hard and soft sciences”, then by 2011, “the use of computer simulation is increasingly becoming central to scientific enterprise”. Separated by almost four decades (writing in 1973 and 2011 respectively) the old and new editors of Simulation both urge us to use proper documentation that demonstrates the validity and reproducibility of simulation so as to maintain credibility. The difference is that McLeod’s successor as editor, Levent Yilmaz, is not concerned about selling simulation or maintaining the credibility of the technique, but about the accuracy of its use in the broader context of scientific application. McLeod wanted to establish the credibility of simulations, and Yilmaz is more concerned about developing adaptable and reproducible models that can be used by scientists who are developing simulations: “progress in simulation-based science requires the ability of scientists to create new knowledge, elaborate and combine prior computational artefacts, and establish analogy and metaphor across

No longer is simulation struggling to turn from a technical skill into a professional area; it is now a widely used scientific technique.

Looking at simulation historically was initially an attempt to understand why simulation became so widespread from the perspective of historical and philosophical studies. It seems like an important question; according to McLeod, “the failure to answer this question [why simulate?] unequivocally when first considering the matter is one reason simulation technology has growing pains”. The motivation for this question was the rapid and seemingly unquestioned acceptance of a technique which seemed to raise, upon inspection, several epistemological concerns. I am not alone in my puzzlement; a long-time simulation expert stated in 1977 that while the wide selection of simulation topics and the increased acceptance of simulation study results was not unexpected, he had been “somewhat surprised by the nonchalant acceptance of the simulation approach by so many users”. Simulation was not associated with a radical shift in thinking, or even a revolution in practice, and yet somehow the face of scientific practice – what it means to do science – has been irreversibly altered.

Simulation’s own virtues certainly play a role in its rapid acceptance – who wouldn’t want to use a technique that performed your calculations quicker than ever before and provided you with visual, evolving results? Simulation provides access to systems that are inaccessible and otherwise incomprehensible. It is undoubtedly successful: “the efficacy of [simulation’s] generic principles in diverse environments, as directly experienced by practitioners and users, has produced a form of pragmatic proof of the robustness of simulation’s validity. It has established the legitimacy of simulation as a concept and tool.” McLeod emphasised the capacity of simulation to be applicable across multiple areas, that it was not a technique restricted to a single discipline. He used, to explain this, the example of a mathematical model where the same integral can describe a model of water flowing into a bathtub or, with different variables, a model of how long a vehicle will reach a certain speed. Like mathematics, simulation can be adapted to a variety of different situations, and to diverse and disparate scientific fields. Terry Shinn argues that the development of C++, a common coding language used in simulation, meant a lingua franca that connected simulation’s large number of ‘markets’ and practices with a stable and standard computer language – “a kind of research technology connected to many and diverse audiences and functions and highly amenable to transversality in the form of boundary crossing and commensuration”. Like with a research technology, the more disciplines simulation is successfully applied to, the more faith its practitioners have in the technique. Part of this flexibility may be due to the fact that simulation’s environment is virtual, one Simulation author

297 Yilmaz, 2011, p. 4.
298 McLeod, 1973a, p. 10.
300 Küppers et al., 2006, p. 20.
301 McLeod, 1972b, pp. 5-6.
answering his own question: “How have we reached this state of versatility in simulation? By
doing experiments on models instead of on real systems.”

In addition to these factors, this chapter has made the stronger argument that simulation became
established through nondisruptive paths, inculcating itself into common practice through the use
of particular methodologies from existing traditions. From the tradition of theory, simulation
borrowed the modelling technique, and thus established epistemic continuity with a very familiar
way of doing science. Winsberg, in a similar vein, argues that when validating a simulation, “the
credibility of that model comes not only from the credentials supplied to it by the governing
theory, but also from the antecedently established credentials of the model-building techniques
developed over an extended tradition of employment.” In addition to this, simulation borrowed
from the tradition of experiment the relationship between experimenter and instrument (and to
a lesser extent between experimenter and system), and thus established another continuity of
practice.

The idea of cultural ‘borrowing’ to establish credibility has also been discussed by Robert Kohler,
who describes field biology as a border culture between natural history and laboratory work. The
new naturalists sought to capture the ‘living interest’ in nature as a cure for microscopists’
myopia, but also required the credibility given by precise and unsentimental labwork. Field
biologists had little success translating lab methods to their own practice, but had more success
when they tried to “give traditional field practices the force of laboratory practices but without
importing the paraphernalia or protocols of laboratory culture.” This meant framing natural
processes and the measurements of them taken by field biologists as ‘nature’s experiments’, an
odd blend of the authority of experiment and the authenticity of nature. Similarly, simulation
is placed between two established traditions and borrows from both of them in order to lend
itself authority, authenticity, or credibility. Furthermore, these borrowed aspects are then
blended in peculiar ways – field biology watches nature experiment, simulation experiments on
models – which then become the hallmarks of a distinct technique, or discipline. As the next
chapter will explore more closely, interpreting simulation ‘as theory’ or simulation ‘as
experiment’ helps legitimise the knowledge that simulation produces – in this chapter, such
positioning helps legitimise the simulation technique.

2.7.2 The Value of a Historicised Perspective

A historical characterisation of simulation has the advantage of drawing attention to particular
aspects of simulation. In general, philosophers tend to characterise simulation using the terms
‘experimental’ and/or ‘theoretical’. Locating simulation with respect to these traditions
historically is useful in understanding what aspects of these traditions simulation inherits, but in
general characterising simulation only with respect to existing traditions is limiting. This

\[304\] Ören, 1977, p. 182.

\[305\] Winsberg, 2010, p. 45.


manifests in various ways in the literature; simulation has been described: *reducio ad* model, as experiment, as *like* experiment, as a hybrid of experiment and theory, as purely theoretical, as pragmatic, and as mediating between theory and experiment (or between models and data). That there are so many characterisations, sometimes polar opposites of each other, states to the unfruitfulness of trying to understand simulation within familiar epistemological categories. This is why simulation has not been fully characterised as a border culture (between theory and experiment), despite having in common borrowing tendencies; ‘border culture’ implies too much reliance on the traditions either side of the border, whereas simulation seems to be on its way to forming its own country, so to speak.

Walter Karplus, who gave the keynote address at the Summer Computer Simulation Conference in 1972, suggested that “every technological speciality goes through a multi-step irreversible evolution”.\(^\text{308}\) Though the applicability or validity of this approach may be questioned, its terminology is indicative. From Karplus’ perspective, there were four stages of development, defined by their relevant actors: a) the inventor or technician (between 1938 and 1953), b) the engineer or mass producer (from 1953 to around 1965), c) the theorist or physical scientist and mathematician (from the late 1960s to the 1970s), and finally d) the ‘ultimate user’ or social scientist (1970 onwards). Karplus noted that while the professions in all stages will obviously coexist at all times, the actor who is *most* important will change. There are similarities between Karplus’ stages and the themes of this chapter: not only is there is an early emphasis on the physical computer, but Karplus also sees the shift from engineer to ‘theorist’, commenting that in 1952, “the first and foremost problem was to get the problem running correctly on the computer ... the problem at the present time [1977] is one of *mathematical modelling*”,\(^\text{309}\) At the same time, Karplus’ ‘history’ reflects his own: the ‘theorist’ was not one who uses simulations to develop theories in scientific disciplines, but one who theorises *about* simulations, called upon to formulate fundamental theories about the operation of the technology, and put simulation on a more scientific and respectable basis. This matches the push to professionalise simulation noticed at the start of the 1970s in *Simulation*, and also Karplus’ perspective of considering simulation as a stand-alone discipline. His fourth stage, too, reflects the state of concerns at the time. The ‘ultimate users’ were “those who are more interested in the ultimate impact of the development rather than in its technical details”,\(^\text{310}\) These ‘social scientists’ were not researchers who use simulations to explore topics of interest to the field of social science, but who instead focused on the broader scales of application – transportation, urban problems, politics, the environment – the so-called ‘world simulations’ made popular in the early 1970s. Ironically enough, this stage may be experiencing a comeback (in *Simulation*, even) amongst the drive for sustainability and the concern with the world’s climate.\(^\text{311}\)

---

\(^{308}\) Karplus, 1972, p. 21. Karplus relates that he has based his ideas on those presented at a symposium by E. Tomash, the cofounder of a corporation in 1962 specialising in computer technology.

\(^{309}\) Karplus, 1977, p. 29. Emphasis in original.

\(^{310}\) Karplus, 1972, p. 22.

\(^{311}\) Bruzzone, 2014.
Karplus’ history, while certainly vulnerable to methodological criticism, is nevertheless useful in revealing how positioning can alter not only the history but the philosophy of simulation. As he was a computer scientist, is it perfectly permissible to accuse Karplus’ history of being whiggish, but we must also avoid this trap as well – even the suggestion that, once upon a time, the ‘model’ was not an integral part of ‘simulation’ (and today the two words can be synonyms!) tells us that the technique has a definite historicity. Not only the themes themselves, but the shift from one theme to another can give us multiple useful philosophical perspectives – the association of simulation with the physical computer was significantly less conscious when the validation of the model became more important, but the embodied aspect of simulating is also an important feature of the man-machine relationship.

As an example of how these themes affect the historiography, consider Tarja Knuuttila and Merz’s account of how understanding can be gained by simulation modelling. The authors make the somewhat strange claim that models have a concreteness, “a material embodiment that gives them the spatial and temporal cohesion that enables their workability”.312 Such embodiment comes about in the form of symbols and diagrams on paper or screen that we use to communicate and develop models. This objectual perspective depicts models “as objects through which we can make our theoretical ideas workable and ’experimentable’,”313 understood in the way that entities can be both objects (epistemic things) and as tools (technological objects). Scientific work (experimental work) revolves around objects, and importantly it is this “interactive relationship between scientists and objects [that] is crucial to conceiving how modelling generated expertise and skills on one hand and scientific knowledge and understanding on the other.”314 The analysis is sound, but inverted. The simulations taken as case studies suggest the same experimenter-instrument relationship outlined in ‘Man and Machine’, and so Knuuttila and Merz conjecture that the subject of study (the model) must be analogous to the subject of study of experiments – an object. Since experiment is intrinsically material, so must this other object be, and hence the stretched claim that models have an epistemically meaningful concreteness: theoretical models incorporated into the experimental tradition. Instead of assigning materiality to the simulation through the use of the computer, Knuuttila and Merz reduce the simulation to the activity of the model, and assign materiality to that – “without materiality, mediation is empty”.315 Again, there is the link between being able to ‘experiment’ with models (through the medium of simulation, though this is underemphasised in the paper) and with those models somehow having to be material. Yet the materiality that Knuuttila and Merz are talking about is the embodied materiality of theoretical modelling – writing, reading, printing, shuffling papers, making notes on touchscreen or whiteboard or notepad. Simulation was similarly embodied in the section on early simulation. However, there is a difference between embodied materiality and object materiality – modelling (simulating) is material, but simulation’s ability to make models ‘experimentable’ is

312 Knuuttila & Merz, 2009, p. 150.
313 Knuuttila & Merz, 2009, p. 151.
not because models are concrete objects *per se* (though they certainly have material dimensions in enactment and application) but because simulations are able to provide *manipulability*.

It is not only philosophers that must deal with simulation’s convoluted associations. A glance at the references for an astrophysical paper published in *Simulation* in 1970 shows a variety of names for simulation: numerical experiments, computer solutions, numerical computations, computer model, and machine calculation.316 This variety of characterisations, all legitimate, speaks of the traditions simulation borrowed from, but importantly they also describe the multitude of roles simulation finds itself playing in actual scientific practice. Putting aside, for a moment, philosophers’ opinions on the subject, the manner in which simulation is described by its practitioners also has an effect on how simulation knowledge is gained and validated – this will be the subject of the next chapter.

---

316 Hohl, 1970c, pp. 139-140. Curiously enough, none of the seventeen references contains the word ‘simulation’ in its title.
3  The Simulacrum of an Experimental Science: The Bar Instability and Dark Halos

3.1  Introduction
The last chapter attempted to understand simulation as a historicised technique, and in doing so drew out several themes that highlighted some general aspects of simulation. This chapter takes a more in depth look at simulation in actual practice, focusing on a small drama in the astrophysics of galaxies. This chapter, broadly, looks at the search for a stable model of a spiral galaxy – more particularly, the case study consists of how the ‘dark halo’ component of the galactic model was introduced and accepted. Early simulations of galaxies tended to model the galaxy as a somewhat unrealistic but highly symmetric system in order to reduce the number of calculations. Examples include representing the galaxy as a large number of one-dimensional mass sheets, or as a two-dimensional cylindrical collection of infinitely long mass rods. The “more realistic case of mass points moving in a plane seems, unfortunately, to be a much more difficult problem,” due to the complexity of the calculation of fields and the limitations of contemporary computers. Nevertheless, in the late 1960s astrophysicist R. W. Hockney, in discussion of a simulation modelling the galaxy as a cylinder of rod-like stars, noted that “some of the structures observed show a remarkable similarity to structures seen in real galaxies. … These calculations are necessarily exploratory, but it is hoped that they will aid our understanding of galactic structure and evolution.”

In his simulation, Hockney ran “some simple experiments” by varying the percentage of rotation that is required to balance centrifugal force against gravitational attraction. He found that the cylinder exhibited an instability whereby ‘flutes’ appeared on the surface:

Attempts to predict this surface instability ... have proved unsuccessful and we present this result here as an experimental observation. It is reassuring that similar surface flutes are also observed by Hohl ... in a model which uses an entirely different method for the calculation of the fields. It is most likely therefore that these flutes represent a physical phenomenon and are not due to some mathematical instability.

Though the ‘phenomenon’ Hockney uncovered was due to the choice of model as a cylinder and not physical after all, it is instructive to look at the language that Hockney uses when discussing his simulation. The simulation runs are directly referred to as ‘experiments’, and the surface instability in the simulation is described as an ‘experimental observation’. On one hand, the simulation results are not ‘predicted’ (by theory), placing simulation separate from theory, and

---

318 Hockney, 1967.
319 Hockney, 1967, p. 797.
320 Hockney, 1967, p. 797.
321 Hockney, 1967, p. 797.
322 Hockney, 1967, p. 804. The reference is to Hohl’s PhD thesis.
on the other hand, there is the question that the results may be the outcome of some mathematical error, placing simulation as distinctly non-experimental. This sort of language is a taste of what’s to come: in looking at how simulation is used in practice to generate scientific knowledge, simulation is placed to play a narrative role similar to that of experiment, but in a manner that maintains simulation as only ever metaphorically experiment.

Rather than answering the question of whether simulation ‘is’ or ‘is like’ experiment, this chapter looks at how simulation is represented as experiment through the use of specific language. By employing the metaphor of experiment, and the narrative of confirmation-of-theory-by-experiment, simulation is placed in a position opposite to theory, and so can act to confirm or test theory, as with the case of Toomre’s criterion. At the same time, simulation ‘predicts’ like theory, and is used to make theoretical statements, as with the Ostriker-Peebles conjecture. This flip-flopping between traditions allows simulation to be either (or both) ‘experiment’ and ‘theory’ where necessary – when simulation ‘confirms’ it is experiment, when it ‘predicts’ it is theory. The use of experimental language has a definite epistemological effect, allowing simulation to confirm a theoretical prediction. Furthermore, the flexibility of simulation in shifting between ersatz experiment and theory is a key component in how simulation helps produce new knowledge based on a limited amount of observational evidence – though it will also be apparent that observation (or other empirical grounding) remains necessary (dark halos could not be reified by simulations alone). Simulation will be shown to be placed as if it were experiment, allowing simulation to adopt roles in the scientific process that are traditionally the purview of experiment. Distance – and prevention of seduction – is maintained with the return to simulation as ‘theory’; this apparently contradictory manifestation of simulation’s dual nature is demonstrated through an analysis of the images, simulated and otherwise, used in the papers relevant to this chapter’s case study. The chapter concludes by suggesting that the overall narrative attempts to reconstruct astrophysics as an experimental science, but simulation’s ineluctable multifarious nature results in a discipline that deals increasingly with virtual, abstract, and flexible objects.

3.2 Miller and the Metaphor of Experiment

In the late 1960s, computational techniques were developed that made modelling an infinitesimally thin disc of point stars confined to motion on a plane – then the preferred model of a galaxy – become viable. One such early simulation, published in 1968 by Richard Miller and Kevin Prendergast, proudly described “a new approach to numerical experiments with self-gravitating stellar dynamical systems ... that permits enough bodies to be included to make the problem interesting.”323 These two authors invited the reader to consider a game – a point moving over a two-dimensional lattice – with a set of rules. The point could only move to locations with integer \((x, u)\) co-ordinates, and must always alternate a move where \(x\) changes with one where only \(u\) changes. The value of \(x^{(n+1)}\) was determined by the previous value of \(x\) and \(u\), and the value of \(u\) was determined by a function \(f(x)\) (see Figure 4). Regardless of the function, Miller

and Prendergast noted that it was impossible for the contents of two cells in the lattice to occupy the same space at the same time, nor could the contents of one cell split to occupy two cells, i.e. particles initially together remained together. The process was also completely reversible – after \( n \) steps, if the operations were performed backwards \( n \) times, the points would return to their original positions.

Figure 4: Systematic moves of many particles in the game described by Miller and Prendergast. This simple example shows the moves in which \( x \) is changed according to the present value of \( u \). For example, for a particle at \( x = 1 \) and \( u = -2 \), on the next move the particle will be at \( x^{n+1} = x + u = -1 \). When \( u \) is determined by a function \( f(x) \), the two moves together \((x \text{ and } u)\) make a complete step (Miller & Prendergast, 1968, p. 701).

If we set \( x \) as a coordinate, \( u \) as velocity, \( f \) as force per unit mass and \( n \) as time, then “this system describes the reversible flow of an incompressible fluid in the (discrete) phase space”,\(^{324}\) which approximates to solutions of the collision-free Boltzmann equation. In galaxy dynamics, physical collisions are rare enough that particles in galaxy simulations are considered collisionless, and the ‘flow’ of matter can be approximated as a fluid behaving according to the Boltzmann equation. Miller and Prendergast were clearly pleased with themselves at such a neat formulation: “the game contains the physical features of the phase-space description exactly. … The game carries out an approximate integration without doing arithmetic. The game clearly can be generalized to more dimensions.”\(^{325}\) With the ability to approximately integrate without arithmetic, Miller and Prendergast had developed an algorithm that they could use to model a disc galaxy on the limited computers of the time: “the game was designed to fit easily into the basic operations of a digital computer.”\(^{326}\) The authors, in fact, specifically set this up from the start of the article, describing

---

\(^{324}\) Miller & Prendergast, 1968, p. 702.
\(^{325}\) Miller & Prendergast, 1968, p. 702.
\(^{326}\) Miller & Prendergast, 1968, p. 703.
the rules of their game in what is sometimes called 'pseudo-code', a qualitative step-by-step description of the procedure that the computer will go through when one presses 'run' in a compiler. Presented in this manner, the readers of The Astrophysical Journal in 1968, presumably experienced astrophysicists themselves, would find the method accessible and transparent. The tone of the article definitely indicated it was directed at an audience who may not have had extensive programming knowledge.

The first program using this method was tested on a single particle in a harmonic oscillator force field "to explore the programming problems."\(^{327}\) Once debugging had been completed, a full program was written for a larger machine at the Goddard Institute for Space Studies in New York; "the properties of the machine that make it well suited to this problem are (1) large memory ... and (2) a fast processor."\(^{328}\) Miller and Prendergast's configuration space consisted of a 128 by 128 grid of cells which was periodic – a particle leaving one boundary was replaced by one entering the opposite side with the same velocity. The force of interest for stellar dynamics is self-gravitation, and was calculated by solving the Poisson equation at each step using Fourier methods. This meant that the time required to calculate the force was independent of the number of particles, which meant that 'the game' could permit large numbers of particles. The program reported "makes about eight complete steps per hour",\(^{329}\) where about half of the time was spent recording the results. At each timestep, the entire phase space was copied onto magnetic tape, so that post analysis was possible without rerunning the simulation. Taking the time to inform the audience of such trivialities of the process is typical of early simulation reports, reminiscent of the 'method' section of an experimental report and matching up with the early theme of the physical computer discussed in the previous chapter.

Miller and Prendergast convinced themselves of the appropriate behaviour of their system through some tests (checking to see if the program maintains an initial symmetry of a system, for example) and summarised:

The approximation conserves its integrals very well and appears to conserve the usual integrals of the mechanical system being mimicked better than we had expected. Because the game matches the physics of collision-free motion exactly, the development of the systems is probably a reasonably faithful representation of a physical system. We are very pleased with the evident good behaviour of the game and anticipate a fruitful application to stellar dynamical problems.\(^{330}\)

Miller and Prendergast, together with William Quirk, expanded on their simulation in 1970 (a paper referred to by multiple later authors, and in this chapter, as MPQ). Their later simulation expanded the grid to 256 by 256. In this version, the authors aimed to determine what affects the persistence of spiral patterns; in their 1968 paper they had observed short-lived spirals in the

---

\(^{327}\) Miller & Prendergast, 1968, p. 703.
\(^{328}\) Miller & Prendergast, 1968, p. 703.
\(^{329}\) Miller & Prendergast, 1968, p. 704.
early stages of their n-body systems but, “because they are so short-lived, these spectacular patterns cannot be valid n-body analogues of the spiral patterns of actual galaxies.” Despite their unreality, the existence of any spiral patterns did give the authors more faith in their model: “since we have now observed spiral patterns in our machine calculations, it seems that the numerical experiments are on the right track.” In order to observe ‘real’ spiral patterns, however, the simulation would need to be more realistic. Most real galaxies are ‘cool’ – particle motions are nearly circular, and there is not much random motion. However, “machine calculations typically produce ‘hot’ systems that are largely pressure-supported. These ‘hot’ systems cannot respond to irregularities in the gravitational field; spiral patterns should be sought in a ‘cool’ (dynamically) stationary self-consistent system.” The velocity dispersion – the amount of random motion, and what determines how ‘hot’ a system is – typically becomes large in simulated systems as they evolve.

Miller et al. found it easier to attempt to cool the systems “already in the computer” than to try and re-model a cool system from scratch. A fifth step was introduced into their ‘game’ outlined in the 1968 paper: the velocities of the particles at a certain location were re-arranged in such a way that the total momentum of those particles were the same before as after, yet the velocity dispersion was reduced. This method of ‘cooling’ meant that the particles involved behaved something like an interstellar gas: “they collide inelastically, but retain their identity. There is no mass exchange. A component of the system that is treated this way is called ‘gas.’” Hence the authors’ new simulation was a two-component system, with a ‘gas’ component that was artificially cooled at each step, and a ‘star’ component that was not. This model was more realistic, and “with a rule for changing ‘gas’ into ‘stars,’ the calculation quite naturally mimics the normal processes in a galaxy with gaseous and stellar components and a means of forming stars out of the gas.” Nine runs were performed with this model, varying the initial conditions (whether the rotation balanced the gravitation force or was half as large, whether there were stars at the start or whether they were formed from the gas, how many stars were allowed to form, and so on). Run 9, carried much further in its evolution than the other runs, displayed persistent spiral patterns in the gas for about three rotations, whereupon the spiral had tightened up too much to be distinguished; “the density pictures of the ‘gas’ component suggest actual galaxies. The ‘star’ component showed much less structure; it looked rather like a photograph of a globular cluster.” The authors concluded that pure ‘star’ or ‘gas’ systems were not good candidates for spiral patterns, but systems where stars formed from gas were.

---

331 Miller & Prendergast, 1968, p. 904.
332 Miller, Prendergast, & Quirk, 1970b, p. 903.
333 Miller et al., 1970b, p. 904.
334 Miller et al., 1970b, p. 905.
335 Miller et al., 1970b, p. 905.
336 Miller et al., 1970b, p. 907. The authors do not specify why they chose to carry Run 9 further than the other runs. This might be taken as an instance of the natural rapport the simulators had developed for their simulation through the processes of construction and tweaking.
Miller et al. referred to their simulation runs as ‘experiments’ throughout the article. Consider the following excerpt (with my emphasis):

The experimental evidence ... indicates that the system is quite sensitive even to rather small departures from the static self-consistent state. ... The computer model of [Miller and Prendergast, 1968] seems to be a useful representation that permits an experimental approach to the study of stellar systems. In particular, the achievement of a fairly long-lived spiral structure in a self-gravitating system means that we now have a laboratory in which to carry out experiments that may provide useful insights into the nature of systems that will support spiral patterns. We confidently expect that this will open up a fruitful research area – one for which we have only reached the threshold.\textsuperscript{337}

Using the words ‘experiment’ or ‘numerical experiment’ to refer to simulations was (and is) common. In a summary of a symposium at which Miller presented the MPQ simulation, the author records that, "Miller commented that these [numerical] experiments are not done to provide at great expense spectacular films, but that we now have a tool for actually experimenting with spiral galaxies and studying problems of internal structure, motion, stability, and evolution."\textsuperscript{338}

As the previous chapter argued, this reference to a well understood and already existing exploratory mode of working is used to connect the simulation technique to legitimate scientific practice. This language continues throughout the case study; simulation, with its experiment-like methodology, is easily spoken about using the discourse of traditional experiment. In the above excerpt, multiple runs of the same simulation each constituted a different experiment with different initial conditions; the simulation is a laboratory in which to run these experiments, where the results of which are experimental evidence (note the dual role of simulation as both experiment and laboratory). And a new research area opens up, through the ability to use something like experiment in astrophysics.

Simulation as experiment is, importantly, only understood as a metaphor. Expressing a vague worry of seduction, the MPQ paper emphasised:

The results are very suggestive; the models, at certain stages, look a good deal like actual galaxies. Many terms that exploit this suggestiveness are enclosed in quotation marks to stress their metaphorical character. But the very suggestiveness may encourage too literal an interpretation of the results. As with any gravitational n-body calculation, only qualitative features of this one are to be trusted. In particular, the initial conditions used, the early history, and the various time scales that emerge should not be taken literally.\textsuperscript{339}

Certainly the astrophysicists do not understand simulation as providing ‘real’ experimental evidence of the empirical sort that material experiments do. The careful use of ‘gas’ and ‘stars’ in inverted commas differentiated the particles in the model under a set of rules intended to mimic the behaviour of real gas and stars from the real thing. Though the appearance of spiral patterns

\textsuperscript{337} Miller et al., 1970b, p. 916. Emphasis added.
\textsuperscript{338} Bok, 1970, p. 466.
\textsuperscript{339} Miller et al., 1970b, p. 903.
meant the model was on the right track, their short-lived nature attested to their very unreality. This distanced the simulation from reality while at the same time attempting to demonstrate that the model was an adequate description of it. The simulation aimed to faithfully represent the dynamics of a real system, but was explicitly portrayed as the product of a theoretical model, perhaps to avoid the worries associated with confusing representation and reality (see section 3.9). The simulation runs were described as experimental because like experiment, they involved multiple parameter variation to test the reaction of the system, providing astrophysicists with an experimental methodology previously unknown to their discipline. Given this, it is unsurprising that simulation finds itself taking up similar roles to those experiment plays in canonical science.

Miller explicitly treated simulation like experiment in a paper analysing the possible sources of errors in simulations. Consider the following issue: simulations of the disc model tended to evolve into hot discs, and Hohl later noted that “this is typical of computer generated galaxies.”340 Is this a property of computer generated galaxies, however, and therefore unphysical? Miller set out to validate the entire approach of using simulation to explore galactic structure. He noted that the discrepancy between the high velocity dispersion observed in simulated galaxies and lower dispersion measured from observation presented a serious problem: “these results have important implications if they properly reflect the physics of flat galaxies”.341 Miller therefore created a program that performed the $n$-body calculations in polar coordinates, and combined his results with earlier results “to establish the validity of the simulations. The goal is to produce a definitive statement that the simulations represent the physics of the model, which implies that any disagreement with observation represents a deficiency in the model.”342 Simulations represented experiments on the theory – if the experiments revealed a disagreement with empirical data (in this case from observation), then the theory would be in fault.

Simulation, metaphorically regarded as experiment, must therefore be validated as producing results which reflect the physics of the real system. Miller’s metaphor, in this case, focused on the need for analogous validation of simulation results:

*We adopt the viewpoint that computation must be regarded as numerical experimentation.* This viewpoint is essential with astronomical systems which do not admit laboratory experiments, but it is useful in other areas of physics as well. An experimenter in a laboratory must be certain that his results reflect the physics of the system under test, rather than some form of experimental error. … The computer program is analogous to the experimental apparatus. The usual program verification checks are equivalent to a laboratory experimenter’s equipment checks. There is a peculiar tendency to accept computed results uncritically; the computational physicist must be even more careful than

---

341 Miller, 1976, p. 401. As we will see later, a halo of large mass might be required to resolve the problem, which if present in all galaxies may provide enough mass of account for the closure of the Universe.
342 Miller, 1976, p. 401.
a laboratory experimenter to satisfy himself that computational results reflect the physics of the problem under investigation.  

Though most authors did not make the analogy as explicit as Miller, their writing often referred to simulations or simulation runs as experiments, but like Miller were also conscious that their models needed to be validated as appropriate representation of reality – this latter move obviously redundant in traditional material experiments. Miller described seeing simulation as experimentation in terms of error analysis as ‘essential’ for the correct validation of the model – in an area like astrophysics where laboratory experiments cannot be performed, it is more important than ever that the theoretical models are shown to be representing the right thing. Taking on the false mantle of experiment is, oddly enough, one way to ensure simulations are properly validated. Having a foot in both theoretical and experimental camps means simulationists often talk of experiment and model validation in the same breath:

An additional danger is that important physical effects may have been omitted in the original design. In this respect, computational physics is closer in spirit to theory, which is usually based on abstractions designed to leave out ‘unimportant’ details in order to produce a manageable system. But the abstraction may be so severe that there is little resemblance between the physical system and the model. The computational physicist can include more effects than can a theoretician, so he can get closer to the true physics; but he still deals with oversimplified systems.

Miller saw simulation as having some sort of epistemic primacy over theory as through the ability for simulation to handle complexity you can get ‘closer to the true physics’. Though simulation can make more complex systems, they are only more theoretically complex; Miller recognised this and went on to suggest that ”computer experiments are most convincing when they can be used to identify those physical processes which are most important in producing the effects observed in nature". Simulations are therefore limited to uncovering the mathematical form by which natural mechanisms can be expressed – Miller stopped short of attributing to simulation the same epistemic power granted to experiment.

One expects nothing else, despite Miller’s free use of ‘experiments’, ‘experimental confirmation’, and even ‘empirical evidence’ with respect to work done solely with simulations. Like metaphors tend to do, however, the metaphor of experiment extends beyond analogies between methodologies (like parameter variation) and the importance of thorough error analysis. In Miller’s own work, and common in others, is a tendency to structure the paper as an experimental.

---

344 Miller, 1976, p. 401.
345 Miller, 1976, pp. 401-402.
346 Metaphors can play an important epistemological role in science, and often this role expands beyond the boundary of mere analogy. For example, the information metaphor – DNA as code, and the genome as an information system – is central to understanding post-war genetics (Kay, 1998). Another computer-related example: changing concepts of intelligence and calculation eventually fostered the metaphor of, and then the theory of, the mind as computer (Gigerenzer & Goldstein, 1996).
report. In a simulation study looking at stable classes of models, Miller had a ‘method’ section, followed by a section on ‘experimental method’, and then one on ‘experimental results’. In one table, he listed comparisons between ‘expected’ (theoretically determined) values for certain models, and ‘observed’ (simulated) values.

In Miller’s metaphor, it was not a galaxy that was being simulated, but an experimental setup. The previous chapter showed how simulation could be understood as simultaneously opaque and transparent by understanding how simulation plays both the role of experimented with and experimented on. Through the metaphor of experiment it becomes apparent how this duality manifests in actual scientific practice: in describing a game of rules that govern the movement of each particle, and in coding the mathematically-physical laws that these particles obey as ‘stars’ or ‘gas’, a system is being simulated that contains both the model of the target system (a galaxy) and the means to draw information out of that system. In simulation, taking ‘measurements’ of velocity, mass, and position occurs simultaneously with the creation of the ‘phenomena’, as the computer stores the values it uses to advance the particles so that these values may be used in later analysis. The simulated setup is then framed in the discourse of experiment to distinguish it from the simulated object; this is what Miller was doing when he treated simulation as experiment in order to validate the simulation.

The metaphor of experiment blurs the line between traditional experiment and simulation, especially in their relationship to theory. As the next few sections will demonstrate, such language places simulation in a position to take up certain roles that experiment is supposed to play in canonical science. In astrophysics specifically, due to the lack of traditional experiment, simulation takes up the role as the ‘other’ way of doing science, the way that is typically contrasted with theory.

### 3.3 A Local Instability

It is necessary, in order to place this chapter’s simulation case study in context, to begin with an analytical study. In 1964, Alar Toomre undertook a study of a self-gravitating, thin, rotating disc of stars, ignoring the effects of gas or dust as a first approximation. The motivation for this study was that “no non-rigid, infinitesimally thin, plane sheet of gravitating matter – not even a perfectly uniform, infinite sheet – could long endure near its original state in the presence of the slightest disturbances if it lacked all stabilizing influences.” Such stabilising influences, including centrifugal forces stemming from rotation and the ‘pressure’ from random motions of individual masses, would have to be enough to counter any disturbances in the system. If they were not, then the flat disc system would become unstable.

Toomre undertook an analytic – that is, non-simulation – study to test the stability of these discs. He first considered a thin disc rotating about an axis of symmetry that was in equilibrium with

---

347 Miller, 1978a.  
respect to both gravitational and centrifugal forces. He then introduced a disturbance in the form of a contraction of a small region of the disc. A particle of matter (a star) near the edge of the region would be, due to this contraction, pulled closer to the centre of the region, feeling an additional gravitational pull. However, there would also be an increase in the centrifugal force on the particle which would act to push the particle out of the region. Toomre's analysis showed that for regions shorter than a characteristic length, the centrifugal force would not overcome the gravitational force. Thus short disturbances would not be smoothed out by rotation and would remain unstable.

In order to smooth out the instabilities, Toomre conjectured, there needed to be a certain amount of random motion of the particles. The particles needed to travel with a motion separate from the common rotation to have a dissipative effect on the instability. As the amount of velocity dispersion increased, the range over which instabilities could form would become smaller, and beyond a finite value instabilities could vanish completely. Toomre, after these order of magnitude calculations, performed a more detailed analysis of the problem which blended analytical approximations and machine calculations; “through the use of an electronic computer”, he relieved some of the mathematical burden of the large-scale examples. Toomre arrived at the conclusion that an arbitrary sinusoidal disturbance of axial symmetry can be avoided only if the radial velocity dispersion is of minimum \( \sigma_{u,\text{min}} = 3.36G\mu/\kappa \) (see Figure 5).

Figure 5: Toomre's analysis showed a threshold velocity dispersion is required. The peak at 0.2857 corresponds to the value of \( \sigma_{u,\text{min}} \) quoted above (Toomre, 1964, p. 1233).

Toomre showed that rotation itself could not ensure the stability of a thin, gravitating sheet. Thus the model of a galaxy as a flat rotating disc of stars necessitated having sufficiently large random

\[ \sigma_{u,\text{min}} = 3.36G\mu/\kappa \]

---

349 Toomre, 1964, p. 1226.
350 Toomre, 1964, p. 1234. \( \kappa \) is the local value of the epicyclic frequency – the frequency at which a region oscillates about its mean position.
velocities in the disc, which, in combination with rotational forces, maintained stability. Toomre's calculations were considered analytical by the astrophysics community, and his conclusion theoretical. It would take until roughly 1969 for his work to be specifically tested with simulation.

3.4 Toomre’s Criterion and The Narrative of Theory and Experiment

Whereas MPQ studied a squadron of galaxies, the other prominent galaxy simulators – R. W. Hockney and Frank Hohl – studied an isolated galaxy.\(^{351}\) Like MPQ, the Hockney-Hohl model involved point particles that moved through time according to a set of rules. The single galaxy was divided into a square array of cells (in this case, 64 by 64) they called the ‘grid’. At each timestep the mass distribution was calculated, and then converted to a potential distribution. This was then used to recalculate the positions of the stars and advance them according to the gravitational force applied to them (see Figure 6). The system then moved one timestep forward, and the calculations were repeated. The authors described their approximations as "the simplest that can be devised and lead to the fastest computer program."\(^{352}\) Refinements were possible but seriously increased the ‘cycle time’ (the amount of time taken to compute the required values at each timestep) and so would have required that fewer stars be included: "since there is considerable premium on having a large number of stars in a galactic simulation we have preferred to use the simpler model."\(^{353}\) A large number of stars provided a better density distribution and a good statistical measure of stellar distribution – the aim of the simulation affected these choices. Like MPQ, Hockney and Hohl took the time to go through the process of their simulation step-by-step in a qualitative description of their program.\(^{354}\)

Hockney and Hohl validated their simulation in a way that would have pleased Miller. The authors compared two methods of calculating the potential – by Fourier transform or by summation. They tested for various codes, methods of storage, and numbers of stars, and established that the Fourier method was three times as fast as the summation method, but required 25 per cent more storage.\(^{355}\) They demonstrated that there was little significant difference between taking 200 timesteps per rotation and taking 400 timesteps per rotation, and that 50,000 stars produced results very similar to 200,000 stars. Considering a 64 by 64 mesh also provided the same results, though a little coarser, than using a 128 by 128 mesh. (Note the focus on the capabilities of the physical computer and the effect this would have on the simulation model.) Systematic variation of parameters and comparison of outcomes increased their faith in their simulated experimental

\(^{351}\) Hockney & Hohl, 1969a and Hockney & Hohl, 1969b.
\(^{352}\) Hockney & Hohl, 1969a, p. 308.
\(^{353}\) Hockney & Hohl, 1969a, p. 309.
\(^{354}\) They also relate how the coordinates are stored in computer memory, 'packed' and scaled to minimise the space used. This section is only in the paper published in Computational Physics and is a good example of how the limitations of the physical computer was an important part of constructing these early galaxy simulations.
\(^{355}\) Hockney & Hohl, pp. 312-317.
system: "hence we believe that the parameters used in these experiments give a good model of a star system."\textsuperscript{356}

The development of this good model, then, "makes possible the experimental verification, on the computer, of theoretical predictions of the behaviour of such systems."\textsuperscript{357} Like in MPQ, the metaphor of experiment was used, through here it moves beyond simple validation methods into the capacity of 'experimental' verification. Like in Miller's work, the layout of the article reflects that of an experimental report: following the introduction, the model was briefly described, followed by initial conditions, and then a results and discussion section. The aim of Hockney and Hohl's 'experiment' was to check Toomre's prediction that $\sigma_{u,\text{min}} = 3.36G\mu/\kappa$ provided sufficient velocity dispersion to prevent instabilities – a 'test' of the theory. The cold initial disc in equilibrium was found to be unstable, "a result which is in agreement with theoretical predictions"\textsuperscript{358} (see Figure 7). By introducing velocity dispersions in increasing percentages of the circular velocity, it was found that the higher the percentage the longer instabilities took to form, and that instabilities were stabilised at smaller wavelengths. At a percentage of 27.2 the

\textsuperscript{356} Hockney & Hohl, 1969b, p. 1102.
\textsuperscript{357} Hockney & Hohl, 1969b, p. 1102.
\textsuperscript{358} Hohl, 1970c, p. 139.
disc was found to be relatively stable, and as smaller than the maximum velocity dispersion required to stabilise the disc "the value of 27.2% found from the computer simulation is in good agreement with the predictions of Toomre".\textsuperscript{359} With Toomre's precise conditions for a minimum velocity dispersion, "the disc is just barely stable",\textsuperscript{360} which is what one would expect from a minimum condition.

Here, simulations played a role in confirming a theoretical prediction. Toomre's analysis was a theoretical one done with a little aid from the computer, and made no use of simulations. Perhaps more importantly, it was considered by his peers to be a theoretical analysis. Toomre's conclusion was necessary for theoretical consistency – if we want a thin disc of stars to adequately model a galaxy, there must be a certain minimal velocity dispersion to maintain stability (and the model

\textsuperscript{359} Hockney & Hohl, 1969b, p. 1104.

\textsuperscript{360} Hockney & Hohl, 1969b, p. 1104.
must be stable if it is to represent a galaxy). Hockney and Hohl coded a thin disc of stars that obeyed standard motion laws, and pressed ‘play’. They found that the disc was, like Toomre concluded, unstable. They increased the amount of velocity dispersion until the disc was sufficiently stable, and their numbers matched up with those of Toomre. If it were not performed on a computer, Hockney and Hohl’s work would be classical confirmation of theory by independent experiment. The simulation was not a traditional material experiment, but it nevertheless fits well into the narrative of theory-confirmed-by-experiment.

In the history of the philosophy of science (post 17th century), the understanding of the relationship between theory and experiment can be seen to follow a familiar path. According to the traditional view in philosophy of science, it was only through experiment that a theory could be confirmed or disproved – experiment was the arbiter of theories. This understanding, however, implied that experiments were only (meaningfully) performed to test a theory, while any actual shift in knowledge came about due to the shift in theory. Duhem’s theory-ladenness of observation, in the hands of philosophers following the historicist turn in HPS, further reduced the autonomy of experiment by calling into question even this singular, but crucial, role. In reaction to this, a new experimentalism emerged, which took up the task of describing, historically and philosophically, the multiple roles of experiment and emphasising its practice (rather than just the results); Steinle’s account of exploratory experiments in the early history of electromagnetism, experiments which did not aim for theoretical explanations, is one such example.

Though philosophers of science have moved on, modern scientists engaging in occasional philosophy have a tendency to present the relationship between theory and experiment in arbitrator terms. Despite being philosophically outdated, it is a pervasive narrative in the more general reflections of scientists and is closely tied to what many feel makes ‘proper’ science, and the crucial role experiment plays in this understanding:

The narrow view of experiment as the handmaiden of theory is not wholly false. The caricature considers only observations in the form of statements, and dismisses the processes of eliciting and producing observable phenomena. But there is an element of truth in this caricature, and here the element of truth is that experiment is important.

Modern scientists certainly don’t limit their use of experiments to only testing theory, but when questions of theory and reality do arise, experiments maintain their narrative role of arbitrator as the “source of the observation statements that confirm or falsify theories.” The Popperian language that David Gooding uses here is important, as Helge Kragh observes “while Karl

363 Steinle, 2002.
364 Gooding, 2000, p. 119.
365 Gooding, 2000, p. 118. This also means that, especially in certain areas of physics, experiment is reduced to being the handmaiden of theory.
Popper’s philosophy of science has only few followers among modern philosophers, it is easily the view of science with the biggest impact on practicing scientists.” Krigh’s analysis of the role of Popper and his theories in cosmology is an example of the ‘experiment tests theory’ narrative. Many of the scientists in this area invoke something like Popper’s idea of falsificationism, whether by name or implicitly; Krigh suggests that “the influence of falsificationist philosophy à la Popper may be particularly strong in the physical and astronomical sciences.” Importantly:

All (or ... nearly all) physicists agree that testability is an epistemic value of crucial importance. They consider it an indispensable precondition for a theory being scientific: a theory which is cut off from confrontation with empirical data just does not belong to the realm of science.

To be properly scientific the theory must be testable, and this is often tied to testing by experiment or comparison with observations. Krigh considers this a naïve falsificationism, born of what scientists know about Popper’s theories but without a more sophisticated understanding, “a simplified folklore version ... boiled down to the seductively simple formula that a nonfalsifiable theory is unscientific per definition.” Furthermore, this simplified Popperianism is typically confined to popular books, review articles, and public lectures, and “had only very limited influence on their actual scientific practice.” Popper is used rhetorically, not as a prescriptive way of doing science. Even physicists’ consensus that testability is important, Krigh emphasises, is a rhetorical consensus given that they disagree wildly on testability in practice. This is important to the present discussion because falsificationism, especially in its naïve form, forms the basis for the narrative that, in proper science, theories must be able to be arbitrated, and what is doing this arbitration is typically the role of the ‘other’ form of science – classically experiment or observation. And when spoken of using the language of experiment, simulation is placed to follow this narrative.

One can now see how Hockney and Hohl’s simulation could ‘confirm’ Toomre’s criterion: already spoken of as experimental in terms of methodology and error analysis, simulation easily plays the same performative role as experiment in the narrative of ‘experiment confirms theory’. As an

---

366 Krigh, 2013, p. 325. Godfrey-Smith makes the same observation (Godfrey-Smith, 2003, chapter 4).
367 Krigh, 2013, p. 343.
369 Krigh, 2013, p. 351. Sovacool, 2005, more optimistically, comes to almost the exact opposite view. In an analysis of astronomers’ use of Popper and his effect on their methodologies, Sovacool concludes that: “none seem to misunderstand or misuse Popper’s concepts. Each of these tendencies seems to suggest a tacit agreement to view Popper as an important authority within the practice of astronomy” (Sovacool, 2005, p. 60). Despite the difference in interpretation, both Krigh and Sovacool identify that the falsificationism is a common theme in cosmology and astrophysics.
372 It is important to stress that this is only an observation of the narrative of scientists – it is valid neither in philosophy, nor in actual scientific practice, despite the reconstructions of the scientists themselves. I am fortunate in that Krigh and Sovacool’s studies focus on cosmology and astrophysics (though Krigh cites similar narratives in biology and biochemistry); it is conceivable that the story would be different in a science based more heavily around exploratory experimentation.
independent study using methodology that was sufficiently far from Toomre’s definitely theoretical approach, Hockney and Hohl’s work certainly provided support for Toomre’s criterion. In terms of the typical scientific narrative, this amounted to a ‘confirmation’. Though the epistemic viability of this move could be questioned, Toomre’s criterion certainly did achieve a sort of ‘confirmed truth’ status from then on. In a 1974 study looking at the effect of Population II stars (older stars) on spiral structure, Hockney and Brownrigg wrote:

Hohl & Hockney (1969) have reported experiments on the stability of an infinitely-thin cold disc galaxy which they found to be violently unstable in agreement with the theory of Toomre (1964). Subsequent experiments by Hockney & Hohl (1969) showed, again in agreement with Toomre’s theory, that the disc could be stabilized by a sufficiently large dispersion in the star velocities.

The authors went on to demonstrate that their initially cold discs were also violently unstable as per Toomre, and that with an (unrealistically) large number of Population II stars – 80% and 90% of the total mass, which increased the velocity dispersion – the structure held well against Toomre-type instabilities. Toomre’s work had become a way of demonstrating that the simulation model was not unphysical – like basic tests of conservation of energy, or maintaining symmetry, the disc model should become stable or unstable on either side of Toomre’s threshold criterion to be a valid model. Of this 1974 paper, Miller noted that: “the experimental results confirm the theoretical expectations”, referring to the ‘confirmation’ by ‘experimental’ results of the theoretical expectations astrophysicists had about the effect of Population II stars on spiral structure. The narrative presented then reads as follows: Toomre made a theoretical prediction about stability; Hockney and Hohl’s simulation then ‘confirmed’ Toomre’s prediction, making it a ‘criterion’; and Hockney and Brownrigg used this criterion to validate their simulation, which in turn provided confirmation of another set of theoretical predictions or ‘expectations’. Knowledge is generated, validated, and used within a circle of simulation; and it is a circle because Hockney and Brownrigg’s model was “a slightly modified version of the Hohl-Hockney disc model”, the same model used to confirm Toomre’s criterion in the first place.

The questionable circularity of the validation process is probably less epistemically suspect than the narrative told by the astrophysicists might imply. Reviewing his criterion a decade later, Toomre noted that it had been corroborated by several independent analyses; he cited four, Hockney & Brownrigg, 1974, p. 351. Miller, 1976, p. 406. Hockney & Brownrigg, 1974, p. 352. The modifications are refinements to how the gravitational field potential is obtained, and to convert the model from two to three dimensions. Toomre, 1974, p. 458. Toomre does, in this review, suggest that the criterion needs to be corrected for a disc of finite thickness. Regardless of the actual numbers, however, the idea of a stability threshold called ‘Toomre’s criterion’ becomes embedded into the theory of galactic discs; this becomes more evident later when the community adopts Hohl’s parameter Q as an easier measure but continues to attribute it to Toomre.
only one of which (Hohl’s 1971 work, see below) was a simulation. Noting that “it has been customary to invoke the Toomre stability criterion ... in discussions of the numerical experiments”, Miller fastidiously determined (via simulation) that Toomre’s criterion, referring to axisymmetric disturbances in axisymmetric systems, did in fact apply to the nonaxisymmetric systems in which it was frequently used. He also set out to answer the related methodological questions: “could we recognize a genuine instability in a calculation? Could we recognize stability? Can an n-body calculation display a stability threshold like that indicated by the Toomre criterion?” Miller did find that numerical calculations can display a stability threshold, validating the technique used to validate Toomre’s criterion. So there was, after all, validation for the simulation and the techniques used that did not come from the simulation itself.

Toomre’s criterion was, however, corroborated only by non-empirical studies – either analytical work, or simulation work. Miller was therefore somewhat sceptical about the validation of the criterion: “an interesting feature of the Toomre criterion is that the velocity dispersion for stars in the solar neighbourhood or our Galaxy is near that required by the Toomre criterion; this coincidence is responsible for the importance attached to the criterion under circumstances in which the criterion is evidently not applicable.” The adoption of Toomre’s work does seem oddly bereft of empirical grounding; Toomre noted of his criterion that “unfortunately this simple local formula remains distinctly more certain than the set of ‘observed’ quantities \( \sigma_u, \mu \) and \( \kappa \) which it invites us to intercompare.” It is peculiar that a formula could be more certain than observational quantities, but this is typical of astrophysics, which has always keenly felt the limits of observational technology. Simulation can help push back that boundary somewhat, by providing another form of validation that, while not empirical, is data-like in being derived from a sufficiently complex system (see Chapter 4). Within the narrative of science (told by scientists), this puts simulation in place to play certain roles traditionally the purview of experiment. This is all the more evident in astrophysics, which relies on observation for its connection to the empirical and the real, but which is otherwise bereft of “the active interrogation of nature, an intervention in natural processes, and a manipulation of nature’s forces” – of experiment. However, as will become apparent later when halos are introduced, understanding how simulation produces knowledge involves more than representing simulation as experiment.

3.5 The Bar Instability

Toomre’s criterion was only the tip of the iceberg. In 1971, with respect to the Hockney-Hohl model, Hohl said “we were overenthusiastic in calling such discs stable.” He took a further look

---

378 Miller, 1976, p. 405.
379 Miller, 1974, p. 539.
380 Miller, 1976, p. 405.
381 Toomre, 1974, p. 458. The word ‘observed’ is in quotation marks because the quantities are not directly observed but are derived from other, more easily observed, quantities.
382 Arabatzis, 2008, p. 159.
at discs of stars with initial velocity dispersions at least equal to Toomre’s stability criterion to ensure that there were no local axisymmetric instabilities. Using a larger grid and increasing the number of stars to 100,000 Hohl found that while the disc was stable to small axisymmetric disturbances of those identified by Toomre, in the long term the disc "is not stabilized against relatively slowly growing large-scale disturbances which cause the system to assume a very pronounced bar-shaped structure after two rotations" (see Figure 8). This central bar-shaped structure was the new steady state of the system. Hohl was "able to generate an axisymmetric, stable disc only with considerable difficulty" by artificially symmetrising the bar structure in the central portion of the disc; attempts to cool the disc had little effect on the structure formed.

Figure 8: Hohl’s simulation demonstrates the formation of an unstable bar-like structure with two arms (Hohl, 1971, p. 347).

Hohl presented his results as revealing a flaw in the galaxy model, but not necessarily a fatal one: "the reason for using this highly artificial procedure of 'symmetrising' was mainly to show that the bar structure, although here persistent, is not necessarily very strong or 'deep.'" The bar was clearly not a real effect – though there exist barred spiral galaxies, the short-lived nature of the spiral arms in the simulation, dissipating as they did in one or two rotations, demonstrated the phenomena to be unstable, and therefore not real. The fault was obviously in the model, but given that somewhat ad hoc adjustments to the simulations (such as forcing symmetry of the bar)

---

did actually produce a stable disc meant that the fault was not theoretically ‘deep’, resulting as a consequence of some ‘higher level’ assumptions rather than a fundamental law.

The quest to develop a stable galactic model could only proceed with simulation. Toomre analytically ‘predicted’ small scale axisymmetric instabilities, but these were only ‘confirmed’ when the galaxy as a whole could be modelled and evolved. The combination of complexity and black-boxing meant that simulation could help scientists explore galaxy dynamics in a more coherent way. Using simulation, scientists could stitch together the various hypotheses and explanations of individual phenomena into a model without requiring a more complete understanding of their interactions. Hohl agreed:

> It is, of course, only with the advent of computer experiments such as those described here that it has become possible to determine the behaviour of discs of stars in the large. Before such experiments, there existed no really adequate theory describing the overall dynamics of discs of stars.\(^{387}\)

Simulation assisted astrophysicists in unifying what they knew about how a galaxy should behave into a coherent, testable, model. Semi-isolated hypotheses like Toomre’s stability criterion were identifiable with theory work, but not wholly testable in terms of the effects they had on other components of the model, or on the evolution of the system. Simulation could, however, perform these ‘tests of theory’, and could even reveal new ‘phenomena’ that emerged as a result of the model, such as the (unrealistic) bars.

Hohl noted that his result “is in qualitative agreement with the unstable two-armed density wave predicted by Kalnajs.”\(^{388}\) Agnis Kalnajs’ work, published in 1970 and then in more depth in 1972, investigated the stability of a flat model galaxy that was derived from the observed rotation curve of M31 (Andromeda), one of the nearest galaxies to ours. Kalnajs wanted to determine whether there could be stable equilibrium states with small random motions that could not be reached by rotationally supported initial conditions – that is, whether the standard self-gravitating disc model of galaxies could have any stable modes at all. He investigated the various modes of a linear family of uniformly rotating discs, each characterised by a different mean angular rotation rate, and determined that all variations were unstable. Instabilities that were smoothed by Toomre’s criterion were ruled out, but some remained: “the instabilities that remain ... are either nonaxisymmetric, or they are axisymmetric overstabilities.”\(^{389}\) In order to construct stable discs, Kalnajs investigated models that were composites of the previous family of models. This analysis showed that discs “which are hot enough to avoid axisymmetric instabilities can still evolve rapidly in a nonaxisymmetric manner.”\(^{390}\) Of these nonaxisymmetric instabilities, Kalnajs pointed out a particular mode \((m = 2)\) that corresponded to a bar-like instability for further study. He

---

390 Kalnajs, 1972, p. 74.
also noted that situating the discs in a rigid spherical halo was one method of stabilising the models.

As with Toomre’s criterion, Kalnajs’ work was viewed as a theoretical prediction of the bar-like instability, which was then later confirmed by Hohl’s simulation – in a later study of the same instability, Hohl stated that Kalnajs’ “stability analysis was confirmed (Hohl 1972) by numerical simulation techniques.” Furthermore, the ‘experiment-simulation tests theory’ narrative is circulatory in the same way as with Toomre’s criterion: while Hohl used Kalnajs’ theoretical analysis to show his bar-like instabilities were not mere numerical artefact, Kalnajs used Hohl’s simulations to provide punch to his own conclusions:

> It would be premature to conclude from our uniformly rotating models that stable discs must possess velocity dispersion in excess of that needed to stabilise them against axisymmetric disturbances. However the computer experiments on differentially rotating models, particularly those carried out by Hohl (1970), point to the same conclusion.392

Toomre’s criterion was theoretically predicted, confirmed by simulation, and then used by both simulations and theoretical work either as an initial condition, or as a validation of the model. Kalnajs’ work was forced to fit into the traditional narrative of ‘the scientific method’: a phenomenon is theoretically predicted, then confirmed (or falsified) by experiment. This was despite the fact that the analytical work done by Kalnajs was arguably simultaneous with Hohl’s simulations, and that both authors used each other’s work to validate their own.

Of more interest, perhaps, is the part simulation played in revealing new phenomena. The bars themselves were a ‘phenomenon’, though the term is used advisedly, in the negative sense that they were there when they should not be, or perhaps, when there was no reason to expect them. The bars were objects that needed to be explained away, and in the process were incredibly epistemically motivating. Miller struggled with the instability, ultimately coming to the conclusion (after a detailed analysis) that “numerical experiments may be telling the truth”;393 that is, in the simulated galaxy, the simulated bars were ‘true’. Later, there was a hint that the bars did have some reality, but of more pertinence to this chapter is the role they had in introducing the halo into the galactic model.

3.6 The Ostriker-Peebles Conjecture

Jeremiah Ostriker and Phillip Peebles provided the next contribution to the discussion, beginning their oft-cited 1973 paper with the statement: “there is some theoretical reason to believe that a highly flattened disc supported mainly by rotation is subject to large-scale (barlike) instabilities,

391 Hohl, 1976, p. 35.
393 Miller, p. 542.
whether the disc is composed of gas or stars.” The purpose of their paper was to investigate under what circumstances a cold rotating disc of stars could be stable, or could be made to be stable. Unlike MPQ and Hohl’s work, this simulation study was specifically designed to explore the bar-like instabilities. As such, Ostriker and Peebles had a different model for their galaxy; they introduced a ‘halo’:

The model is based on the numerical integration of the equations of motion in three dimensions for N mutually interacting particles. In addition to this flattened (but not flat) system, we suppose that there is a spheroidal component, which we call the halo, having an assumed mass distribution designed to produce a relatively level rotation curve. One issue of definition is important here: by ‘halo’ we mean ‘spherically symmetrical component,’ without prejudice as to whether the correct astronomical term would be ‘halo,’ ‘galactic bulge,’ ‘galactic nucleus,’ or some combination thereof. This galactic system was modelled as a uniformly rotating, uniform-density, fluid body. The spheroidal component – the ‘halo’ – was modelled using an axisymmetric sequence called the Maclaurin sequence. The reason for the inclusion of this halo was that the authors knew from previous work in fluid dynamics that a Maclaurin spheroid would display barlike instabilities beyond a critical value, instabilities similar to that uncovered by the previous simulations. The pedantry behind the definition of ‘halo’ is due to the looseness with which terms denoting galaxy components are used.

Like the other simulations discussed here, Ostriker and Peebles presented a brief description of the procedure at each timestep. Unlike the MPQ and Hockney-Hohl model, Ostriker and Peebles had a finite thickness for their disc, used fewer stars, and dealt with gravitational interaction in a different manner (through acceleration). The position, velocity and acceleration of each point at time $t$ was used to calculate the position at time $t + \Delta t$ via a fourth order polynomial. This new position was used with the previous numbers to allocate a new acceleration, which allowed a

---

394 Ostriker & Peebles, 1973, p. 467. Here one assumes that the authors mean there is theoretical, as opposed to observational, reasons for the bar-like instabilities.
396 Apparently Maclaurin spheroids have something of a reputation; Toomre began his 1964 paper by musing that: “the well-known instabilities of those Maclaurin spheroids whose rotational flattening exceeds a certain fairly moderate value suggest that other sufficiently flattened, rotating, and self-gravitating systems might in some sense likewise be unstable” (Toomre, 1964, p. 1217). Toomre at this point was not thinking about the bar instability per se, but rather the tendency of flattened discs to exhibit gravitational clumping.
397 The meanings of core, halo, nucleus, and bulge depend on context, and they are occasionally used as synonyms. Sometimes the core of the galaxy is called a bulge, due to the way that it bulges out beyond the plane of the galaxy on which the disc of stars lies, and for this reason it is also sometimes referred to as a halo. ‘Halo stars’ are older population stars which have non-circular angular orbits around the galactic centre, and which therefore do not lie on the galactic plane (but they are not part of the ‘halo’). Later, a ‘central’ component is added to the model, which is sometimes called the galactic core or nucleus, and refers to quasi-stellar objects (QSOs, or quasars). In modern parlance, the ‘halo’ (or more commonly ‘dark halo’) is a large, dark, mostly-spheroidal component in which the galactic disc is embedded, and which extends far beyond the luminous disc.
397 Hohl, 1975, p. 364.
corrected position to be obtained via a fifth order polynomial. This method was used to increase accuracy, which “is checked by comparing the results for different time steps”.\textsuperscript{398} In assigning velocities to the particles, Ostriker and Peebles added a velocity dispersion “designed to fit the Toomre (1964) criterion for stability against the development of small-scale irregularities”;\textsuperscript{399} Toomre’s criterion having achieved a cemented status in galactic models.

To assist in determining how stable their models were, Ostriker and Peebles defined:

\[ t \equiv T_{\text{mean}}/|W| \]

where \( T_{\text{mean}} \) was the systematic rotation energy in the model (kinetic energy) and \( W \) was the sum of the potential energies (of interaction of the particles in the field of the halo plus the self-energy of the halo).\textsuperscript{400} As the system evolves, energy is transferred between states and the ratio \( t \) changes. By tracking this change, \( t \) can be used as a direct measure of the tendency to instability (when the rotational and potential energies are not in equilibrium).

Ostriker and Peebles ran their basic model – Model 1 – and discovered that, as expected, a large-scale bar quickly developed. The disc heated up rapidly and \( t \) fell to half its original value in only half a rotation period. What followed was classical experimental procedure: parameter variation, which allowed comparison with the ‘control’ of Model 1. Models 2-8 in which the parameters velocity dispersion, initial random distributions, cutoff value, surface density, and number of particles were individually tested, showed little difference in their overall outcomes: “among the parameters we varied, the only one that markedly changes the course of evolution of \( t \) is the halo mass \( M_H \).”\textsuperscript{401} Models 9 to 12 only varied the halo mass parameter, by increasing the amount of mass in the halo. Ostriker and Peebles found that the higher the ratio of \( M_H/M_D \) (halo mass to disc mass) the flatter \( t \) (see Figure 9), meaning that the more mass in the halo, the more stable the disc. There appeared to be a threshold value at which the system was stable, and when \( t \) began above this value it would be expected to approach the threshold asymptotically. Though Ostriker and Peebles could not unequivocally demonstrate this, “in any case, it appears that models having \( t \) initially \( \lesssim 0.15 \) change slowly and are only weakly unstable if at all.”\textsuperscript{402} Models 11 and 12, for example, appeared quite stable and exhibited no bar-like instabilities. In summary: “the cold models without halo exhibit a violent instability that we cannot relate to any peculiarities of the model save the absence of a ‘hard’ component in the potential. . . . For the chosen forms of density distribution in disc and halo components, a halo mass of 1 to 2½ times the disc mass appears to be required to reduce the initial value of \( t \) to the stable range \( t \approx 0.14 \).”\textsuperscript{403}

\textsuperscript{398} Ostriker & Peebles, 1973, p. 470.
\textsuperscript{399} Ostriker & Peebles, 1973, p. 470.
\textsuperscript{400} Note that this \( t \) is different to time \( t \), which is later redefined by Ostriker and Peebles as \( \tau \) and measured in units of orbit time for particles in the outermost part of the disc.
\textsuperscript{401} Ostriker & Peebles, 1973, p. 476.
\textsuperscript{402} Ostriker & Peebles, 1973, p. 477.
\textsuperscript{403} Ostriker & Peebles, 1973, pp. 477-478.
The outcome of the work was essentially that a halo component with not insignificant mass needed to be added to the model of a galaxy to ensure stability – the halo was 'hot', and so reduced the bar instability in a similar qualitative way to increasing the velocity dispersion in the disc, though it had the advantage of not directly contradicting observational evidence. Ostriker and Peebles were aware of the contentious nature of this proposition, and assured their readers that "on the basis of the various checks we have made and the parameter studies described, we believe that the instability is not an artefact of some special errors in these calculations". Having asked MPQ and Hohl to provide their own calculations of the critical value of $t$, Ostriker and Peebles were also able to demonstrate that their conclusions were in general agreement with the previous notable simulations that modelled the galaxy as a whole. In addition, Ostriker and Peebles admitted that, given the theoretical motivation for including a halo, it was possible their critical value should be closer to that for Maclaurin spheroids (0.2738 rather than 0.14) and that:

Numerical errors and small values of $N$ combine to give so much relaxation that the computed $N$-body systems simulate a high-viscosity fluid. Thus it is extremely important

---

404 Recall that higher the amount of random motion, or the greater the velocity dispersion, the more 'hot' a system is.

that there exists one exact study of the stability of a disc of stars to nonaxisymmetric modes, that of Kalnajs.⁴⁰⁶

Kalnajs was summoned once again to play the role of theoretical prediction. This followed a similar pattern to Toomre's criterion, where Kalnajs' analytical work was sufficiently 'external' to lend weight to the results of the non-analytical simulations.

The problem was that a cold disc model of a galaxy, one that matched observations of the stellar rotations in flat spiral galaxies, exhibited instabilities. To smooth out the axisymmetric instabilities, the disc had to have a minimum velocity distribution, to be at least a little 'hot'. However, to smooth out nonaxisymmetric instabilities, the disc would have to be far too hot (with \( t = 0.14 \)) to match observations. Ostriker and Peebles proposed a solution that added a hot component to a cold disc in the form of a halo with a mass that was a significant portion of the whole system's mass. The methodical variation of parameters over several models is a clear example of the 'experimenting on models' procedure with which simulation is associated. The simulation actually dealt with several different models, variations on a theme that helped identify which parameter effected the bar forming tendency the most. However, unlike a normal experimental system, the simulation introduced a new (virtual) object in order to resolve the (virtual) phenomena. This was not a popular move.

3.7 The Problem with Halos

The problem of galaxy stability was a significant one, as while computer simulations were invaluable in exploring the theory, they also revealed how disparate, piece-meal, and inadequate our understanding of galactic dynamics was. Toomre, in a review article in 1974, described the problem as "a near-scandal in our understanding of galactic structure that has surfaced unmistakably only during the past year or two. ... It raises doubts even about fundamental assumptions [and] it seems unlikely to vanish overnight."⁴⁰⁷ He identified four main "analyses or numerical experiments"⁴⁰⁸ that revealed the bar-like instabilities, and which have been discussed in this chapter – Miller, Prendergast, and Quirk, 1970; Hohl, 1971; Kalnajs, 1972; and Ostriker and Peebles 1973. The last article Toomre described as unifying the other three investigations, as the Ostriker-Peebles result of \( t = 0.14 \) characterised the coolest of all the largely pressure supported discs of the studies: "though of course it is no proof that every conceivable model disc must be as hot to be fully stable, this astonishing numerical agreement means at the very least that the Ostriker-Peebles criterion is an excellent rule of thumb summarizing all the available evidence."⁴⁰⁹

The Ostriker and Peebles article had the specific aim of attempting to model a galaxy without a bar-like instability. MPQ demonstrated that 'machine calculations' of the existing model produced

---

⁴⁰⁷ Toomre, 1974, pp. 460-461.
⁴⁰⁸ Toomre, 1974, p. 461.
⁴⁰⁹ Toomre, 1974, p. 462.
discs that were too hot, and had to be artificially cooled. Hohl uncovered a bar-like instability in existing models, and was only able to stabilise the disc by making it unrealistically hot. Kalanjs provided a theoretical basis for the existence of bar-like instabilities, but could only demonstrate the instability via analysis, and not resolve it. Ostriker and Peebles were able to provide a criterion for how much heat (on one interpretation of the $t$ ratio) was required for a disc to be stable, and it was here that Toomre saw the work’s real value. Miller agreed with this: “the Ostriker-Peebles conjecture appeared to systematize the results of various studies which had been difficult to intercompare. The stability question could no longer be dismissed”,410 Ostriker and Peebles had helped astrophysicists realise the gravity of the problem of finding a stable, realistically cold, galaxy model, a problem that could “seriously plague all large-scale galactic dynamics”411 until it was resolved.

The Ostriker-Peebles criterion (or ‘conjecture’, as Miller would have it) would never achieve the same status as Toomre’s, remaining, at best, a good rule of thumb. The reason for this was the proposed introduction into the galactic model of “major stellar halos, such as Ostriker and Peebles have already suggested tentatively”.412 The suggestion of a halo remained contentious largely because the halo required such a large percentage of the total mass to stabilise the system. Hohl, in 1975, produced further simulations that involved a rigid core or halo component.413 Hohl added to his original model an axisymmetric, fixed radial gravitational field in order to simulate the effects of a hot and time-independent core or halo component. Varying the mass of the halo component, Hohl found that “the bar-forming instability is prevented for systems with up to 40% of the mass in the disc stars”.414 More than half the system mass must be in the halo, therefore, to prevent bar-like instabilities.

With regards to the plausibility of large mass halos, Hohl was reserved but optimistic:

> Even though there is little direct observational evidence for massive halos (Freeman 1974), there appear to be no difficulties with the assumption that a large fraction of the galactic mass is contained in the halo (Oort 1965), likely consisting of high-velocity-dispersion objects.415

Hohl meant that there are technically no theoretical difficulties with such an assumption – Jan Oort had argued, with regards to the total mass of (our) Galaxy, that:

> There is no way for estimating how much more mass there may be in the form of intrinsically still fainter stars. The real mass of the halo therefore remains entirely unknown. It is quite possible that there might be enough halo stars to make the halo an

410 Miller, 1978a, p. 812. a
412 Toomre, 1974, p. 463. Arguably, there is little that is ‘tentative’ about the manner in which Ostriker and Peebles present their idea of the halo.
413 In Hohl’s simulations, “since the radius of a fixed spherical component is varied over a large range, the terms core and halo are used interchangeably at times” (Hohl, 1976, pp. 31-33).
414 Hohl, 1975, p. 364.
important contributor, or even the most important contributor to the mass of the Galactic System. The uncertainty concerning the relative contributions of the halo and the disc to the total mass is the greatest obstacle in the way of constructing a model of the mass distribution in the Galaxy.\textsuperscript{416}

This was in 1964, and the state of observational knowledge was little changed a decade later: it was theoretically possible that there was large amounts of non-luminous matter with high velocity dispersion (‘hot’) in the galactic core or halo. Possible, because there was no evidence suggesting it was not possible. Ken Freeman, however, saw the lack of observational evidence in a very different light. He argued, in 1973, that “we see flat systems ... with no significant visual bulge component. ... The smallness of the bulge component argues against the presence of a massive stabilising bulge, unless the $M/L$ [mass-to-light] ratio for the bulge is very large.”\textsuperscript{417} There was evidence that this mass-to-light ratio (for the central bulge) is not very large. Oort countered, in discussion to Freeman’s paper:

\textit{Oort:} I do not know of any observational evidence opposing the idea that a large fraction of the mass of a galaxy is contained in a halo. ... Who knows what the missing mass is made up of? It may be very faint sub-dwarfs.

\textit{Freeman:} How faint would you want to go?\textsuperscript{418}

Oort, of course, wanted to go very faint indeed. The missing mass he referred to was from studies of galaxy dynamics where mass determined from observations of stars and other luminous objects was insufficient to hold the observed galaxies together gravitationally.\textsuperscript{419} From Oort’s point of view, there was already the long standing suggestion that there was a large amount of so far undetected matter within galaxies (this later became ‘dark matter’), and so the idea that this missing matter was located in a halo postulated to resolve stability issues was not so problematic. For Freeman, the two problems (of missing mass and galaxy stability) were not necessarily connected, and he was not alone in his discomfort with the idea of a massive halo. Toomre summed it up in a discussion to his 1974 summary paper: “Yet it also seems to me that the really tough nut to crack is not whether a massive halo is plausible, but to establish that it indeed must exist.”\textsuperscript{420}

Miller, in 1978, also seemed to be unhappy about the idea of a halo. Firstly, he stated that “no satisfactory dynamical model of the Galaxy has ever been constructed.”\textsuperscript{421} While Ostriker and Peebles “forcefully argued for the halo possibility”,\textsuperscript{422} Miller considered the entire situation fraught with questions and implications for galaxy studies. What about dynamical effects, the

\textsuperscript{416} Oort, 1965, p. 793.
\textsuperscript{417} Freeman, 1974, p. 134.
\textsuperscript{418} Freeman, 1974, p. 136.
\textsuperscript{419} Oort, 1932 and Oort, 1960. See also Zwicky, 1937.
\textsuperscript{420} Toomre, 1974, p. 464. Emphasis in original.
\textsuperscript{421} Miller, 1978a, p. 811.
\textsuperscript{422} Miller, 1978b, p. 32.
mechanism by which the halo provides stability, or develops spiral patterns? In addition, “if a massive halo can stabilize a disc galaxy, how massive must it be? If it must be many times as massive as the disc itself, the hypothesis is less appealing.”

In looking at the fraction of mass and velocity dispersion in the active disc, Miller’s simulation found that the halo would stabilise a galaxy, though he remained uncomfortable with how large the values of halo mass and velocity dispersion would need to be. Miller in fact arrived at a similar formal conjecture to that of Ostriker and Peebles, though he was sceptical about the degree to which this validated the \( t = 0.14 \) threshold: “the remarkable numerical agreement must be fortuitous, but it provides a welcome check on both sets of experiments. It also shows that the numerical values found here are surprisingly model-independent.”

Independent analyses which corroborated the Ostriker-Peebles criterion were not sufficient to validate it, as they were for Toomre’s criterion; at best Miller saw his work as validating the procedure and model involved in the simulation.

While Zang and Hohl agreed that the addition of a rigid halo of roughly the same mass as the disc was a “less drastic cure for this violent instability” than allowing velocity dispersions in excess of that observed in the solar neighbourhood, the halo remained an unpopular choice. In their study, Zang and Hohl referred to the possibility of retrograde stars providing stability (an idea resulting from an analysis done by Kalnajs) as “a more attractive cure than massive halos ... there seems little question that at least the central regions of galaxies contain substantial numbers of retrograde stars, in contrast to the more debatable observational evidence for massive halos”.

The outcome of their simulation supported Kalnajs’ theoretical conclusions – that is, more retrograde stars reduced the severity of the bar-forming, though the distribution had to be selected carefully. While halos were to be preferred over demanding high velocity dispersions which directly contradicted observational evidence, there was still insufficient empirical evidence for such halos, which drove astrophysicists to consider other possibilities. Another study suggested “the first hopeful sign that very massive halos may not be essential for stability” involved immobilising the central regions of the disc “in some plausible way”. Despite the fact that it was found that the rest of disc had no unstable formations, “exceptions to the stability criterion postulated by Ostriker and Peebles remain distressingly difficult to find.”

One of the reasons for this reluctance to the Ostriker-Peebles criterion and its implied halo was the puzzle of spiral structure, which astrophysicists had been grappling with at around the same time – what was the mechanism behind the persistence of spirals, and how did they form? The galactic model had to support spiral structure, but “an initially axisymmetric system stabilized by a halo cannot develop the nonaxisymmetric form, so spiral features cannot form.” Let’s take a

423 Miller, 1978b, p. 32.
424 Miller, 1978b, p. 37.
430 Miller, 1978b, p. 36.
brief excursion into this issue of spiral structure. In the mid to late 1960s, it was suggested (by Chia-Chiao Lin and Frank Shu) that the spiral arms in galaxies were actually density waves, mostly affecting the gas (and young stars) in the disc of the galaxy.\footnote{As a start, see Lin \& Shu, 1966. The idea also has origins in Bertil Lindblad's work (see Toomre, 1977, pp. 438-442).} Density waves, as a lecturer once explained to me, function like traffic jams on the highway of the galactic plane. Stars and gas traverse their orbit, but when they come to one of these areas of higher density they bunch up. Once clear of the traffic jam, the stellar objects resume their more spaced out distribution. Stars do not, therefore, orbit in a spiral pattern, but pass through spiral density waves on circular orbits. The idea caught on quickly, but at the time of bar-instability studies "it is not just the origin but even the persistence ... of these waves that remains distressingly unexplained".\footnote{Toomre, 1977, p. 449. Emphasis in original. To some extent, this appears to remain true to today.} The best guess was that the density waves were the result of either galaxy-wide disturbances, or smaller scale axisymmetric instabilities which were hard to detect but whose net effect was spiral density waves.

Toomre, in a 1977 review, looked at the state of galactic knowledge with the aid of a simple diagram (Figure 10). Point B was where researchers landed when they conjectured that there was a wavelike spiral forcefield that somehow exists and keeps on existing. The path up to C was learning how the gas in galaxies behaved given these density waves, which seemed to be reasonably well understood. The path up to D – star formation – was poorly understood, as was the path down to A. This latter point, which Toomre described as "first principles", concerned the origins of the spiral structure. The simulations discussed in this chapter were attempts to shed light on A – to describe a galactic model that was stable and could host persistent spiral patterns.

This image has been removed by the author of this thesis for copyright reasons.

\textit{Figure 10: "Status of the wave theories. As yet, only the thickened portions of this schematic route seem reasonably secure." (Toomre, 1977, p. 463).}
The problem of finding an adequate solution to the bar-instability problem remained, further compounded by issues of spiral structure. Any resolution, furthermore, was not going to be found in the tried-and-true mathematical analyses, but rather in the simulations that had revealed the depth of the problems in the first place:

With the sole exception [of three mode studies] using only pencil, paper, and Legendre polynomials – all existing studies of the very large-scale or ‘global’ behaviour of disc-like model galaxies have been heavily numerical. Some of these studies, devoted again to linear instabilities and/or modes, have at least made a mild pretense of analysis. Others have amounted frankly to brute-force time-integrations of the nonlinear equations of motion of up to about $10^5$ gravitationally interacting particles, with their only elegance consisting of various ‘checkerboard’ schemes for rapidly solving the Poisson equation. As might be expected, the mode searches have tended to be much less expensive and, within their limited domain, also more accurate than these large N-body experiments. Yet it is indeed the latter that have taught us far more of what we really needed to know.\(^{433}\)

Despite the inelegance of brute force simulations, Toomre could not deny that the way forward was through simulation, which was the only tool available to astrophysicists that could fully evolve the theoretical ideas of (in this case) galactic models, providing them with “the kind of firm deductive basis that one likes to associate with problems of dynamics”.\(^{434}\) Furthermore, Toomre’s problem seemed to be one of aesthetics, rather than epistemics: simulations are large and cumbersome, and therefore mathematically inelegant, but it is this complexity that allows ‘experimentation’ and for seeing the full implications of one or more theoretical assumptions on the rest of the model, or further down the line in time.

In this case study, simulations began by following an experimental narrative, providing confirmation of a theoretical prediction of axisymmetric instabilities in the early galactic model. From then on, models of galactic discs were not complete without a minimum velocity dispersion matching Toomre’s criterion. Following this, simulation work then revealed a further substantial bar-making instability, which followed less closely the experimental narrative by both confirming and being confirmed by more traditional analytical work: “it was suddenly beyond all reasonable doubt that the various investigations were telling essentially the same story – and that it was not a story of numerical errors.”\(^{435}\) The search for a stable model, from then on, occurred almost entirely within the realm of simulation – arguments for and against a halo, studies of possible stabilising forces, and modelling alternative structures all occur within the simulation arena. Genuine scientific knowledge about the composition and evolution of galaxies was being generated solely by means of simulation techniques.

\(^{433}\) Toomre, 1977, p. 464.


\(^{435}\) Toomre, 1977, p. 469.
3.8 Resolving the Instability

Introducing a gas component, adding an inert halo, increasing velocity dispersion, and adding retrograde stars, “these modifications, alone or in combination, substantially reduced the barlike behaviour. Nevertheless, simulation of a stable dynamical Milky Way-like galaxy has not been achieved.”\footnote{Comins, 1983, p. 595.} How then, was a stable model achieved? And on a related note, if the halo was not a popular addition to the galactic model to resolve stability issues, how did we end up with a modern-day model that features prominently the existence of a massive dark halo?

Figure 11: Evolution of a leading density wave in Zang’s simulation, showing the growth of spiral pattern through swing amplification (Toomre, 1981, p. 125).

The answer to the first question came via studies of spiral patterns. In 1981, Toomre wrote on several simulations that explored a mechanism for the bar instability in spiral patterns called ‘swing amplification’.\footnote{The amplification process was first explored in the mid-1960s by Goldreich and Lynden-Bell, and Julian and Toomre, as part of explorations of spiral formation.} Essentially the idea is as follows: density waves can reflect off both the centre of the galaxy and the corotation circle (the circle around the galactic centre where the stars
move at the same speed as the spiral arms). This allows a standing wave to be set up. Standing waves are composed of two travelling waves: leading waves propagate outwards, trailing waves travel inwards. The trailing waves reflect off the galactic centre and turn into leading waves, and at the corotation circle switch back to trailing waves. As the waves reflect off the corotation circle they are sheared by the differential rotation into spiral waves (see Figure 11).

The decay of the density wave in frames 7-9 is due to inner and outer Lindblad resonances (ILR and OLR) which occur at certain radii from the centre of the galaxy and, broadly, act to maintain the spiral patterns. If a star’s orbital speed around the centre of the galaxy is greater than that of the spiral wave through which it is passing, the ILR acts to move the star outward. If the star is slower than the wave, OLR acts to move the star inward. ILR tends to absorb arriving density waves, and so suppresses the excessive growth of the spiral. However, in galaxies without ILR, there is no mechanism to suppress the swing amplification:

A key aspect of the instability mechanism is that amplified, ingoing, trailing waves are able to reach the centre where they can reflect into outgoing, leading waves. Toomre therefore proposed that if the centre of the galaxy should be made inhospitable for density waves, the feed-back loop would be cut and the disc would avoid this particularly virulent instability.438

Thus by “simply placing an absorbing plug in the centre of the disc”,439 the bar forming instability is prevented from occurring, and with “no extra forces from an imagined halo”,440 Toomre’s simulations showed clearly how swing amplification could develop and maintain spiral structure, the persistence of which had been an elusive goal for most earlier simulations of galaxies. The Ostriker-Peebles criterion, now downgraded to “something of a numerical fluke”441 resulting from an understandable but misguided intuition about similar instabilities in Maclaurin spheroids, was fully defeated. In support of his mechanism, Toomre demonstrated that translating high velocity dispersions or the addition of rigid halo – two ways of creating a stable disc – into the language of swing-amplification resulted in the same requirements his own study had come up with: “in retrospect, it is ironic that most of those experimental ‘confirmations’ of the Ostriker-Peebles conjecture can also be recognised as fine testimonials to swing amplifications!”442 Though the narrative of experimental ‘confirmation’ worked for Toomre’s own criterion, it failed for the Ostriker-Peebles conjecture because that latter postulation required that simulation ‘confirm’ not only a threshold value, but also the existence of a massive halo object. While simulation could play the role of experiment persuasively, it could not reify an entirely new object. Simulation alone could not provide sufficient reason that a halo was necessary and plausible. In fact, it seemed to indicate the opposite: swing amplification implied

441 Toomre, 1981, p. 131. Toomre does admit that he himself had fallen prey to this intuitive association of disc instabilities with those in Maclaurin spheroids.
that the halo solution was unlikely to apply to galaxies with bisymmetric features because, as simulations demonstrated, a dominant halo component inhibited swing amplification of all two-armed patterns, and instead caused nonbarred galaxies to prefer multi-armed patterns. Since the scientific community did not consider the Ostriker-Peebles conjecture valid, simulation ceased to play an experimental role and returned to being treated as theoretical. Toomre placed “experimental ‘confirmations’” in inverted commas because he recognised the metaphorical use of ‘experiment’ consciously.

Having uncovered the mechanism by which barred features form, and therefore knowing how to prevent their formation (by using a dense, central mass impassable to the density waves), why then did the halo stick around? The answer came in the form of observational evidence. In 1980, Vera Rubin and her collaborators published a study of the observationally determined rotation curves (velocity over distance) of 21 spiral galaxies, the same type of galaxies for which simulationists had been trying to obtain a valid model. This study showed something surprising: the rotation curve remained flat out to large distances from the galactic nucleus, far beyond the luminous disc of the galaxy (see Figure 12).

This image has been removed by the author of this thesis for copyright reasons.

Figure 12: “Superposition of all 21 Sc rotation curves.” ‘Sc’ refers to the morphological class of galaxies with a loose spiral structure. The top curve represents the average shape of a typical Sc galaxy. (Rubin, Ford, & Thonnard, 1980, p. 480).

Rotation curves were expected to decline at the edge of the luminous disc (approximately 10 kpc from the centre), because it was assumed that there was an approximately equal correlation between mass and light – if there was no light, there would be no mass (broadly speaking), and so the rotation curve should decline. That the curve did not decline, however, implied that there was some invisible mass in the galaxy (and not an insignificant amount) beyond the luminous

\[^{443}\text{Sellwood & Carlberg, 1984.}\]
Previously, the mass-to-light ratio was assumed to be constant; now it was found that the mass-to-light ratio in the further parts of the galaxy was much larger than in the interior parts.

Claims to an observationally determined flat rotation curve can be found as early as 1939, and would continue to crop up periodically, especially after the discovery in 1951 of the 21-cm line. In 1974, one year after their contentious halo-for-stability paper, Ostriker and Peebles (with Amos Yahil) published an article reviewing the evidence for large mass-to-light ratios, that is, the evidence that there was large amounts of mass where there was little stellar content. The authors suggested that:

The very large mass-to-light ratio and the very great extent of spiral galaxies can perhaps most plausibly be understood as due to a giant halo of faint stars. Such a structure, which appears superficially so improbable, has been proposed recently ... for quite different reasons concerning dynamical stability. The least troublesome way of ensuring the stability of a cold disc of stars against nonaxisymmetric disturbances is to suppose that the spherically distributed (halo) mass interior to \( r \sim 10 \text{ kpc} \) is substantial. ... If, further, the spherical halo exterior to 10 kpc is as extensive as the halo of an elliptical ... and the mass-to-light ratio stays large, then the mass associated with the outer parts of spiral galaxies must be large on these grounds.

The argument for the halo from galaxy stability was doomed, but the argument from mass-to-light ratios remained. Prior to 1980, in spite of this latter 'evidence', the presence of a halo had been so easily dismissed in stability arguments because there was little consensus at the time that the rotation curves were in fact flat; one astrophysicist later recalled, "the general feeling was that all instrumental and systematic effects had not been correctly taken into account." In the second half of the 1970s however, instrumental techniques improved and allowed rotation curves to be determined with higher resolution and sensitivity. The impact of these results – observations of multiple spiral galaxies all producing non-declining rotation curves – was strong, and, "in a period of two or three years (1978-1981), the phenomenology of flat rotation curves went from being a somewhat dubious result of radio astronomers to an increasingly accepted view of spiral-galaxy kinematics."

This was not, of course, unambiguous evidence in favour of galactic halos. The flat rotation curves implied a large amount of unseen mass at large radii, but did not provide information about its

---

444 This is derived from the formula for the mass of a point particle \( M(r) = \frac{v^2 r}{G} \) where \( M(r) \) is the mass enclosed in radius \( r \) and \( v \) is the velocity (\( G \) is the gravitational constant). If \( v \) is constant, mass increases linearly with radius: \( M(r) \propto r \).
445 Babcock, 1939.
446 The 21 cm line is an emission line of neutral hydrogen that allowed observation of rotation curves as far as the hydrogen extended, which was beyond the luminous disc of the galaxy. It was the 21-cm line that in fact allowed astronomers to determine that the Milky Way had a spiral form.
The historical primacy of the halo suggestion, however, combined with the new requirement for a space for all this missing mass to reside, meant that astrophysicists continued to use the halo in their analyses. As an answer to the bar instability, the Ostriker-Peebles criterion suffered defeat. Yet the halo found unexpected application in explaining the flat rotation curve. Simulation could not ‘confirm’ the existence of dark halos, but it could provide evidence that a galaxy embedded in a dark halo would produce the same flat rotation curve that had been observed. Now playing the role of ‘theory’, the simulated dark halo galaxies were supported by observational evidence. Simulations, in conjunction with that crucial, almost singular, piece of empirical evidence, were able to produce both data-like evidence and theoretical outcomes that resulted in the development of a galactic model that was stable and explained observed phenomena.

It is the capacity of simulation to play different epistemic roles – some more experiment-like, others more theory-like – that is crucial to its success as an epistemic tool.

---

This image has been removed by the author of this thesis for copyright reasons.

Figure 13: The total rotation curve for the galactic model is the top curve. The other curves show the contributions to the total curve from the four mass components. (Sellwood, 1985, p. 129.)

---

450 In attempting to provide answers to these questions, simulation continued to play a major role.

451 The case study in Chapter 4 has a very similar shape. In this chapter, simulation blends the roles of experiment and theory; in Chapter 4, simulation blends observational data and the ‘empirical content’ of theory.
Consider a simulation performed in 1985 by Jerry Sellwood. The model he adopted had three components: a disc of stars, a halo, and a central component (the bulge) which was modelled in two parts as a luminous spheroid and a dense central core. With a combination of all three stabilising mechanisms (halo mass, a small amount of velocity dispersion in the disc, and the dense central core), the model was “simultaneously bar-stable and able to develop gentle bisymmetric spiral structure.” Furthermore, the three components “are all essential, since when any one is removed, the distribution of matter within the disc is quickly rearranged.” Figure 13 shows the rotation curves of each component, and how they are all necessary to produce the total rotation curve of the general form determined by observation. The central component is responsible for the initial peak, near the centre of the galaxy. The disc contributes the most in the regions interior to 10 kpc, while the halo only contributes significantly in regions beyond this radius.

With all these components, a stable system with weak two armed spirals was formed. Sellwood demonstrated that removing the core resulted in a bar instability. Removing the halo resulted in a strong one armed instability, which Sellwood suggested was probably developed through same swing amplified feedback loop as the bar instability, but the nature of the mode being such that it could not be inhibited by central mass concentrations. The halo was included in this model in order to make up the right rotation curve, and also proved important for resolving a different stability though Sellwood noted, “it is rather ironic that the halo component of [the model] appears to be irrelevant to bar formation ... since control of bar instabilities was originally one of the prime reasons for postulating its existence”.

By 1999, “the mechanism for the bar mode ... and ways in which it can be suppressed are well understood.” At the beginning of this chapter, galaxies were modelled as self-gravitating, thin discs. Half a century later, galaxies were widely considered to have a dense centre and a massive dark halo in addition to that luminous disc (which of course, obeys Toomre’s criterion), and in all of these aspects simulation played a crucial role.

Sellwood and Moore’s 1999 simulation demonstrated another important aspect of simulation in astrophysical work. Using a halo-bulge-disc model, the authors this time evolved the galaxy over the growth of a central mass. Beginning with the disc and halo, it was found (unsurprisingly) that the disc formed a bar, and that as the central mass grew, the bar weakened and became shorter. Later in the evolution, it disappeared completely and did not return. Interestingly, “a notable feature of this experiment is that the matter rearranges itself in such a way that the rotation curve becomes approximately flat except perhaps for the sharp inner peak”. Sellwood and Moore cited Rubin et al. in proudly demonstrating that their galaxy created the rotation curve typical of spiral galaxies. Here again is the reversal of the role of simulation in the confirmation narrative –

---

452 Sellwood, 1985, p. 128.
453 Sellwood, 1985, p. 128.
454 Sellwood, 1985, p. 145.
455 Sellwood & Moore, 1999, p. 337.
456 Sellwood & Moore, 1999, p. 130.
in the presence of actual empirical data, simulation can perform more philosophically comfortably as 'theory' and simply demonstrate agreement with observation rather than being required to simulate an experiment.

Sellwood and Moore then went on to suggest “that every massive disc formed a bar in its early stages and that many, though probably not all, of these bars were destroyed by the development of a massive object in their centres.” The massive central component is a QSO – quasi-stellar object – or quasar. These are incredibly luminous objects that emit large quantities of radiation. Modern consensus is that the quasar is powered by the large amounts of mass accreting into the central supermassive black hole that is thought to exist at the centre of most galaxies; the idea has its own history which would no doubt be complementary to this one. Sellwood and Moore suggested that some of the mass that forms the bar collapsed to power quasars, which would mean that the quasar epoch is coincident with the formation of bars in discs. Therefore Sellwood and Moore suggested supermassive objects should be found at the centre of almost every bright galaxy.

Part of what this case study reveals is that developing a galaxy model involved understanding the galaxy’s evolution – for which simulation is almost sine qua non. Sellwood and Moore chose to evolve a continuous growth of the central mass in their simulated galaxy, revealing how matter rearranged itself to develop into the correctly flat rotation curve, and thereby incorporating historical explanation into the model. At the end of the published conference proceedings of a 1989 meeting focussing on the dynamics of astrophysical discs the ‘theorist’s summary’ concluded that the perspective of astrophysics has shifted: “an important consequence of this change in perspective is that there is now less emphasis on a priori predictions of the spiral structure that should be present in a given galactic disc.” We know broadly how spiral structure can come about, “but the prediction of the strength, shape and origin of the spiral pattern in a given galaxies is usually too difficult a task”. There had been a shift towards a more historical or evolutionary style of explanation.

On partial differential equations (PDEs), M. Norton Wise describes the explanatory narrative that connects the mathematical quantities with real world phenomena as traditionally deductive. When using simulation to solve the more complex PDEs, however, the explanation is obtained only through a time evolution, creating a historical or evolutionary narrative. The simulated

---

457 Sellwood & Moore, 1999, p. 125. This is not enough to explain the presence of barred spiral galaxies that occasionally occur: the bar instability has too short a lifetime for us to expect to observe it ‘in action’. The authors therefore suggest that a second, longer lasting bar forms in some galaxies due to large perturbations from tidal encounters (with other galaxies), or with the accretion of extra material that could ‘re-excite’ the initial bar. Though this second bar lasts longer, it is still a temporary phase in the early evolution of the galaxy, and is expected to be stabilised into a regular spiral pattern.

458 With the exception of galaxies observed to have a gently rising rotation curve, not a peaked one. These galaxies are dark matter dominated, because they would not have formed the bars required to power the quasar.


460 Tremaine, 1989, p. 236.
solutions to the PDEs consist of all possible solutions, without proof of existence or uniqueness, "they are objectified and materialized, both in the mathematics and on screen, through the languages in which their evolution is written. In this sense, the distributions that emerge (the explananda) are essentially historical objects; furthermore, they can be understood only through their histories (explanans)." This is a long way from the physical object understood as a timeless, mathematical thing; replacing the language of deduction with the language of evolution – consistently seen through the construction of the galactic model – means that the simulation object is a historical object, and its virtual history its explanation (see section 5.4.1).

3.9 Persuasive Pictures: Simulation and Observation
This brings the narrative of how the bar instability was resolved and how the halo came into our modern galactic model to an end, though much more work was subsequently done on the shape, distribution, and content of the halo, all within the realm of simulation. The main twist of this narrative was that the halo, although originally suggested to prevent the bar instability, was found to do nothing of the sort, and instead remained in the model thanks to the intervention of observational evidence; in 1977 Toomre, with respect to stability studies, wrote that "the need for proper halos has still not been established firmly in our subject – though this is neither to imply that they would be unwelcome, nor to dispute the growing evidence from flat rotation curves ... in favour of much extra mass of low luminosity." It seems clear that simulation has a good deal of epistemic force. Apart from revealing the severity of the bar instability (which theoretical analyses like Kalnajs' could only suggest was a problem), simulation was also instrumental in the resolution of the problem. Now part of the toolbox of astrophysicists, simulation was called upon to provide compelling visuals demonstrating clear, bar-less spiral structure that was the result of a central mass, cementing the addition of this component into the galactic model. However, the demonstration that a halo was not required for this purpose was not quite forceful enough to eliminate it from the model (requirements for preventing other instabilities – like that in Sellwood's 1985 simulation – aside). Simulation does not have the empirical force of observation. Simulation was granted 'confirming' power in the case of Toomre's criterion because the result of the criterion neither contradicted the (scant) empirical evidence nor was overly speculative. Confirmation was reserved, however, in the case of the Ostriker-Peebles conjecture, for while it resolved an instability that did contradict observation, it also implied a highly speculative and empirically unconfirmed object – the halo.

It was only until compelling observational evidence arrived – in the form of flat rotation curves – that the halo began to gain firmer ontological footing. Though there was evidence for the presence of a good deal of non-luminous matter, flat rotation curves demonstrated that most galaxies had a large amount of this matter within them, at larger radii. The halo adopted the identity of a dark halo, providing a location and approximate distribution for all this dark matter. There is now

consensus that the dark halo is real despite the fact that both the halo and its contents are only indirectly empirically confirmed.

Empirical grounding is necessary – this is science, after all – but the relationship between empirical evidence and simulation ‘evidence’ remains complex. Rotation curves were compelling, thoroughly empirical evidence (though indirect). To demonstrate that the model worked, however, especially for other components of the galaxy (that a dense central core generates the flat, stable, bar-less disc), simulation provided the only ‘evidence’ in the form of images that matched what galaxies are supposed to look like based on our observations. Throughout the simulation reports, visualisation was incredibly important. Wise emphasises that the remotely performed mathematics cannot deliver the simulation’s historical explanation by itself:

Visual images have always been crucial in physics to guide intuition and reasoning and to illustrate problems and solutions. But the role of visualization, and of languages of visualization, in the explanatory narratives generated by simulations is qualitatively different, for it typically serves as the direct representation, and the only effective means for understanding, the growth process and its intricate results. The simulation, to be understandable, must incorporate a technology for converting the calculations performed by the cellular automaton into an object accessible to the senses.463

Simulation must, in simulating experiment, also simulate our observations of it, if you will, “engaging the visual gestalt capacities of embodied humans.”464 Actually seeing bars form, or spiral structure emerge through the evolution of the model turns a disc of dots into a galaxy, and convinces the audience that the outcomes of the analysis pertain to real galaxies much more vividly than arguments defending the choice of particular assumptions, or error analyses. Importantly, these simulated pictures resemble those taken of actual galaxies with telescopes, observations of the reality astrophysicists are trying to understand. Hockney noted, in one of the earliest and most unphysical simulations of a galaxy that:

Comparisons between real galaxies ... and the results from the computer model of rod-like ‘stars’ ... must be viewed with considerable caution. However, one cannot help but note the striking similarity between the computed structure after one rotation and the whirlpool galaxy (M51).465

Though a galaxy made of rod-shaped stars was a plainly unphysical approximation, the pictures still remain persuasive. The pictures are even more persuasive if they move: “the most dramatic display of the computed results is in the form of a 16-mm computer-made movie which shows the space and time development of the star system.”466 This film was shown at a meeting of the Astronomical Society of the Pacific and at an International Astronomical Union (IAU)
colloquium, and Hockney gained special mention in the meeting report as giving “one of the most entertaining talks”. A couple of years later, at another IAU symposium, both Miller, Prendergast, and Toomre (on behalf of Hohl), presented their results as a ‘motion picture’. At such conferences, the films would have enabled the witnessing of the evolution of the model into something distinctly galaxy-like to be extended beyond the authors and the mere stills of their paper. In a summary of the 1969 symposium, Bok describes the presentations as ‘spectacular’, and writes that “in the film we saw trailing spirals, bars and rings, forming and disintegrating before our eyes.”

Keller makes the observation that “CD-ROMs and the Internet, I suggest, give new meaning to Shapin and Schaffer’s notion of ‘virtual witnessing.’” Experimental biologists, in Keller’s work, use the computer’s representative capabilities to convey experimental results in a format “that is accessible and persuasive to an audience of experimental biologists who may be unable to follow the underlying technical analysis”. Hockney’s film would have both showed the astrophysics qualitatively, and shared with the audience the simulators’ own belief that their simulations are faithfully galaxy-like. The movement of these images is even more compelling; Miller made several references to the moving images that his simulation produced, and which he could only reproduce in print with a series of selected stills. These motion pictures “show the evolution of the systems more effectively” and “leave a distinct impression of persistence and of a nearly rigid rotation of the spiral structure”, and “especially in the motion pictures can individual ‘particles’ be seen passing through the pattern.” In the publication of his 1970 presentation, Miller (probably needlessly) writes that “the spiral patterns look like real galaxies.” The experimental biologists in Keller’s work use moving images in much the same manner: “the illusion of veridicality is made even more compelling by the presentation of these three-dimensional images in time.” Miller’s motion pictures have done a good job of convincing their author of the reality or appropriateness of his model; sharing them invites an audience to have this same conviction.

In the case of simulation, however, virtual witnessing extends beyond the audience. Robert Boyle, in Steven Shapin and Simon Schaffer’s well-known historical work on the air pump, used the technique of virtual witnessing to generate belief in his hard-to-replicate experiments. The experiment was described sufficiently vividly to develop in the reader’s mind an image of the experiment, clear enough that it was as if the reader was there: “the validation of experiments,

467 Hockney, 1968, p. 667.
468 Scarfe, 1967, p. 520.
470 Bok, 1970, p. 466.
473 Miller et al., 1970b, p. 905.
474 Miller et al., 1970b, p. 907.
475 Miller et al., 1970b, p. 907.
476 Miller et al., 1970a, p. 366.
and the crediting of their outcomes as matters of fact, necessarily entailed their realization in the laboratory of the mind and the mind’s eye.”478 The snapshots of the virtual galaxies taken in sequence in an astrophysical paper serves the same virtual witnessing purpose as Boyle’s detailed etchings of an air pump experiment: “by virtue of the density of circumstantial detail that could be conveyed through the engraver’s laying of lines, they imitated reality and gave the viewer a vivid impression of the experimental scene.”479 Hockney’s film made his distinguished audience witnesses of his simulation experiment because they were shown the evolution of a galaxy, the replicated image – or rather, the exact image – of what the author created, manipulated, and witnessed themselves. But the visual capabilities of simulation also produce in the scientist’s mind “such an image of an experimental scene as obviates the necessity for either direct witness or replication.”480 Simulation must be persuasive in this manner to be able to stand as a metaphoric experiment – the images of the technique must be compelling enough that direct witnessing of the ‘experimental scene’ is not necessary for justified belief in the results of the simulation. The simulated images function as empirical output, as observations of the evolved galaxy; observations which can only be conducted via virtual witnessing, as we cannot control and evolve galaxies in real life.481 Naturally, this is not the main reason simulationists have such faith in their own simulations. Through creation and continuous adjustment, the scientists have built up months of familiarity with their code, but this faith is not readily communicable. Anything new that comes out of the code is, from the perspective of the scientist, less of a surprise than an anomaly they have, through familiarity, grown to believe in. From an outside perspective, however, it will be a surprise, and needs to be justified.482 Sharing the images or film of the simulation is a way of briefly inviting the audience to witness the exact same images the author had, and convey some of the faith in the results. Simulations produce images that are unavoidably galaxy-like, vividly empirical and highly persuasive.

So galaxy-like, in fact, are these images that sometimes simulation authors compare the simulated images with photographs of actual, specific galaxies taken from an atlas:

It is useful to compare the computational results with real galaxies in order to check that the model has some validity and to gain insight into the chronological evolution of real galaxies by comparison with the model results. In referring to real galaxies, the orientations considered are those of the pictures in the Hubble Atlas.483

---

478 Shapin & Schaffer, 1985, p. 60.
480 Shapin & Schaffer, 1985, p. 60. Shapin and Schaffer, concerned with traditional experiments, are here referring to the reader of Boyle’s correspondence, the original virtual witness. In simulation, however, even the experimenter is a virtual witness.
481 Necessarily ‘witnessing’ in the case of the simulator expands beyond passive observation – the ‘witnessing’ of a virtual galaxy involves not only the manipulation but the creation of the object. One may forgive this stretched use of the word, but I wished to preserve Shapin and Schaffer’s notion of the virtual witness constructing the experiment in their mind’s eye, without ever actually seeing the real thing conducted.
483 Hockney & Brownrigg, 1974, p. 356.
Hockney went on to compare photographs of various real galaxies to stills taken from various stages in the simulated galaxy's evolution, intending both to demonstrate support for his simulation and provide more information on the stages of galaxy (real and simulated) evolution. Hohl did the same thing; Figure 14 shows quite persuasively the similarities between the computer generated galaxy (left) and an Sc type spiral galaxy from an atlas (right).

Figure 14: Hohl compares the simulated image (left), and the atlas photograph (right) by placing them side by side (Hohl, 1975, p. 364).

After demonstrating, however, that a snapshot of a dynamic computer galaxy can look like an instance of a static observed galaxy, Hohl then cautioned:

Such a comparison should not be taken seriously, since for any computer generated galaxy one can find a similar looking object in the various collections of galaxy photographs. Note that the computer generated galaxies considerably change their structure during the following rotation whereas we expect the actual galaxy to preserve its overall structure.484

Presenting the images alongside this text may have been intended to provide a cautionary example of how the persuasive visual power of simulations may mislead a scientist into drawing erroneous conclusions about his model and how it represents reality. After all, simulated galaxies – as evidenced by the bar unstable galaxies – can radically change their shape over the course of a few rotations if the model is faulty. This interpretation of Hohl's words would be in line with the fear of seduction discussed in the previous chapter that is subtle but persistent in simulation discourse; it could even be some sort of foreshadowing of the faulty intuition that lead Ostriker and Peebles to equate bars in Maclaurin spheroids with bars in their galaxies, an intuition that

484 Hohl, 1975, p. 365.
was almost certainly visually motivated. The compelling nature of the comparison, however, 
immediately followed by the words 'such a comparison should not be taken seriously' makes it 
seem like Hohl was contradicting himself – his actions belying his words.

Why, then, do these scientists make comparisons between simulation pictures and photographs 
taken by the Hubble telescope, if they recognise the specious nature of the gesture? In 
aviastrophysics, the majority of empirical data comes from various forms of observation, but 
unfortunately data is often ambiguous, poorly resolved, or simply absent. The theoretical 
ambitions of astrophysicists sometimes overreach the available observational technology, and 
they cannot always subscribe to the scientific ideal of testing theory with real data. In the absence 
of observational evidence, astrophysicists do what they have always done, and make do with 
other forms of validation, which after the advent of powerful computers includes simulation. This 
was evident in the search for a stable galaxy model:

Unfortunately, there are few observational results that bear directly on stability problems. 
Two approaches are available: theoretical investigations and numerical experiments. 
Theoretical investigations involve more drastic compromises than numerical experiments, 
so they are not as close to the physics of real galaxies. ... Much sharper comparisons can be 
made with theoretical investigations than with observation. 485

In the absence of observation to provide information one way or the other, theory and simulation 
are the main options for astrophysicists. Simulation, in particular, is able to provide the non-
empirical equivalent of empirical evidence, which is the next best thing when faced with a scarcity 
of real data. The countless simulation movies and stills in research papers attest to this, 
generating a sort of empirical-like evidence; the visual comparison of real and simulated galaxies 
could go one step further and provide some sort of indirect empirical support.

In understanding the role of rhetoric in developing convincing simulations in astrophysics, 
Mikaela Sundberg notes that tests of the simulation’s stability against different resolutions, 
particle numbers, and parameters "is sufficient for the producer of a simulation to recognize and 
trust its results, but it is insufficient as evidence to convince others."486 Therefore, other tactics 
must be used to increase belief in the feasibility of a simulation to external viewers; an important 
part of this is convincing the audience that the results are 'real' rather than 'numerical'. This may 
be done through connecting the simulation output with equations or by comparing the results 
with observation.487 One way of doing this in print is to compare simulation results to atlas

485 Miller, 1978a, p. 812. a
486 Sundberg, 2012, p. 73.
487 Sundberg differentiates between idealised and realistic simulations - in the former, more focussed on 
understanding the physical processes for a phenomenon, connecting the results of the simulation to 
equations is sufficient to make the underlying model convincing. In the latter, which is intended to capture 
complexity, connection to observation is necessary to 'justify' the simulation. From Sundberg's point of 
view, this would make the simulations in this chapter 'realistic', though they are also idealised in the sense 
of ignoring factors like gravitational effects from nearby galaxies, neglecting gas, reducing the number of 
particles, etc.
pictures, but from Sundberg’s point of view this sort of tactic is more rhetorical than practical: “astrophysicists speak of observations as principally synonyms with ‘truth,’ ‘reality,’ or ‘Nature,’ but scarcity of observations commonly justifies the need for numerical simulations in the first place, and observations cannot always be used as guidance.”\footnote{Sundberg, 2012, p. 76.} Another study noted that the authors of a supernova simulation used more traditional rhetorical devices to convince others of the reasoning behind their ad hoc decisions, as physical evidence, while useful to scaffold the work in terms of previously established facts, was insufficient to prove or justify their mechanism by itself.\footnote{Roundtree, 2010.} In differentiating between ‘numerical’ and ‘real’ effects, astrophysicists are understanding ‘real’ as either what the effect derives from – the demonstrative result of a physical equation – or “whether the effect is plausible in relation to what could be expected from observations.”\footnote{Sundberg, 2012, p. 79.} Hence Hohl and Hockney’s reluctance to ascribe too much persuasive power to the similarity between atlas pictures and the simulation results is due to the scientific tempering of classical rhetoric: to convince their audience that the simulation is realistic, there must be some connection to an empirical reality, but due to the dearth of actual observation evidence, it must suffice to demonstrate that the results are empirically plausible (see Chapter 4).

This is a small shift in understanding the notion of ‘empirical evidence’; common throwaway appeals to ‘the future’ where technology can improve our observations and so provide proper evidence suggest that it is a temporary measure to accept simulation data like this. However, one may suggest such compromises are typical of simulation results and their acceptance into the wider body of scientific knowledge. Daston and Galison’s impressive history of images and objectivity suggests this in its final pages. The age of simulation and computerised science is the shift from image-as-representation to image-as-process, where the image itself is a tool, a hybrid of simulation, mimesis, and manipulation. Digitised atlases are not intended to solely represent, but are meant to be used, “cut, correlated, rotated, coloured.”\footnote{Daston & Galison, 2007, p. 383.} The simulation image is a presentation rather than a representation – an image that is constructed, bred of a created and manipulated model constrained by physics and computational structure. This is similar to Ihde’s more recent commentary on simulation, where “imaging in the context of simulation and modelling is more analogous to a critical, interpretive instrument.”\footnote{Ihde, 2006, p. 86. Emphasis in original.} The image is non-representational in the sense that: “there is no original from which to copy. Yet the end result is image-like; it is a gestaltet pattern which is recognizable, although it is a constructed image.”\footnote{Ihde, 2006, p. 86. Emphasis in original.}

More like a map than a picture, the simulation image is intended to convey the interpretive results of the complex calculations, making the data intelligible.

The Sloan Digital Sky Survey is one such example of these constructed virtual images, a combination of human visual interpretive abilities and algorithmic analysis. Other sky surveys are used in a similar way; for example, astrophysicists Press and Davies set out to ‘weigh’ clusters
found in the Center for Astrophysics (CfA) Redshift Survey data. The first step in doing this was to first find, among the mess of raw data, the clusters that suited their method of determining mass and to do this the authors developed an algorithm based on their required parameters. However, there was the possibility that this algorithm may have statistical biases, such as determining a high mass when this was not the case: "for these reasons, we view it as essential to calibrate our method, or any other method for that matter, against cosmological N-body experiments of known [mass].... We have prepared simulated catalogues from a snapshot of a 20,000 body simulation."\(^{494}\) The authors compared a catalogue of real data with a catalogue of simulated data, which was not simply velocities and masses taken from the simulation, but a simulation of an *observation* (see also section 4.7). An 'observer' was positioned in one corner of the simulation, and the velocity 'observed' by this entity was calculated from the degrees of freedom known absolutely in the simulation. Press and Davies used this simulated catalogue to test their algorithm, comparing the masses they calculated from 'observations' with the 'true' values known from the simulation to determine whether there were any significant biases.

The simulated catalogue can also be compared directly to observation; this provides information on what we should see, given the theory, with what we do see. Another study found there was a difference in the clustering of galaxies between the simulated observation and the real thing:

The n-body simulations should include all the essential physics of [the clustering] process, and yet they fail to mimic the observed distribution in any real sense. Is this because some additional physical process, such as dissipation, has been left out? ... Perhaps the initial conditions did not properly match those of the early universe.\(^{495}\)

As with galaxy models, simulation displays a coherent picture of our theoretical universe, allowing us to see where it fails to live up to the real thing, and providing us with a dynamic means to line the two up. Trained judgement was required to set sensible parameters on what made a cluster, but the observational data also needed to be algorithmically manipulated before producing anything that could be useful. On occasion, observation needs to be tested against theory, in a manner of speaking – using simulations to calibrate selection algorithms which then calculate the masses of real clusters, processed values which fall under the heading of 'observational evidence'.

The blurring of the line between representation and construction characterises much of the halo narrative told here. In order to play the role of experiment, simulation relies heavily on visualisation in demonstrating the virtual laboratory, and in showing the manipulation of the objects therein. The pictures that emerge from this process are used in conjunction with the scarce and indirect empirical evidence to convince the audience of the 'reality' of the results. In doing so, the boundaries of what constitutes 'empirical evidence' are re-sketched to lean less heavily on the absent empirical and more on the empirically plausible. Observation plays a crucial, epistemically superior role to simulation, but is often not definitive or unambiguous.


\(^{495}\) Davis, Huchra, Latham, & Tonry, 1982, p. 443.
enough to preclude other forms of evidence, indeed “it has been agreed for more than a century that ... the empirical elements need not be very important, or can be wholly absent, in the creative or constructive phase of a scientific theory.”  In the construction of the modern galactic model, then, it is unsurprising that empirical considerations played only a peripheral role; but in the evaluation of the model it seems that it sufficed to have a combination of indirect empirical evidence and the simulation image. The latter, as a constructed image, is not merely representational, but instead stands as evidence for the efficacy of the theoretical model that derived it, the computational code that produced it, and the legitimacy of physical assumptions that allowed for it.

3.10 Reconstructing Astrophysics as an Experimental Science

The aim of this chapter was to demonstrate how simulation can do science; the case study adopted for this purpose was the search for a stable galaxy model, and how the halo came to be included in this model. Through various twists and turns, the modern model of a galaxy has developed three main components – a luminous disc, a central mass, and a dark halo. In following how this came about, some answers to the question of how simulation does science and produces knowledge have been offered. In astrophysics, it seems that one way in which simulation finds its place is by adopting several of the roles that experiment traditionally plays in other sciences. Due to the similarity in methodology, simulations are often referred to as ‘experiments’, in which parameter variation and error analyses are conducted in much the same way as in the laboratory. Miller makes conscious use of this metaphor of experiment, but the language is common in the simulation papers studied in this chapter. Extending the metaphor, simulation takes on the role of ‘the other’ form of science, in opposition to theory, and this is framed using the Popperian narrative of ‘confirmation of theory by experiment’. In this way Toomre’s criterion was established as ‘fact’. Simulation is only ever metaphorically experiment, however, as evidenced by the way it lacked the power to overcome the strong doubts of the community and ‘confirm’ the Ostriker-Peebles conjecture. This is how the duality of simulation – as both ‘experiment’ and ‘theory’ – manifests in actual scientific practice, with scientists easily switching from experimental to theoretical language based the aim of their argument. Simulation also demonstrated its more unique features in how it cohered several disparate pieces of information about galaxies into a ‘testable’ model. Through this, simulation work revealed an anomaly (the bar instability) and was instrumental in understanding and resolving it. Arguably, simulation also discovered new phenomena, as the bar moves from being an instability – a theoretical problem – into a stage in the early evolution of galaxy formation, becoming an occurrence that is at least in principle observable. Along the same theme, simulation also introduced (and at the same time almost defeated) a new object – the halo.

---

496 Kragh, 2014, p. 49. Kragh refers to agreement in cosmology and other areas of fundamental physics, not science as a whole.
There is an odd dependence on simulation when providing empirical evidence, one that suggests a reworking of what ‘empirical’ might mean. Simulation alone couldn't validate halos, but observation in conjunction with simulation could. The rotation curves kept halos in the galactic model when they weren’t required for stability, but simulation kept them there as halos (massive, spherical), as well as introducing them in the first place. Empirical grounding is necessary, but plays a looser role than one might expect. Empirical data comes from observation, but this latter approach has none of the manipulability or intervention that are the hallmarks of experimental knowledge. Simulation does, but has none of the empirical grounding that is the epistemic force of experimental knowledge.

We cannot accuse modern astrophysics of dismissing the importance of empirical grounding, but the practicalities of sparse observational evidence means that simulation finds a place as a technique that provides supplementary empirically plausible evidence in an experiment-like manner. Astrophysics, then, by having a metaphor and narrative of experiment, effectively reconstructs itself as a traditional natural science following, “a nostalgia for the lost original points of reference which can then, of course only be simulated and reconstructed.” However, because astrophysics has never been an experimental science, the ‘reconstruction’ is done in the postmodern sense of restoring an original that never existed. If a simulacrum is “an exact copy for which no original exists”, then astrophysics is a simulacrum of an experimental science. Simulation does not take the place of experiment because in astrophysics experiment has no place to begin with; rather it takes on the mask of experiment when required to ‘test’ or ‘confirm’ (in conjunction with what observational evidence is available), and removes it when required to build or alter theory. This flexibility of role is rather fruitful, as seen in this chapter’s case study, enabling the construction, testing and at least partial confirmation of a new model of spiral galaxies. The implication of this mode of doing science, however, necessarily blurs the lines between empirical and simulated data, leading to notions of virtual experiments and virtual observation.

---

497 One can perhaps draw a link with simulation and other types of techniques used to supplement or interpret experimental data. Ursula Klein, in looking at the ‘paper tool’ of Berzelian formulas in chemistry, describes that if a reaction product was reasonably well known and its Berzelian formula accepted, the formula was taken as a substitute for the actual measurement of the mass – "here, the formula fulfilled a wholesale function otherwise performed by a laboratory tool or experimental inscription device" (Klein, 2001a, p. 275).


4 The Case of the Missing Satellites: Technoscience and Astrophysics

4.1 Introduction: Technoscience in Astrophysics?

While the previous chapter focused more on the positioning of simulation with respect to theory and experiment, this chapter takes a closer look at the actual interplay between simulation and observation, and how this results in new knowledge. This interplay will be understood as a negotiation around both technical and theoretical limits, aided and allowed by the dual technologies of observation and, particularly, simulation. Because of this focus on technology and the production of knowledge, it makes sense to draw on the portion of literature in HPS that styles itself as technoscience. Technoscience is generally understood to refer to some sort of interaction between science and technology. However, philosophers vary on what constitutes ‘science’ and ‘technology’, and where and when the division between the two occurs, if it does at all. The scope of these debates is beyond this chapter, which will disregard questions around the ‘epochal break’ claim, since that would require a far more concerted historical thesis. Another division in the technoscientific literature is between external and internal: “first, the alliance of modern science, technology and industry; and, secondly, the technical shaping and production of scientific objects within the experimental sciences.” This chapter, when talking of technoscience, refers to the latter understanding, taking as its departure the notion that scientific knowledge is inextricably linked to the technologies that aid in producing it.

In fact, the central motivating question of this thesis – how simulations generate knowledge – begins with the very premise that knowledge is entangled in its technologies. This approach is in common with Alfred Nordmann’s formulation of technoscience, in which:

> The business of theoretical representation cannot be dissociated, even in principle, from the material conditions of knowledge production and thus from the interventions that are required to make and stabilize the phenomena. In other words, technoscience knows only one way of gaining new knowledge and that is by first making a new world.

It is the ‘even in principle’ that is important to Nordmann, who likewise brackets questions of historical breaks. Typical examples of technoscience tend to focus on specific technoscientific objects – like onco-mouse – or on the material technologies used in the production of technoscientific knowledge – like the scanning tunnelling microscope. That is, technoscience is

500 See, for example: Rheinberger, 2005a; Barnes, 2005; Klein, 2005b; and the essays in Nordmann, Radder, & Schiemann, 2011.

501 Klein, 2005a, p. 139. Similarly, Pickstone identified three levels on which the term ‘technoscience’ can be used: level one, to describe interactions of knowing and working (‘internal’, with science and technology converging); level two, to describe ‘sci-tech’ fields that exhibit dense intertwinings of science and technoscience (‘internal’, with science and technology exhibiting interdependence); and level three, to describe academic, university, and state interpenetrations (‘external’) (Pickstone, 2005).

seen from the perspective of the technology (or the technology-science hybrid), as opposed to that of the more representative, theoretical 'science'. Indeed, Nordmann argues that an 'epochal break' would only be visible from the perspective of classical 'science'; from the point of view of technoscience, the history of science and engineering has always been technoscientific. Technoscience is associated with the making of its objects of study, and naturally this has a heavy material component which emphasises experimentation and engineering if not outright manufacture. So where does that leave astrophysics – and more to the point, where does that leave the non-material simulation?

Astrophysics is not a technoscience; at least, not in the same sense as genetic engineering and nanotechnology. It does not materially produce its objects of study, it has little to do with industry. Astrophysics does, however, make use of sophisticated technologies. Observation technology – not only ground and space telescopes, but the whole gamut of receivers and interferometers that measures the variety of signals from space – has pride of place when it comes to empirical say-so in astrophysics. And as should hopefully be evident by now, simulation technology – computers and algorithms – has an incredibly important part to play in the production of theoretical knowledge. At the same time, despite simulation being commonly included in references to modern 'technoscience', the use of the determinately virtual simulation (see section 5.5) to generate knowledge does not seem to gel with the materially focused aspects of technoscience.

On the other hand, there are many aspects of how simulation functions with relation to the knowledge it produces that are reminiscent of technoscientific literature: the embeddedness of science and technology, of knowledge production and generation; the creation of self-contained systems; and the closeness between the ‘object’ and ‘instrument’.

Just how the technologies of observation and simulation fit together with each other, and how this interplay produces knowledge is the subject of this chapter. It will be demonstrated that astrophysics can indeed be conceived of as a technoscience, in which computational technologies particularly are essential to understanding how theory is enacted through simulation. In fact, beginning with astrophysics forces a classical perspective in that the chapter looks at theory rather than focussing on material technologies, a reversal of the usual approach found in technoscientific literature. Through this perspective, simulation is shown to effect a negotiation between observation and theory; in particular, this chapter emphasises simulation as a 'technology of theory'.

This approach has something in common with the portion of HPS literature that looks at the ‘paper tools’ of theoreticians. David Kaiser, who studies the use and dissemination of Feynman diagrams in post-war physics, reminds us that theorising is most often not purely ‘in the mind’

---

503 Nordmann, 2011.
504 Though to be fair, given the wide variety of understandings of what 'science' and 'technology' mean in the context of technoscience there is plenty of scope for a non-material, epistemically focused understanding of the modern interactions between science and technology. It is also for this reason that the technoscientific framework is only borrowed for this chapter, rather than attempting to make a wholesale adoption.
but involves the act of calculating, for which theoreticians develop a wide range of tools.\textsuperscript{505} Ursula Klein similarly looks at how the paper tools of Berzelian formulas in organic chemistry in the first half of the nineteenth century were adapted for a multitude of purposes in order to help scientists interpret experimental results and answer theoretical questions.\textsuperscript{506} While not a paper tool per se, simulation is also a tool of theory. Feynman diagrams were used persistently and widely because, as Kaiser describes it, they were dispersed, in the sense of the word that "emphasizes the diagrams' plasticity as physicists refashioned and redeployed them for new applications".\textsuperscript{507} Simulation similarly owes its ubiquity to its wide usefulness and applicability. Likewise, Klein suggests that "on many levels paper tools are fully comparable to physical laboratory tools or instruments",\textsuperscript{508} and this is a very similar argument to that found in the HPS simulation literature in Chapter 1. Importantly, both paper tools (Feynman diagrams, Berzelian formulas, partial differential equations)\textsuperscript{509} and simulation do not physically interact with the object under investigation but still represent "visible marks which can be manipulated to create representations of scientific objects – sometimes even to co-produce inscriptions with laboratory instruments – and to explore their relationships with other objects as they appear in the light of concepts and theories embodied by these marks."\textsuperscript{510} Tools of theory may be abstract in target, but are materially present. For simulation, this extends beyond paper or blackboards into the area of instrumentation – simulation is a \textit{technology} of theory.

The case study through which the interplay of simulation and observation technology will be explored is that of the missing satellites problem. To set this problem up, consider a quote from a 1999 astrophysics paper:

\begin{quote}
If the observed galaxies have large DM [dark matter] halos, then N-body simulations can, in principle, be used to predict the distribution of the dark matter component, to associate the simulated DM halos with galaxies, and to predict these galaxies' bulk properties, such as position, mass, and size. One should be able then to make predictions about the spatial distribution and motion of these simulated galaxies and to compare these predictions with corresponding observations.\textsuperscript{511}
\end{quote}

The authors of the paper state quite clearly the underlying motivation for many dark matter simulations. Due to both technological and theoretical limitations, astrophysicists are unable to fully simulate both baryonic (‘ordinary’) matter and dark matter in parallel. Therefore most astrophysical simulations are dark matter only, based on the common assumption that the dark matter component, being dominant, contributes the most to galaxy formation. The hope is then to associate the activity of the dark matter structures to observable quantities, and so indirectly ‘test’ the underlying model or theory.

\textsuperscript{505} Kaiser, 2005, pp. 8-9.
\textsuperscript{506} Klein, 2001a. See also the edited volume Klein, 2001b for other examples.
\textsuperscript{507} Kaiser, 2005, pp. 357-358.
\textsuperscript{508} Klein, 2001a, p. 292.
\textsuperscript{509} Wise, 2011.
\textsuperscript{510} Klein, 2001a, p. 292.
\textsuperscript{511} Klypin, Gottlöber, Kravtsov, & Khokhlov, 1999a, p. 530.
As one might imagine, associating dark matter with its possible observable effects is no easy task. What astrophysicists call ‘the missing satellites problem’ arose due to the difficulties in comparing simulation results with observational data, and how this problem came to be (mostly) solved is the subject of this chapter. The missing satellites problem occurred when, in the late 1990s, simulations based on the standard cold dark matter model predicted several hundred more satellite galaxies orbiting the Milky Way and Andromeda than are actually observed. More precisely, what the simulations actually predicted were several hundred small dark halos, while what we observe are the luminous galaxies supposedly hosted in those dark halos. Yet, and especially in the early simulations, there was no way of telling which of the simulated dark halos hosted a luminous galaxy; it was generally assumed that if there was a halo in the simulation, we should see corresponding starlight in observation. However, the predicted number of halos in the simulation was an order of magnitude more than the number of observed dwarf galaxies – something had gone wrong somewhere.

Two types of suggested solutions emerged: (i) astrophysical solutions, which assumed that a good chunk of the halos were somehow being kept from forming stars; and (ii) cosmological solutions, which assumed that most of the halos did not form in the first place and that the simulations were wrong. What followed was a complex process of negotiation between better simulations and a new set of observations, that eventually resulted in a consensus of a ‘solution’ at the start of the last decade. Though it is still early days, it seems to be generally accepted that a combination of astrophysical mechanisms acts during galaxy formation to keep most of the satellite halos dark, which explains the discrepancy with the observed satellite galaxies. New hydrodynamic simulations, that are capable of telling which halos host luminous galaxies and which do not, demonstrate the plausibility of this solution – plausibility being an important attribute, as later sections will show. In helping ‘solve’ the missing satellites problem, simulation constructs a universe that, in a technoscientific manner, is by itself a legitimate epistemic target.

4.2 The Problem

The structures we see in the universe today (galaxies, clusters, superclusters, and so on) are expected to have formed by hierarchical clustering. This means that low mass systems collapse early and gravitationally cluster to form larger mass systems over time. A large halo, such as that of the Milky Way galaxy would have formed from several smaller halos. These smaller halos collapsed early when the Universe was very dense, so they have correspondingly high core densities. When they merge into larger halos, these high densities allow some of the smaller halos to resist the strong tidal forces that act to destroy them during the merger. Gravitational interaction with the larger halo will serve to unbind most of the mass attached to the original smaller halo, but some will survive this process and remain bound to the progenitor. The remnants of these merged halos are known as ‘subhalos’ or substructure, and most cold dark matter (CDM) halos are expected to host a large amount of substructure.
Some of these structures can be observed today as very faint dwarf galaxies of low mass and luminosity, acting as satellites to larger galaxies like the Milky Way and Andromeda (M31). The latter two galaxies and their satellite galaxies make up what is known as the Local Group. Because of its proximity, observations of the Local Group are often used as a ‘cosmological probe’, which usually means simulating a pair or group of galaxies and comparing the results to observations of the Local Group to test the viability of the underlying cosmological model, on the assumption that our local galaxies behave much like any other galaxies. In terms of the experimental narrative explored in the previous chapter, one might say that "the Milky Way (MW) environment provides an excellent laboratory for astrophysics. It has been used extensively in the past decades to test theoretical models of galaxy formation." 512 Yet there is still much to be learnt: "despite significant effort, theoretical predictions of the abundance and properties of the satellites are far from being complete." 513 Modelling the survival of satellites inside halos of large galaxies is a numerically demanding problem that requires very high resolution from a simulation in order to track the growth and accretion of such small structures, without losing them among the particles of the larger halo. In order to predict abundances of satellites, both the dynamics of the satellites and the formation of the parent halo need to be modelled in a cosmological setting.

Typically a semi-analytic approach is the key to the solution to such issues of computational resolution (recall section 1.3.2). One 1993 paper adopted a semi-analytic model that used Monte Carlo techniques to construct merger history trees for dark matter halos based on a network of theoretical assumptions and statistical treatments. The model seemed to function quite well, correctly matching observations of galaxy luminosity, colour, gas content, and morphology: "although it would be premature to attempt a detailed quantitative fit to specific cosmological models, the qualitative agreement between the data and the general picture that we present is already very encouraging." 514 Though the authors admitted that their model was necessarily a simplified depiction of no doubt very complex processes, the semi-analytic approach did well to highlight underlying trends of the model and their broad effects on predictions. Happily, it seemed that understanding galaxy formation within the framework of the clustering model "provides an elegant and natural explanation for the qualitative trends seen in systems of galaxies." 515 There was one hiccup: the model over predicted the number of faint galaxies within a standard galaxy-sized halo by a factor of 5 to 10. Increasing the efficiency of dynamical friction would destroy these excess galaxies, but would also destroy the larger mass systems which are observed. Therefore, "the only apparent solution to this problem is to assume that many halos remain observationally undetectable" 516 and unless this could be demonstrated, "cosmological models that predict fewer low-mass halos are favoured". 517
Returning to this issue a few years later, a 1999 paper commented: "although the above study is suggestive, there is clearly a need for a more detailed study using direct numerical simulations".\textsuperscript{518} It was only in the late 1990s that such direct simulation became possible to do to a sufficiently high resolution. The title of this particular 1999 paper (hereafter KKVP, for its authors) – ‘Where are the Missing Galactic Satellites?’ – is responsible for what later came to be known as the ‘missing satellites problem’\textsuperscript{519}. MSP, as it is sometimes abbreviated, refers to the problem that simulations of galaxies similar to the Local Group over predict by a considerable margin the number of low mass satellites actually observed. To determine this, KKVP had to perform counts for both the observed satellites and the simulated ones. In the first instance, this was not trivial. One 1998 review, in listing forty-one candidates for members of the Local Group, noted that half of these candidate galaxies have been discovered since the early 1970s, and that it was plausible that there were yet more to be found.\textsuperscript{520} Indeed, a 2010 review noted that a few dozen more very faint satellite galaxies had been discovered in the preceding decade, with more discoveries expected in the near future.\textsuperscript{521} Therefore KKVP could only count those satellites that had been discovered; to mitigate this uncertainty, in their simulation they only considered satellites with a velocity dispersion of larger than $10 \text{ km s}^{-1}$, on the basis that it was unlikely that larger galaxies had been missed (the larger the velocity dispersion, the higher the mass and corresponding luminosity, broadly speaking). They also only considered the number of satellites per central galaxy – in this case the Milky Way and Andromeda – simplifying the problem by not including the few isolated or remote galaxies, and by counting any galaxies within a certain radius of a host galaxy without attempting to determine whether it was indeed bound or not. The authors chose to use circular velocity to characterise the dark halos and their satellite galaxies, as this was well-defined in the simulation and a measurement for which they had observational data for most satellites.

Counting the number of simulated satellites required the use of a finding algorithm. This piece of code identified by various means the location of a simulated halo and its bound particles, "then computes various properties and profiles for each of the bound halos and constructs a uniform halo catalogue ready to be used for analysis."\textsuperscript{522} This catalogue of simulated observations could then be compared to the real data set. Halos were identified as such if they had more than 20 particles, which put a mass constraint on the particle of the simulation to be quite small (to match the mass of the small satellites of the Milky Way and Andromeda). Since computational power put limits on the number of particles, this meant that the ‘box’ or volume of space that the simulation could generate is quite small, only a few megaparsecs across. Thus the study could only apply to

\begin{footnotesize}
\begin{enumerate}
\item Klypin et al., 1999b, p. 83.
\item Another paper published the same year (Moore et al., 1999) performed a similar but independent simulation, but I only examine KKVP in detail as both studies achieved similar results (and are almost always cited together as the origin of the ‘missing satellites problem’).
\item Mateo, 1998, p. 447.
\item Kravtsov, 2010, p. 6.
\item Klypin et al., 1999b, p. 86.
\end{enumerate}
\end{footnotesize}
‘field population’, that is galaxies not in the vicinity of massive clusters which, fortunately, the Local Group can be counted among.

Figure 15 shows the simulated group of halos from KKVP. Within the radius simulated (the solid circle enclosing the halos), the authors identified 281 satellites. This, compared with the approximate 40 observed in the actual Local Group, was a large discrepancy. While the simulation overall predicted very well observed abundances of larger satellites, it over predicted rather drastically the smaller satellites. KKVP were not overly concerned: “although the discrepancy between observed and predicted satellite abundances appears to be dramatic, it is too early to conclude that it indicates a problem for hierarchical models. Several effects can explain the discrepancy and thus reconcile predictions and observations.”

KKVP specified two of these possible solutions. The first identified the missing satellites with high velocity clouds (HVCs) observed in the Local Group. While this was plausible, “at present the observed properties of HVCs (mainly the abundances, distances, and line widths) are so uncertain that a more quantitative comparison is impossible.” This solution gained little attention during the subsequent wider search for an explanation to the missing satellites problem. The other option KKVP discussed was the presence of dark satellites – either the halos were completely dark, or they hosted small galaxies with luminosities so low they were quite invisible to observation. Satellites containing such a small amount of luminous matter had to have been formed by some mechanism that prevented the formation of stars, inhibiting the collapse of gas

---

523 Klypin et al., 1999b, p. 89.
524 Klypin et al., 1999b, p. 90.
clouds. Though it was likely that some low luminosity satellites were missed by observational surveys, it seemed unlikely that they missed so many, therefore any of these mechanisms would have to be able to produce a large amount of very faint systems. This option, of course, would be even more resistant to observational confirmation (though it is observationally plausible – if it’s invisible, of course we can’t see it); in this case KKVP emphasised that “more theoretical work needs to be done”, though there were “potential observation tests for the existence of dark satellites” in the works.

In either case, KKVP’s main conclusion was that the simulation model over-predicted the amount of small satellite galaxies in a system similar to the Local Group, in a numerically and physically robust manner. KKVP had satisfied themselves that the excessive satellites were not a numerical fluke; more specifically, that one could not model a hierarchically forming galaxy cluster and avoid large satellite abundances. Given this, KKVP chose to assume that the hierarchical clustering model was accurate, that therefore the predicted satellites did exist, and so attempted to answer the question “where are the hundreds of predicted ... halos” with suggestions of high velocity clouds or dark satellites.

4.3 Possible Solutions
An independent study, which was published in the same year as KKVP and showed the same over prediction of satellites, summed up the problem bluntly: “either the hierarchical model is fundamentally wrong or the substructure lumps are present in the Galactic halo but contain too few baryons to be observed.” The suggested solutions to the problem following the 1999 papers tended to follow this rough binary by either altering the cold dark matter model or finding some way to favour dark satellites. Joshua Simon and Marla Geha therefore labelled the first category of solutions ‘cosmological’, and the second ‘astrophysical’, based on which area of theory was being modified. The cosmological solutions involve the suppression of substructure formation below a certain threshold, so the satellites do not form in the first place, while astrophysical solutions involve the suppression of luminosity in certain subhalos, so the satellites...
are mostly dark. Put simply, in the cosmological class the missing satellites do not exist, and in the astrophysical they do.

Solutions that fell into the ‘astrophysical’ category generally attempted to demonstrate the plausibility of dark satellites. The missing satellites of the simulation could be identified with real satellites if they were very faintly luminous, or entirely dark, thus evading all our present technological capabilities of detection. Mechanisms to make (or keep) a satellite dark are mechanisms of suppressing star formation. Stars are understood to form as gas clouds collapse gravitationally. In order to do this, the gas cloud needs to cool and become denser (and form other, heavier molecules). If this cooling is prevented or slowed down, stars would not form. For example, if the gas in low mass galaxies was ionised by background photons (‘photoionising’), the gas in the cloud would remain hotter for longer. Gas accretion would be effectively suppressed in low mass halos after the epoch of reionisation, which meant that typical galaxy halos would contain many low mass dark matter subhalos. In proposing this sort of astrophysical solution, Bullock, Kravtsov, and Weinberg reconfigured the missing satellites problem as “a mismatch between the expected number of dark matter subhalos orbiting within the Local Group and the observed number of satellite galaxies.” The problem was how to account for dark satellites, rather than how to account for extra satellites.

Naturally, cosmological solutions took the opposite view. These solutions began with the assumption that since we do not see them, the extra satellites probably do not exist. In seeking mechanisms to suppress the formation of small halos in the first place (before a galaxy halo finishes accreting), cosmological solutions therefore sought to alter the model that had given rise to the presence of small halos in the final instance. Thus cosmological solutions are so named because they modify the cosmological model – the CDM (cold dark matter) model. Using warm dark matter (WDM) instead of cold, or reducing the small scale power (fluctuations which, through inflation, grow into halos and their galaxies) would prevent low mass objects from forming early on in the Universe. Another alteration to the CDM model was to make dark matter self-interacting, which would have the effect of making small halos more likely to be destroyed during accretion to the larger halo. In any of these cases, the present day Local Group would

533 In fact, the problem for those preferring astrophysical solutions was to in fact account for how any of the satellites could be seen at all, corresponding to those few bright satellite galaxies that had been observed. It had been demonstrated, by simulation, that star formation in low mass halos was rather inefficient in the first place. Some authors therefore attempted to provide an astrophysical mechanism – tidal stripping – that explained how some of the small halos did form stars. This suggested solution therefore takes as its starting assumption that any small halo would be dark and sought then to explain the few bright (Kravtsov, Gnedin, & Klypin, 2004). This is an inversion of the cosmological approach, which takes as granted the bright and seeks to explain (away) the dark
have a low population of satellite galaxies that approximately corresponded to the luminous galaxies that were actually observed.

Since models of all the various suggested modifications were in line with current observations, it seemed that there was no compelling empirical (or otherwise evidentiary) reason to prefer one category of solutions over the other. There are, of course, other motivating forces – other than logic or empirical evidence – in play when it comes to theory choice, a fact well documented in HPS literature. A variety of epistemic or theoretical values are known to assist in judgement between theories – virtues like simplicity, elegance, unifying power, fruitfulness, and explanatory power – this is not new philosophy. Astrophysical solutions tended to be motivated by the explanatory success of the CDM model – it made more sense to theory craft in areas where the physics was poorly understood (like in the astrophysical processes of galaxy formation) than to assume a model with well-demonstrated explanatory power was at fault. In addition, adopting astrophysical mechanisms to suppress star formation was seen as a more “natural, almost inevitable” because they used existing physical processes to explain the missing satellites problem, creating new relationships rather than new physics.

Cosmological solutions could not lay claim to such virtue, but then they tended to see far more value in empirical fit – developing a theory that doesn’t conflict with observation even if that requires some fine-tuning (presumably the hope is that these modifications will turn out to be quantitatively justified). Cosmological solutions sought an integrated way to account for the dissonance between simulation and observation within the context of wider cosmological theory; astrophysical solutions could, on the other hand, be accused of using post hoc or worse, ad hoc, tactics to save the phenomena. One set of authors felt that though the CDM model was quite successful at universe-scale, “standard cold dark matter (CDM) theory appears to be in conflict with observations of clustering on subgalactic scales”, of which the missing satellites problem was only one. The authors of another paper, which suggested self-interacting dark matter as a

---

537 For overviews, see Worrall, 2000 and McMullin, 2008.
538 This is not to suggest that those proposing astrophysical solutions were blind to the deficiencies of the CDM model – the very fact that the missing satellites problem existed was enough to suggest flaws in the explanatory power on small scales. Those suggesting astrophysical solutions were confident, however, that the problems could be resolved within the framework of the CDM model, without having to change any of the basic assumptions which had so far proved useful. Even those suggesting cosmological solutions did not suggest any major overhauling of the CDM model as their modifications were usually restricted to the small scale and were deliberately non-disruptive on other scales.
539 Bullock et al., 2000, p. 520.
540 Bois et al., 2001, p. 93.
541 Bode et al. also referred to the ‘core-cusp’ or ‘cuspy halo’ problem, another mismatch between simulation and observation – simulations predict that dark matter density forms a cusp at the centre of the galactic halo, but observations seem to imply there is no cusp in the dark matter distribution at all. Other (later) problems for CDM include the ‘too big to fail’ problem (galaxies do not form in the most dense substructures, undermining the assumption that the bigger the dark halo the brighter the galaxy) (Boylan-Kolchin, Bullock, & Kaplinghat, 2011), and the ‘planes of satellites’ problem (bright galaxies seem to orbit their hosts in the same plane, while simulations predict we should see substructure wherever we look) (Pawlowski & Kroupa, 2013). The latter two problems only came to the attention of astrophysicists in the early part of this decade and are therefore not relevant to this section’s overview of cosmological solutions, but are relevant to those suggesting alternatives to CDM after 2010.
solution, were also motivated by inconsistencies of the CDM model: they “are lead to consider self-
interactions because ordinary astrophysical processes are unlikely to resolve the problem with
standard, weakly interacting dark matter.”542 Cosmological solutions sacrificed theoretical
aesthetic – “in the absence of an obvious astrophysical mechanism, it is natural to think of more
exotic explanations”543 – in search of a model that, self-consistently, explained all the phenomena
on both large and small scales.

Whether one preferred an astrophysical or a cosmological solution, at this stage, was born of
preference for particular epistemic virtues as both options explained the observed phenomena
and no solution was immediately testable. This sort of narrative – underdetermination of theory
by data, theory choice governed by epistemic virtues – is a common one: it is the classic story of
theory versus observation. The interaction between the theory and data is mediated by
simulation, as this is where the suggested solutions are built and tested, but if this is
technoscience, it is a very weak version. And since the data – counts of satellites – is rather sparse,
there is very little progress in knowledge to be made from this limited perspective, even with the
presence of epistemic virtues. Following the classical narrative, then, this impasse must be
resolved with the introduction of some new information, preferably better and more complete
observations that will tell us one way or another where and if the satellites exist:

If these problems with CDM are real, they represent a remarkable opportunity. The
observed pattern of gravitational clustering may be revealing the physical properties of the
dark matter. If so, this will be an invaluable clue to physics beyond the standard model.
Deciphering this clue represents an exciting challenge in which both more refined
observations and numerical simulations will be needed.544

4.4 ‘Future Observations’
In 1971, Paul Hodge reviewed the objects of the Local Group that could be classified as dwarf
galaxies. There were fourteen in total: Sculptor, Fornax, Leo I, Leo II, Ursa Minor, and Draco are
all companions to the Milky Way, and a further four are companions to Andromeda (with the
memorable names of M32, NGC 205, NGC 185, and NGC 147). Added to this count were two
irregular galaxies (NGC 6822 and IC 1613) and the Large and Small Magellanic Clouds.545 These
dwarfs are referred to by later literature as the ‘classical’ satellites, though they are almost never
picked out specifically. In 1998, Mario Mateo wrote an updated review that increased the number
of observed satellites to around 38 (give or take a few uncertain cases), almost tripling the
previous count: "more dwarfs have been confirmed or identified as Local Group members in the

542 Spergel & Steinhardt, 2000, p. 3760.
543 Kamionkowski & Liddle, 2000, p. 4525.
544 Bode et al., 2001, p. 94. Note the reference to improvements in both simulation and observation; this
statement will be important in the following sections.
545 Hodge, 1971. This number fourteen was approximate, as there were six or so uncertain cases, and the
Large and Small Magellanic Clouds may be too large to be considered ‘dwarf’.

135
past 27 years than during the previous 222 years beginning with Le Gentil's discovery of M32!".\textsuperscript{546} In a section somewhat prophetically entitled ‘What’s Missing?’, Mateo mused that since over 50% of the Local Group satellites have been found since 1971, “the era of discovery within the Local Group probably is not yet over”.\textsuperscript{547} Assuming a uniform galaxy distribution, Mateo predicted as many as 15 to 20 more galaxies would be found at low latitudes.

The year following that review, of course, saw the publication of the 1999 simulation studies that suggested there was an entire order of magnitude more than 38 satellites. There was little to arbitrate between suggested solutions to the discrepancy – it was not even clear whether the culprit lay in astrophysical or cosmological theory. Ideally, the arbitrator would be observational data; some solutions attempted to provide observational predictions of their models that future observations might corroborate. In general, there is often a door left open for ‘future observations’ in astrophysics, which are expected to refine or verify a model, to refute a contrary idea, to validate an assumption, and generally to provide more substantial grist for the theoretical mill. A small sample of examples from the astrophysical papers looking at the missing satellites problem reveals this trend; the examples extend up to 2014 to demonstrate how little this habit changes even in the face of said ‘future observations’:

“This picture ... can be either confirmed or falsified by the upcoming observations”,\textsuperscript{548} and “these properties ... should be testable in the near future with new upcoming surveys.”\textsuperscript{549} “The signatures of these events may be detectable in the velocity data of future astrometric missions and several deep halo surveys that will soon be completed.”\textsuperscript{550} “Future large sky surveys like LSST, DES, PanSTARRS, and SkyMapper should be able to see these satellites if they do exist and thereby provide unprecedented constraints on the nature of galaxy formation in tiny halos.”\textsuperscript{551} “Future observations and quantification of the masses of the newly discovered MW satellites will enable us to do precision tests of the viable MSP solutions.”\textsuperscript{552}

\textsuperscript{546} Mateo, 1998, p. 439.  
\textsuperscript{547} Mateo, 1998, p. 447.  
\textsuperscript{548} Klypin et al., 1999b, p. 90.  
\textsuperscript{549} Klypin et al., 1999b, p. 91.  
\textsuperscript{550} Zentner & Bullock, 2003, p. 51.  
\textsuperscript{552} Strigari et al., 2007, p. 682.
“Future imaging surveys of stars in the Milky Way will provide a more complete census of low-luminosity Milky Way satellites”.553

“Future observations will tighten constraints on the galaxy formation models”,554 and “we can expect to find many more systems at larger radii in the future deep wide area surveys.”555

“Whether or not our input assumption about galaxies tracing subhalos is correct, future surveys will provide a means to test it and thereby provide an important constraint on the formation processes of these extreme galaxies.”556

“At low redshift, most of the baryons have not yet been detected. Improved instrumental sensitivity may lead to their detection in the foreseeable future.”557

This appeal to future observations is an appeal to improvements in technology – similar statements can be found appealing to improvements in simulation capabilities.558 Astrophysicists expect the new data to be put to a variety of uses; for the missing satellites problem, this manifested as the general hope that the missing satellites would be found. In the early 2000s, it seemed that this hope might be realised when data began to be released from the Sloan Digital Sky Survey (SDSS), a comprehensive imaging and spectroscopic survey of the optical sky.559 Spectroscopic surveys of the SDSS data measured the velocity dispersions of possible new systems in order to confirm their galactic nature, resulting in the discovery of several new, very faint dwarf spheroids.560 These new satellites seemed to be dark matter dominated with very high mass-luminosity ratios, and were among the darkest known stellar systems in Universe.561

In 2007, Simon and Geha wrote: “in the past three years, at least 20 of these galaxies have been identified, nearly doubling the previously known population.”562 Around 12 of these new very dark satellite galaxies were bound to the Milky Way. Simon and Geha looked at eight of these twelve, and compared the observed galaxy properties to the properties of subhalos in the Via Lactea simulation via estimates of circular velocities. Their conclusion was that “the ultra-faint
dwarfs significantly fill in the gap for satellites in the two lowest mass bins, but have masses that are too small to affect the satellite deficit at higher circular velocities.” Future observations had arrived, but the new counts were still too low to match up adequately with simulations. Yet the SDSS data was crucial for developing a consensus of solution for the missing satellites problem.

Figure 16: Luminosity function as observed (red dot-dashed lower curve), as corrected for sky coverage (green dashed middle curve), and with all corrections included (blue solid upper curve). The abscissa represents brightness (left to right is darkest to brightest) and the ordinate the number of satellites. Colour version of fig. 6 in Tollerud et al., 2008, p. 282.

Though ‘raw’ data was not especially useful for resolving the issue, it did increase the plausibility of finding more faint satellites. Furthermore, it became plausible that entirely dark satellites existed in a significant amount, but this conclusion required a bit of work. Though SDSS had made it possible to probe the faint end of the luminosity function of Milky Way satellites, the raw data set by itself was not strong enough to be directly comparable to the simulation catalogues. Therefore in order to make the data set more complete, corrections for both the undetected and

---

564 The authors demonstrate that if dwarf galaxy formation was suppressed after reionisation, “the circular velocity function of Milky Way satellite galaxies approximately matches that of CDM subhalos” (Simon & Geha, 2007, p. 327, emphasis removed). This statement comes with five caveats: the extrapolation of dwarf spheroid abundances from the observational data must be reasonable; the observed velocity dispersions must provide a reasonable estimate; the primary physical mechanism for suppressing dwarf spheroids must be reionisation; the cosmology used in the Via Lactea simulation must be a good match to ours; and the main halo in the Via Lactea simulation must be a reasonable approximation for the Milky Way.

138
’undetectable’ satellites needed to be made. Erik Tollerud and his collaborators set about conflating existing corrections to the data set and making some more:

Unfortunately, given the inherent faintness of the newly discovered [dwarf spheroids] and the magnitude-limited nature of SDSS for such objects, a derivation of the full luminosity function of satellites within the Milky Way halo must include a substantial correction for more distant undetectable satellites. ... Our aim is to take the detection limits for SDSS Data Release 5 ... and combine them with a CDM-motivated satellite distribution. In order to provide a theoretically motivated estimate for the total number of Milky Way satellite galaxies, we adopt the distributions of subhalos in the Via Lactea simulation.565

The SDSS data was incomplete in two ways: first, it had angular incompleteness as it only covered about one-fifth of the sky. Second, beyond a characteristic heliocentric radius, SDSS would not be able to ‘see’ a satellite below a certain magnitude even if it did exist. Statistical completeness limits on the data set itself gave the values of the characteristic radius and magnitude, and using this the catalogue of observations could be filled out for sky coverage – this would account for the undetected satellites (green dashed line in Figure 16). This completed the data set to make it as if the SDSS survey that looked at the entire sky. Speculative, but in a statistically robust sense. However, the data set corrected for sky coverage would still not account for satellites that were too faint or too distant to be seen by the SDSS instruments.

This was where the simulation Via Lactea came in. Via Lactea, as suggested by its name, simulated a Milky Way-sized galaxy. Like all simulations based on a CDM model, Via Lactea predicted several dark satellites whose placement relative to the galactic centre were determined by the spatial distribution of dark matter. By placing an ‘observer’ near the centre of the simulation, Tollerud et al were able to conduct several mock surveys in which the simulated observer recorded the satellites that would be seen with no instrumental restrictions: “in the end, we produce 18,576 equally likely mock surveys, each with their own correction factors, and use these to correct the Milky Way satellite luminosity function for angular and radial incompleteness.”566 Using the simulation to make a more refined correction, the authors further ‘completed’ the SDSS data set with those satellites that might be below the characteristic magnitude or beyond the radius limit – this accounts for the undetectable (by SDSS) satellites (see blue solid line in Figure 16). The corrected curve indicated there are several hundred more satellites at very low luminosities, while agreeing with the raw SDSS data (red dot-dashed curve in Figure 16) at the more visible end of the spectrum. Tollerud et al stressed that the outcome did not depend on the total number of satellites in the simulation, only the spatial distribution of subhalos, and that furthermore “numerical effects tend to undercorrect, producing a conservative satellite estimate.”567

Tollerud et al did not use the SDSS data for exact satellite counts, which wouldn’t have helped in the resolution of the missing satellites problem. Instead, they corrected a luminosity curve using

565 Tollerud et al., 2008, pp. 277-278.
566 Tollerud et al., 2008, p. 282.
first, the technical limits of the observational technology used to conduct the survey; and second, the technical un-limitations of simulation. In the secondary corrections applied to the SDSS data, simulation played on both side of the theory/observation fence. As a technology of theory, simulation dynamically modelled the galactic satellites in much the same way the 1999 simulations did, creating a theoretical prediction against which observation is compared. Except in order for observation to be compared to the theoretical data (expressed in simulation), the observational data itself needed assistance from simulation. In Tollerud et al, simulation temporarily took on the role of 'observation', producing mock surveys whose false data is used to bolster the observed luminosity curve. Simulation then reassumed a more theoretical approach, producing a simulated luminosity curve which could then be compared to the 'observed' curve since it has been corrected for completeness. There is something of a dialectic here, a negotiation between the simulated data and that produced by observation, one governed by virtual limitations and the other by material, in order to make an empirically robust claim about the presence (and number) of unobserved objects.

Thus Tollerud et al were led to conclude that there are likely between 300 and 600 very faint satellites within the Milky Way's virial radius (possibly more depending on assumptions):

As it was first formulated, the MSP referred to the mismatch between the then ~10 known dwarf satellite galaxies of the Milky Way and Andromeda, and the expected count of ~100-500 subhalos ... Our results suggest that the recent discoveries of ultrafaint dwarfs about the Milky Way are consistent with a total population of ~500 satellites, once we take into account the completeness limits of the SDSS. In this sense, the primary worries associated with the MSP in CDM are alleviated. Nonetheless, it is critical that searches for these faint galaxies be undertaken, as the assumptions of this correction must be tested.568

Solving the missing satellites problem, in the context of the CDM model, required future observations that would reveal the same amount of very faint satellites that the 1999 simulations predicted. The first round of improved observations only partially lived up to this hope; in count, the results were disappointing. Instead, the chief value of the SDSS data lay in increasing the plausibility of the existence of more dark satellites. Rather than the likelihood argument expressed in Mateo's review – that since we have discovered so many satellites in the last few years (to 1998) it is possible there are several more to be found – the SDSS completion limits gave a much stronger statistical argument. That is, the number of satellites left to be found had been reduced to a degree by direct observation, and the rest of the missing satellites had increased in

568 Tollerud et al., 2008, p. 288. The spatial distribution from Via Lactea is based on a standard CDM model, which means that the spatial distribution Tollerud et al used was therefore not independent of a CDM model. This, however, is rather the point – observations completed using correction factors based on a CDM spatial distribution showed the correct amount of satellite galaxies, which meant that the missing satellites problem could be resolved within the existing CDM model. In addition, the correction factors were independent of the total count of satellites from such a model – important, since this was the cause of the missing satellites problem in the first place – and there was nothing to indicate that the spatial distribution was wrong.
plausibility. Tollerud et al., after claiming that the missing satellites problem was mostly solved, ended their paper with a sentiment that is by now very familiar:

Fortunately, future deep large sky surveys will detect very faint satellites out to much larger distances and hence firmly observe the complete luminosity function out beyond the Milky Way virial radius ... Nonetheless, the current data are not deep enough, and until the new survey data are available, there will be no way to put the spectre of the MSP completely to rest.\textsuperscript{569}

On one hand, the completed data set alleviated the main worries associated with the missing satellites problem – that the CDM model needed to be overhauled in some manner. It was quite plausible that we could observe the correct amount of dark satellites, and no doubt we would in future. On the other hand, the fact remained that these satellites had not actually been observed, the pesky requirement of empirical confirmation preventing a satisfactory resolution.

Overall, the conclusion of Tollerud et al was an optimistic one, both from a theoretical perspective (the CDM model is not flawed) and from an observational one (the existence of the required amount of dark satellites is quite plausible). Andrey Kravtsov, however, was not so easily satisfied. He agreed that “we can reasonably expect that at least a hundred faint satellites exist within 400 kpc of the Milky Way\textsuperscript{570} from future surveys, especially given that the new satellites discovered by SDSS are all within 50 kpc of the Milky Way and there are probably more at greater distances. However:

Even though the discovery of the ultra-faint dwarfs implies the possible existence of hundreds of them in the halo of the Milky Way (this fact has been used to argue that the substructure problem has been ‘alleviated’), the most recent simulations show that more than 100 000 subhalos of mass $m_{\text{sub}} > 10^5 M_\odot$ should exist in the Milky Way ... The substructure problem stated in the actual numbers of satellites is therefore alive and well and has not been alleviated in the least.\textsuperscript{571}

Via Lactea II, according to results published in 2008, produced 1.7 to 2.6 times more substructure than its predecessor Via Lactea (the simulation used by Tollerud et al), “probably because of the improved mass and time resolution of Via Lactea II, which enabled better resolution of inner substructure”.\textsuperscript{572} Via Lactea II did not produce more substructure because it had a different cosmology, but because it had improved time stepping, higher resolution, and more precision than Via Lactea, enabling the newer simulation to more clearly count the amount of expected sub halos.\textsuperscript{573} The 2008 simulation Aquarius, which simulated six Milky-Way sized dark halos, likewise showed “substantially more substructure than reported for Via Lactea I. This is particularly

\textsuperscript{569} Tollerud et al., 2008, p. 288.
\textsuperscript{570} Kravtsov, 2010, p. 7.
\textsuperscript{571} Kravtsov, 2010, p. 9. Emphasis in original. Kravtsov does not cite Tollerud et al directly, but it seems plausible that that was the paper he had in mind (particularly the reference to ‘alleviated’).
\textsuperscript{572} Diemand et al., 2008, p. 737, in caption to Figure 3.
\textsuperscript{573} Supplementary information to Diemand et al., 2008.
evident at lower subhalo masses." Aquarius produced 3.1 times as many subhalos as Via Lactea.

Like the authors of the two Via Lactea simulations, the authors of the Aquarius simulations agreed that "the substantial difference ... must have a systematic origin", one that emerged from the difference in computational capabilities of the simulation rather than from a difference in the model.

It seemed, therefore, that while we had better observations that increased the number of actual low luminosity satellites observed, we then had better simulations that increased the number of satellites required to be observed. The appeals to ‘future observations’ to find the missing satellites went largely unrealised, as SDSS found some, but not nearly enough. ‘Future simulations’ undermined what might otherwise have been very encouraging data, but even without the increased predicted numbers of satellites, SDSS did not actually provide direct empirical proof that there is existed a very large amount of dark (not merely faint) satellites.

Yet despite this, and despite Kravtsov’s cautionary pessimism, the SDSS data, suitably corrected, did prove that the existence of many dark satellites was sufficiently plausible. This is not ‘plausible’ simply in the sense that we haven’t proved that the satellites don’t exist, or even in the weak sense that the detection of more faint satellites suggests that better observations may find even more in the future. That the dark satellites are plausible becomes an empirically robust claim. Simulation plays an important role in this: on one hand, simulation as a technology of theory ‘worsened’ the theoretical prediction, but on the other hand, in a dialectic with observation technology, bolstered the observational data enough that the dark satellites appeared on the ‘observed’ luminosity curve. The corrected curve was the result of the dialectic between two technologies: first, the raw data gathered in the ordinary observational manner. Second, the data corrected for coverage completeness based on statistical analysis of the material limitations of said observational instruments. And third, depth of field completeness based on simulated observations of satellites in a CDM universe with an assumed spatial distribution.

The missing satellites are not (quite) reified through this process. Tollerud et al’s analysis, and the more restrained corrections of the SDSS luminosity curve used in other studies, does not provide proof of ontology – the very use of simulation prevents such a bold realist claim. Despite this, the missing satellites problem does have a resolution. Through the brief decade or so that this section has covered, both the technologies of observation and that of simulation have improved. In the case of observation, this can be characterised as progress in instrumentation, improved techniques of analysing data (sometimes, ironically, involving simulation) and better theoretical understanding in data interpretation. In the case of simulation, progress in the realms of computer science and engineering have resulted in both improved computational capabilities and algorithms to make more efficient use of those capabilities. Ten years after the missing satellites problem was formulated, despite these improvements in technology, by all account the problem in numbers remained essentially the same. Yet there has been recent work that claims, on the strength of the same data described in the previous section, that the missing satellites

---

574 Springel et al., 2008, p. 1695.
575 Springel et al., 2008, p. 1696.
problem has been solved, even though the numbers remain as discrepant as ever. The next two sections will describe first the ‘solution’ that has been developed, and then how this solution came about.

4.5 The ‘Solution’

Even if the SDSS data established that a large number of dark satellites was empirically plausible, what remained was developing a physical mechanism for the creation of those dark satellites that was equally theoretically plausible. It will perhaps be unsurprising that the most popular modern ‘solution’ to the missing satellites problem is in the form of an astrophysical solution, featuring the combination of several of the mechanisms proposed after the 1999 simulations. In 2010, to take an illustrative example, the authors of a study on the luminosity function of satellites galaxies within Milky Way sized halos wrote:

In light of the discovery of the new ultra-faint dwarf population and the improvements in the numerical modelling of galaxy formation, it is now timely to revisit the issue of whether the basic properties of satellite galaxies around the MW, such as their number density, radial distribution, and mass-to-light ratios, can be reproduced within current cosmological $\Lambda$CDM-based models. It is also interesting to ask what physical processes might plausibly give rise to this population of extremely low-luminosity galaxies.\textsuperscript{576}

The discovery of the new faint dwarf population that Andrea Macciò and collaborators referred to is that of the SDSS data described in the previous section. The improvements in numerical modelling that the authors made use of were mainly related to an increase in resolution. In order to test the effect of various physical processes on their models, the authors used a combination of merger trees from very high resolution simulations with various semi-analytic models. The merger trees described the hierarchical assembly of the Milky Way-like halo, while the semi-analytic models were used to predict the relationship between the dark matter subhalos and observable galaxy properties, “allowing us to make a direct and detailed comparison with observational data.”\textsuperscript{577} In order to test the effect of various mechanisms for developing (or suppressing) galaxies in the low mass substructures, the semi-analytic models parameterised in different ways the mechanisms acting on the baryonic component. This use of semi-analytic models allowed the authors to understand the effect of the astrophysical mechanisms on galaxy formation, which in this case involves a heavy baryonic component, within the context of a dark-matter only simulation.

Since one of the main goals of this study was to compare the simulation results to actual data, Macciò et al used observational selection criteria to determine which of the satellites in the simulation were likely to be observed. All satellites brighter than a certain luminosity, for

\textsuperscript{576} Macciò et al., 2010, p. 1996. The model has of course changed from ‘CDM’ to ‘$\Lambda$CDM’ due to the work on the cosmological parameter and dark energy that occurred in the late 1990s.

\textsuperscript{577} Macciò et al., 2010, p. 1996.
example, were understood to be included in the SDSS sample, but the detection limits for fainter satellites depended on satellite distance and luminosity: "it is also interesting to apply the observational selection criteria to our simulations and compare 'in the observational plane', i.e. with the raw data from the SDSS without completeness corrections applied."$^{578}$ Randomly selecting one-fifth of the simulation satellites (since SDSS only covers one-fifth of the sky) seemed to give an "agreement between the data and models [that] implies that the distance–luminosity relation of our satellites is similar to the observed one."$^{579}$

The outcome of the study was that the luminosity function "is shaped by a complex combination of different physical processes including tidal destruction, photoionization and [supernova] feedback."$^{580}$ Supernova feedback, broadly speaking, aids in star formation suppression as the winds from nearby supernovae reheat cold gas and expels it from the small halos. Macciò et al found that all these mechanisms were required: no mechanisms gave no very faint satellites (and we do observe some); and photoionisation and supernova feedback worked together, with the former reducing the supply of baryons available for cooling and the latter reheating cold gas – using only one of these mechanisms gave an excess of intermediary luminous satellites and few faint ones. The tidal destruction of satellites had important effects for satellites at the faint end of the spectrum. In a 2011 study, Andreea Font and collaborators come to essentially the same conclusion with the observation that the relative contribution of the mechanisms varied across the range of dwarf galaxy sizes – supernova heating was more important for massive systems, while photoionisation played the main role for the very faint dwarfs;$^{581}$ see Figure 17. Note that the light blue solid line with represents the best combination of astrophysical processes extends beyond the black line representing the observed data, into the ‘dark’ area of the graph and thus implying the presence of unobserved satellites.

A combination of these astrophysical mechanisms, therefore, was necessary for matching observation:

In final summary, our results show that not only is there no longer a ‘missing satellite problem’, but that well-known and well-motivated astrophysical processes working within the $\Lambda$CDM framework naturally predict the form of the observed [luminosity function]. Indeed, it may be that convincing proof of the existence of the large predicted population of dark subhalos via one of the methods suggested above (or one not yet discovered) is one of the last remaining major challenges for the $\Lambda$CDM paradigm.$^{582}$

Macciò et al were confident enough to state that the missing satellites problem had been solved (or demonstrated not to occur) within the CDM framework – in fact, the inclusion of the astrophysical processes naturally predicted the observed luminosity function. Though it is too

$^{578}$ Macciò et al., 2010, p. 2001. Emphasis in original.
$^{579}$ Macciò et al., 2010, p. 2001.
$^{580}$ Macciò et al., 2010, p. 2006.
$^{581}$ Font et al., 2011.
$^{582}$ Macciò et al., 2010, p. 2007.
early to unequivocally say that the missing satellites problem is solved, it seems that in recent years this sort of astrophysical solution has been the most popular; Font et al, in 2011, considered that “a consensus seems to be emerging then that the abundance and other properties of the MW satellites can be understood as a consequence of the known physics of galaxy formation in a ΛCDM universe”.

4.5.1 Hydrodynamic Simulations

With the increasing popularity of the “‘baryonic’ solution of the ‘missing satellite problem’”, it has also become apparent that despite the successes of the CDM model on large scales, to obtain equally convincing success on small scales necessitates the inclusion of astrophysical processes. In order to adequately model such processes alongside the dark matter, hydrodynamic simulations are required. Most of the simulations studied in this and the previous case study are dark matter only simulations that treat all matter components as a collisionless fluid, with the

---

583 Font et al., 2011, p. 1261.
584 Macciò et al., 2010, p. 2007.
assumption that baryonic effects are mostly negligible. That is, the total mass within a simulation could be effectively modelled as governed by gravity alone, with dark matter naturally making the largest contribution. On large scales, this assumption seemed quite valid. On smaller scales, baryonic processes like star formation and feedback were modelled as limited 'subgrid' processes, described using simple models as the simulation did not have adequate resolution on such scales. It was difficult to simulate both the stellar and gaseous components at once: "many of these processes occur on scales too small to be meaningfully resolved, and the physics of star formation and energy feedback in particular is poorly understood; the simulators simply include ad hoc rules to mimic them."\textsuperscript{585}

There was the ever pervasive difficulty in balancing the demands of physical calculations with the volume covered by the simulation. Poor theoretical understanding of small-scale baryonic processes also played into the difficulty in developing an accurate model. As a result, "no single, self-consistent simulation of the Universe was able to simultaneously predict statistics on large scales ... together with galaxy properties on small scales ... The challenge lies in following the baryonic component of the Universe using hydrodynamic simulations."\textsuperscript{586} High resolution hydrodynamic simulations are very computationally expensive and have only begun to appear in the last two or three years.\textsuperscript{587} They model gravity and fluid motion (hydrodynamics) across a wider dynamic range that was previously possible, attempting to express baryonic and dark matter interactions on both large and small scales. This allows the implementation of more realistic, physically-motivated models for the baryonic processes.

One such hydrodynamic simulation is Illustris, for which "rapid advances in computing power combined with improved numerical algorithms and more faithful models of the relevant physics have allowed us to produce a simulation ... that simultaneously follows the evolution of dark matter and baryons in detail."\textsuperscript{588} Illustris is a cosmological simulation, following the evolution of a statistically significant volume with high resolution over a long time scale, up to the present epoch. The simulation is able to demonstrate a mix of galaxy morphologies (previous simulations had trouble obtaining a representative mix of spirals and ellipticals), with observationally compatible metallicities, as well as predicting the large scale distribution of neutral hydrogen, and the radial distribution of satellite galaxies within galaxy clusters. Another hydrodynamic simulation, EAGLE (Evolution and Assembly of GaLaxies and their Environments) is able to demonstrate similarly excellent results.\textsuperscript{589} Processes that are modelled in these simulations include the cooling and heating of gas due to the presence of stars; the formation, evolution and aging of stars; the distribution of energy and metals generated by the stars and injected into surrounding gas; the explosion of supernovae and distribution of their energy; the formation of

\textsuperscript{585} Sellwood & Moore, 1999, p. 127.
\textsuperscript{586} Vogelsberger et al., 2014a, p. 177.
\textsuperscript{587} Previous hydrodynamic simulations "either did not cover a large enough portion of the Universe to be representative, lacked adequate resolution, or failed to reach the present epoch" (Vogelsberger et al., 2014a, p. 177). EAGLE and Illustris are to hydrodynamic simulations what Millennium was to dark matter only simulations.
\textsuperscript{588} Vogelsberger et al., 2014a, p. 177. See also Vogelsberger et al., 2014b.
\textsuperscript{589} Schaye et al., 2015.
supermassive black holes; and the accretion of mass and ejection of energy around black holes. Very few of these processes are modelled in dark matter only simulations, and are almost never directly implemented. Astrophysicists usually have to make do with simple parameters for star formation and energy redistribution.

The choice of astrophysical 'solution' to the missing satellites problem seems to demand that these processes be explicitly modelled alongside dark matter in order to fully describe the situation. Using the EAGLE code, Till Sawala and collaborators recently simulated a Local Group population, making full use of our nearest and richest 'cosmological probe': "applying state-of-the-art hydrodynamic simulation techniques to a realistic local environment allows us for the first time directly to confront ΛCDM predictions with observations in the critical, subgalactic regime."

Focusing on pairs of halos that corresponded to the velocity and separation of the Milky Way and Andromeda, Sawala et al were able to contrast the dark matter structure with that of luminous matter (see Figure 18). The simulations gave an excellent count of the number of luminous satellites, with an average of 90 within the simulation, which matches the count of around 60 observed (after SDSS) very well within error bars. Of course, there was also a very large number of dark matter halos forming substructure, but the simulation was able to demonstrate that astrophysical processes have kept most of these halos dark, and that therefore we would not observe them. Sawala et al proudly stated that "our simulations [are] free of the 'missing satellites' problem".

This seems to be the culmination of the missing satellites problem. The SDSS data had made the existence of dark satellites sufficiently plausible that counts of satellites was no longer a robust way to compare observations and simulations – since dark satellites could not be observed and therefore not counted empirically. Instead, a luminosity function was constructed out of observation and simulation data so that the predictions of the cluster simulations could be more

---

590 Sawala et al., 2014, p. 2.
591 Sawala et al., 2014, p. 3.
rigorously compared to observation. With the inclusion of baryonic processes, however, we can
once again perform this comparison with counts, because the hydrodynamic simulations tell us
which of the dark matter structures are sufficiently luminous to be observable. This solves the
original formulation of the missing satellites problem, which was a difference in counts between
simulated satellites and observed ones, with the caveat that there must be numerous extra
satellites – which have since transitioned from 'missing' to 'dark'. Sawala et al in fact described
the missing satellites problem as: "there are far fewer satellite galaxies that dark matter halos".\footnote{592}
This formulation makes the problem one of mechanisms of star formation suppression rather
than one of observation versus simulation, and includes the implication that there are numerous
dark halos, ruling out cosmological solutions that would suppress this number.

The authors of Sawala et al, many of whom are naturally participants in the EAGLE project, heavily
emphasised how important including baryonic processes is in modelling galaxy formation. The
authors of Illustris also "stressed that a direct modelling of baryonic physics is essential to
soundly connect cosmological predictions to galaxy observations."\footnote{593} In a dark matter only
version of the Illustris simulation, the radial distribution of satellites galaxies was shown to be
too flat towards the centre, which "directly demonstrates that neglecting baryonic physics causes
inaccuracies",\footnote{594} possibly leading to even greater errors down the road. The inclusion of baryonic
effects is one step closer to "realism of the simulated galaxies",\footnote{595} which EAGLE considers an
essential part of the "theorist's armoury",\footnote{596} especially for confrontation with future observations.
This seems difficult to deny, as astrophysicists have struggled for many years to develop
unambiguous predictions from their models for comparison with decidedly ambiguous data. The
more complex simulations become (and it is impossible to obtain a realistic galaxy model without
simulation), the better we can understand the theory behind galactic processes, and the better
we can compare this theory to the messiness of actual galaxies. As the authors of the OWLS
(OverWhelmingly Large Simulations) project put it, "the advantages of the simulation approach
include the much reduced (though still present) risk of getting the right answers for the wrong
reasons [and] the ability to ask more detailed questions due to the tremendous increase in
resolution."\footnote{597}

4.6 Technical Negotiation

If we have yet to explicitly find the dark satellites (or sufficiently demonstrate their
nonexistence), how then can the hydrodynamic simulations boast the absence of 'the missing
satellites problem'? This section will look at this achievement from the perspective of the
technologies involved, and the next section will explore in more depth the epistemology of the
matter (which, of course, also includes the intersection with simulation and observation). In order

\footnote{592}{Sawala et al., 2014, p. 1.}
\footnote{593}{Vogelsberger et al., 2014a, p. 180.}
\footnote{594}{Vogelsberger et al., 2014a, p. 178.}
\footnote{595}{Schaye et al., 2015, p. 522.}
\footnote{596}{Sawala et al., 2014, p. 4. a}
\footnote{597}{Haas et al., 2013, p. 2932.}
to more clearly understand just how the technologies of observation and simulation worked together to achieve the ‘solution’ in this section, it will be instructive to take a brief look at the problem that can perhaps be called the forerunner to the missing satellites problem.

The hierarchical clustering model, as described at the very beginning of this chapter, implies the existence of substructure. More than this, it can be observationally determined that “clusters in the real universe contain many galaxies that have retained their dark halos (at least within their optical radii).” Simulation models of the formation of rich galaxy clusters in the early 1990s, however, showed that no halos survived within half the radius of the cluster centre. This tendency of simulated galaxies to merge far more efficiently than the real thing was known as the ‘overmerging’ problem: “in the central regions of a cluster ... the overmerging erases not only large-scale substructure, but also any trace of small halos that could be associated with ‘galaxies,’ leaving a smooth giant lump of dark matter.” This problem was very similar in formulation to the missing satellites problem in terms of the mismatch between observation and simulation, but somewhat more clear-cut in its objects – we definitely do see galaxies! The problem was therefore either in the simulation or the model. The ‘traditional explanation’ was to assume that because the simulation model was dominated by dissipationless dark matter (and the dissipational baryons were not included), the halo was modelled with a much looser core that was more prone to destruction.

If this solution was adopted, it would imply that baryons played a much more important role in the formation of galaxies than generally assumed. Actually enacting a non-dissipational model, however, was prevented by the limitations of simulations – it was one of those solutions that appealed to later advancements in technology, this time simulation. Another proposed solution to the overmerging problem made the same appeal, but in a far more pragmatic sense, arguing that the dark matter halos by themselves are dense enough to survive and be identified observationally (if they host luminous galaxies):

With a sufficient computational effort the overmerging problem can be tamed (at least to some extent) even in purely dissipationless simulations. The computational costs are higher than the cost of an average cosmological N-body simulation. However, they may be considerably lower than the computational expense of the corresponding N-body+hydro simulations.

‘N-body+hydro’ refers to N-body simulations that include hydrodynamics, and are therefore able to model baryonic and dark matter components simultaneously. The computational cost of such simulations is very high; from the perspective of 1999 appealing to that technology to test the dissipational solution to the overmerging problem would very much ‘in the future’ – indeed, it would take another fifteen or so years before hydrodynamic simulations were realised on such a

---

599 Klypin et al., 1999a, p. 521.
600 Klypin et al., 1999a, p. 531, as the authors cite papers from 1992 and 1995.
601 Klypin et al., 1999a, p. 531.
scale. In the late 1990s, it sufficed to demonstrate that “the overmerging problem is mostly due to the inability of a numerical code to provide sufficient numerical resolution”.

Though the resolution required was still reasonably high, it was not impossibly so; workarounds to this limit of simulation technology included better halo-finding algorithms and using smaller simulation boxes. KKVP, for example, simulated a pair of large galaxies rather than a cluster, reducing the size of the simulation to a few megaparsecs (rather than the hundred or so that would be statistically significant for cluster size).

Thus the conflict between observation and simulation was, for the overmerging problem, resolved through improved simulation technology. In the course of determining that this was the case, physicists increased their understanding of the sorts of mechanisms that resulted in the destruction or survival of halos. However, the price for this improvement in simulation technology was the trade of one problem for another. As simulation technology improved in the early 2000s, physicists were able to create counts of substructure halos. This revealed the missing satellites problem, a problem born in part from technical limitations as much as technical affordance. Astrophysical effects were not being rigorously modelled in simulations because even obtaining enough resolution to reveal substructure was difficult – early simulations of substructure still made use of semi-analytic models, and KKVP had quite severe restrictions on resolution and space. If astrophysical effects had been included, perhaps the missing satellites problem would have been formulated differently: not in terms of a difference between simulation and observation, but as part of the ongoing problem of detecting dark matter structures.

As it was, limitations from observation technology also played a role: first, in the negative sense that "it is probably unreasonable to attempt a detailed fit to the satellite population of the Local Group because of observational incompleteness and poor statistics." Any comparison between observation and simulation could only be done on the basis of counts, and was open to large uncertainties. Second, in the positive sense, there was plenty historical evidence to suggest that poorer instruments (older observation technology) had missed satellites before, and might do so again. There was therefore every expectation that improvements in observation technology would resolve the missing satellites problem, either by finding the missing satellites, or convincingly demonstrating their lack. Observation dutifully stepped up its technological game and delivered the SDSS data, which did find some darker satellites – but not enough. The new evidence failed to follow the classical narrative, and then further improvements in simulation technology increased the required satellite count.

It is perfectly accurate to say that despite these technological improvements, the missing satellites problem in 2008 was the same in terms of numbers as it was a decade earlier in 1999. Yet undoubtedly the nature of the problem has changed – going so far as to become ‘solved’. This is because each instance of technological improvement progressed both the debate around missing satellites, and the physics itself. When simulation technology improved enough to resolve

602 Klypin et al., 1999a, p. 549.
603 Kauffmann et al., 1993, p. 208.
the overmerging problem,\textsuperscript{603} it also answered the question of whether the problem lay in the assumption that baryonic effects were negligible. This is a very common and fairly fundamental assumption made for both technological and theoretical simplicity, and gaining the technical capacity to include baryonic effects may have important implications for theory crafting (somewhat ironically, baryons then later turned out to be very important for resolving the missing satellites problem). For the missing satellites problem, the improvements in technology were accompanied by the reworking of theoretical assumptions and understandings.

The release of the SDSS data was limited in its usefulness as direct evidence. What the SDSS did not do was convincingly resolve the missing satellites problem according to the original way it was stated. Counts of observed satellites versus counts of simulated satellites remained too different. Attempts to complete the SDSS data (beyond coverage completion) may have been enough to solve the original 1999 problem, but a further iteration of simulation technology made such arguments weaker than they would perhaps otherwise have been. What the SDSS data did do, however, was drastically increase the empirical plausibility of large numbers of dark satellites. While this data was not robust enough by itself to solve the missing satellites problem, the SDSS survey, like the rotation curves of the previous chapter, did provide the crucial piece of evidence required to demonstrate that the existence of dark satellites was plausible in an empirically meaningful way. Astrophysical solutions became more popular than cosmological ones, and claims began to emerge that the missing satellites problem was adequately solved within the CDM model. The hierarchical CDM model, which gave rise to the missing satellites problem, also provided a solution for it!\textsuperscript{605} On the one hand, the missing satellites problem was solved without changing anything about the CDM model. On the other hand, only using the CDM model to simulate galaxy formation would lead to the missing satellites problem. Dozens of simulations and rigorous interpretation of observational data were required precisely so that nothing changed; Font et al summarised the situation quite neatly:

\begin{quote}
There have been significant and rapid advances over the past 5 years in both observations of dwarf satellites and our theoretical understanding of how they form. Despite this rapid progress, the conclusions of the previous generation of models remain essentially correct. The overall results seem independent of the details of any specific implementation of galaxy formation physics and lead to the conclusion that the broad visible properties of the satellite population are reproducible within the current CDM paradigm. This represents an important success. Furthermore, the CDM cosmogony makes clear predictions for the DM
\end{quote}

\textsuperscript{604} Pun intended.

\textsuperscript{605} This may be an instance of Hasok Chang's epistemic iteration, where successive stages of knowledge build on the preceding one in a manner that is not straightforward but still progressive. This sort of knowledge generation achieves justification through coherentism: beliefs are justified insofar as they belong to a system of beliefs that is mutually supportive (Chang, 2004). There is a clearer example of this process in the previous chapter, where Toomre's criterion was 'confirmed' by simulation, and then used to justify a different but similar simulation. This form of knowledge production is not exclusive to those using simulation, and seems common in situations where there is little in the form of arbitrating evidence.
content of dwarf satellites that are mostly independent of the baryonic physics and are eminently testable by observations.\textsuperscript{606}

It seems curious to write that the broad visible properties of the satellite population are reproducible in the CDM model, independent of the galaxy formation process, when it is precisely these latter processes that ensure that the luminosity function of the simulated satellites matches that observed. To make sense of such a statement, two things need to be borne in mind. The first is that throughout the debate theory concerning overall cosmology is kept separate from theory about galaxy formation. It’s difficult to draw this distinction in any meaningful way, as naturally cosmology affects how galaxies form – in fact, this was one of the main reasons the missing satellites were such a problem. However one can see how astrophysical mechanisms like supernova feedback and tidal stripping occur somewhat independently of the cosmological model; they happen whether we think dark matter is cold or warm. This is the sense in which Font et al understand their results as independent to and therefore reproducible in the CDM model, because it is purely by varying the astrophysical processes that they can obtain the observed luminosity function.

However the fact remains that if these processes are left out of a simulation of galaxy formation – that employs nothing but a CDM model – the correct luminosity function will not be obtained. To claim a success for CDM, and to understand the missing satellites problem as resolved within CDM, it must also be understood that part of the reason the original problem emerged was because our simulations did not include the quotidian processes of astrophysical phenomena. Technically speaking, there was nothing wrong with CDM to begin with; if only the 1999 simulations had been able to more realistically model all local processes in addition to the cold dark matter cosmology, there would have been no conflict with observation. The extra satellites in the simulation would appear as dark satellites, instead of halos of unknown luminosity (in fact, this is now possible; see Figure 18). Claiming that the missing satellites problem is solved within CDM becomes a technoscientific claim, in the sense that it is a claim about the technologies involved in generating the solution. Simulation technology revealed, first, a problem with the theory (discrepancy between theoretical prediction and observation). Solving this problem required the addition, to the theory of galaxy formation, of more complex models of baryonic astrophysical processes. But implementing this solution required simulations with better resolution and better algorithms for modelling the subgrid processes – a technological advancement.

It is the combination of observation and simulation technologies that sufficiently demonstrated the plausibility of both the dark satellites and the physical mechanism for making them. The luminosity function provided a measuring stick against which various combinations of astrophysical processes could be measured. A plausible mechanism of star formation suppression (one that reproduced the correct luminosity function) emerged from these comparisons, comparisons which could only be made rigorously with simulation. In return, the good fit of such

\textsuperscript{606} Font et al., 2011, p. 1275.
models to the observed function made plausible a good number of dark satellites that extended beyond the empirically determined luminosity data.

This apparently circulatory justification of a ‘completed’ data set and the new model is actually the outcome of an extended dialectic between simulation and observation, bracketed by the capabilities of their technologies, negotiating progress in the absence of equivocal evidence. Simulation, where required, forms a data-like substitute; the same tendency was observed in the previous case study. When not required, simulation plays the role of a technology of theory, embedding the production of knowledge in the technology used to produce it. The missing satellites problem has ‘disappeared’ from certain recent simulations only due to a complex process of simulating and data analysis. This process is one of negotiation, a back-and-forth between simulation, observation, and theory as physicists attempted to work within technological limitations while still progressing knowledge in some way. The earliest simulations used a model that did not include astrophysical effects like reionisation and feedback from supernovae. Their comparison with observational data could not be rigorous, and therefore relied on counts. The difference in counts meant there were ‘missing’ satellites. In later simulations, there were still satellites we did not observe, but this time they were not ‘missing’. There was both a plausible mechanism for keeping them dark and a plausible assumption that there existed dark satellites: in the first simulation is a technology of theory, and the second it plays both the latter role and one of empirical substitution.

Developing this plausibility (and in fact, understanding the existence of the satellites in terms of plausibility at all, see next section) required that simulation data and observation data become intimately intertwined. Especially in the case of the luminosity curve: observation needed simulation to help complete the data set, and simulation needed observation to give empirical force to the existence of its predicted data satellites. More than this, the inclusion of the astrophysical processes, the demonstration that these processes gave the correct luminosity function, and the legitimacy of the luminosity function in the first place could not have been done without simulation. Hélène Guillemot points out that in climate simulations, “the referent to which simulations are compared is not ‘the climate’, it is a set of remodelled data that are carefully selected from hundreds of thousands of data sets furnished by instrument networks … Only an instrumented world is capable of providing the data that are used to test models.” As we need data to test our simulations, so we need simulations to make that data legible. Simulation allows, in effect, the construction of a world in which the dark satellites exist, and then also enables the superposition of that world with ours in order match representations and cover both theoretical and empirical gaps. Here superposition implies the overlap and comparison of both worlds, but in an additive sense – gaps are filled in, and existing knowledge is increased. The new model of galaxy formation is a simulation model – not only created in simulation but which cannot be accessed outside simulation (for sheer complexity) – and one whose theory is explored, produced,
and intervened with by simulation, whose empirical data is woven together with theoretical models: this is the ‘technoscience’ of astrophysics.

4.7 Simulation as an Epistemic Target

To better understand how simulation technology has helped generate knowledge in this case study, it is instructive to draw an example from the (material) technoscience of nanotechnology. The scanning tunnelling microscope (STM) makes its observations by interacting with its target through the means of a probe. Though this near-sightedness is a direct contrast to the scale on which astrophysical simulations usually operate, the way in which STM data is used has some similarities with the current case study. First, the STM experimental setup produces a set of rough data. Second, a simulation of this setup, in which basic models of nano behaviour are incorporated, produces its own ‘data’, and then:

For the purpose of ‘explanation’ the ectoplasm-like images and the manipulation signals that constitute the rough data are then compared with those produced by a simulation of the whole experimental setup. These simulations use models that are calibrated to make theories fit with the experiment. They do not test theories. Rather, they simulate the interpretation of theories in a back-and-forth process with the experiment, until the two present a sufficiently satisfactory likeness.

As the simulated STM and its objects are governed by theoretical models, comparing their behaviour to that of the real system assists in theoretical explanation of phenomena. Yet, as Bernadette Bensaude-Vincent and her fellow authors underscore, this is not theory testing, but more akin to what has been called in this thesis theory crafting. That is, the construction of a model of phenomenon or object works in negotiation with empirical data and the simulation of that data to produce a sufficient likeness. This chapter began with a poor understanding of the physics of galaxy formation, but has shown how simulation has, in negotiation with the luminosity function, developed a galaxy formation model that has a complex balance of various astrophysical effects and an increased understanding of these effects on small scales.

What is important here is that this negotiation is a dialectic. The simulation technology in the missing satellites problem speaks both ways: to theory, and to the empirical data. In simulating the universe based on CDM theory, astrophysicists do more than provide an arena in which to test their models. They also, in the technoscientific sense, construct a self-sufficient world. In order to rigorously compare simulation and observation, the simulation must have sufficient complexities that it can mimic, in its own right, an independent system that can form a legitimate epistemic target. It is also because simulation can construct this sort of (theoretical) world that simulation data can contribute to the correction of an observed luminosity function, and the outcome of the missing satellites problem is the outcome of this sort of superposition of simulated

---

608 Not to be confused the other STM of technoscience literature: (studies of) science, technology, and medicine.
and observed data. Manipulating this virtual target – ‘experimenting’ upon it, ‘observing’ and ‘measuring’ it – through simulation enables us to draw out ‘empirical content’ of a theory. This is not merely the implied content that emerges from working out a theory’s equations, but a set of simulated measurements and observations that can actually function as empirical data in its own right.

Nordmann contends that in modern science, there is understood to be distance between the representation and its object of aboutness. However, for ease of metaphysics, we assume that there is some kind of immediate way we can determine the ‘aboutness’ relation between the two. Preventing the representation from being collapsed into the object, however, requires that this immediacy be constructed (and hence ‘artful’). Such artful constructions of immediacy between model and scientific object is the hallmark of modern science, with the positivist disclaimer that “we know nothing of reality because all we have are models that are constructed by us.”  

Essentially, Nordmann seems to be making the relatively uncontroversial claim that scientists construct representations of their objects, and take these models as adequate representations about said objects, but do not actually mistake such models for describing, god-like, reality itself. This Nordmann contrasts with technoscience, where the awareness of the fictional nature of the model is easier to lose and harder to maintain. Whereas in other areas of science, the immediacy between representation and object is consciously understood as constructed, technoscience effects an entirely involuntary ‘collapse of distance’:

> Technoscientific work is accompanied by the vague suggestion that we know everything of reality because the dynamic system in front of us is a self-sufficient reality in its own right – where this self-contained, non-referential dynamic system may be a computer simulation, a self-organizing algorithmic structure, or a model organism.  

It is essentially this collapse of distance that allows simulation to ‘insert’ dark satellites into the observed luminosity curve. There is a lack of representational distance between the simulated dark satellites and the observational data, because the simulation of the Local Group is not a representation of the real thing in the mimetic sense. It is Daston and Galison’s presentation, a constructed galaxy that forms a system in its own right.  

One symptom of this collapse of distance is replacing the ‘immediacy’ of quantitative agreement between predicted and measured values with qualitative agreement between calculated and experimental images. That is, between the visual output of a set of measurements, and the very similar output of the simulation of the measurement system. The hydrodynamic simulation Illustris provides a very powerful example of the lack of distance between simulated and observed that Nordmann is describing (though as will be emphasised, this collapse is, contra Nordmann, temporary). Figure 19 appears in the Illustris publication and is a comparison of

---

610 Nordmann, 2006, p. 20.
611 Nordmann, 2006, p. 20.
612 Though as we shall see, the representational capacity of the simulation is recalled through the comparison of the simulation’s evidential data with that produced by observations (or experiment).
“virtual mock observations that mimic the conditions that the Hubble Space Telescope encounters as it images galaxies” and an actual Hubble deep space image. The two images are not identical in content, but they are indistinguishable in type – it is impossible to tell, without the caption, which image was produced with a simulated telescope aimed at a simulated universe, and which was produced with a real instrument.

This image has been removed by the author of this thesis for copyright reasons.

Figure 19: Left: Hubble Space Telescope (HST) Ultra Deep Field (UDF) image, an observation. Right: HST mock observation from Illustris, a simulation. (Vogelsberger et al., 2014a, p. 178)

There is a difference between simulation being used as ‘experiment’ (a large preoccupation of this thesis) and a simulation of an experiment, like that of the STM simulation. It is in the similarity of output – especially visual images – between the simulation of the experiment and the experiment itself that Nordmann sees the collapse of distance. While astrophysics does not have any material experiments to simulate, it is not foreign to simulating both the act and the instruments of observation. The authors of Via Lactea, for example, used their simulation and “constructed all-sky maps of the expected annihilation γ-ray flux reaching a fiducial observer located 8 kpc from the center of the Milky Way halo.” These simulations are intended to be complementary to observation. One way in which dark matter could be detected is by searching for the products of pair annihilations of dark matter particles. If dark matter is a weakly interacting massive particle (WIMP) like the supersymmetric particle the neutralino, then pairs of these particles should ‘light up’ by their annihilation into gamma-rays (γ-rays).

In 2008, the Fermi Gamma-ray Space Telescope (also called Fermi-LAT – Fermi Large Area Telescope) was launched in order to conduct a sky-survey on the gamma-ray spectrum. There is some hope that by the end of its mission in 2018, Fermi-LAT will have detected signals that can

613 Vogelsberger et al., 2014a, p. 178.
614 Diemand et al., 2007a, p. 268.
be associated with gamma-ray annihilation from dark matter particles.\textsuperscript{615} In 2010, using the updated simulation Via Lactea II (VL2), Brandon Anderson and collaborators furthered the work done in the first iteration of Via Lactea by conducting “a study of the ability of the Fermi Gamma-ray Space Telescope to detect dark matter (DM) annihilation signals from the Galactic subhalos predicted by the Via Lactea II N-body simulation.”\textsuperscript{616} The Galactic centre has the biggest concentration of dark matter in the Local Group, but it is also a very active region. The signal from the Milky Way host halo is difficult to disentangle from background cosmic-ray interactions – the most detectable and least ambiguous sources for dark matter signals are therefore “either subhalos hosting dwarf satellites ... or one of the many dark subhalos predicted by numerical simulations”.\textsuperscript{617} Unlike the analysis published with Via Lactea, whose observer could ‘see’ the gamma-ray signals without issue, Anderson et al were interested in whether the Large Area Telescope (LAT) aboard Fermi was sensitive enough to detect the potential dark matter signal and differentiate it from other sources in the first place:

The present paper updates and improves upon that analysis by running the same VL2 all-sky emission maps and the diffuse background predictions through a Monte Carlo simulation of the Fermi–LAT instrument, taking into account the time-dependent sky coverage and an energy-dependent instrument response and angular resolution.\textsuperscript{618}

In other words, Anderson et al not only simulated the Milky Way and placed an observer in the system, they also simulated the observational instrument itself. In a section entitled ‘Observation Simulation’, the authors report that "the Fermi–LAT observation simulation program ... uses parameterized instrument response functions (based on detailed Monte Carlo simulations backed up by beam testing) to approximate the response of the LAT instrument in orbit."\textsuperscript{619} This simulated telescope was then pointed at the subhalo sky maps generated by Via Lactea II, simulating 10 years of viewing at 10 random viewpoints at an 8 kpc radius around the VL2 galaxy. Since the ‘real’ dark matter subhalo signal would need to compete with signals from extragalactic and Galactic sources, as well as host-halo and unresolved dark matter, this extra ‘noise’ was also simulated and therefore ‘observed’. Given a realistic treatment of LAT’s abilities, Anderson et al come to the sober conclusion that, contrary to other predictions, their simulated observation gives "less optimistic predicted sensitivity ... these factors leave room for very few expected detectable subhalos over the LAT’s lifetime, given a conventional DM candidate."\textsuperscript{620}

Like any good technoscientific technique, simulation constructs its target: simulation has the capacity to construct a self-contained system and produce measurements of that system. In its role as a technology of theory, this self-contained system is generally used to compare theoretical

\textsuperscript{616} Anderson, Kuhlen, Diemand, Johnson, & Madau, 2010, p. 899.
\textsuperscript{617} Anderson et al., 2010, p. 899.
\textsuperscript{618} Anderson et al., 2010, p. 899.
\textsuperscript{619} Anderson et al., 2010, p. 900.
\textsuperscript{620} Anderson et al., 2010, p. 904.
predictions with observation. When there is not (yet) data available for comparison, astrophysicists can occupy themselves with predicting what the data might look like, what to expect from future surveys, and how to analyse the empirical information. For example, for the authors of Millennium the goals of observational surveys are to shed light on the formation of galaxies, structure formation, dark matter, and dark energy, but “these goals can be achieved only if the accurate measurements delivered by the surveys can be compared to robust and equally precise theoretical predictions.”\textsuperscript{621} While these surveys have not yet been performed, astrophysicists can prepare ahead of time the tools required to interpret and analyse the data. These tools include: “detailed modelling of the selection criteria of an actual survey, and a thorough understanding of the systematic effects that will inevitably be present in real data. These issues can only be properly addressed by means of specially designed mock catalogues constructed from realistic simulations.”\textsuperscript{622} Simulations can assist in the interpretation of observational data because they can simulate the act of observing and in doing so track possible sources of error by comparing what is ‘observed’ to what is ‘actually there’. The advantage, of course, is that the ‘actually there’ is also simulated and therefore the ‘true’ value is known and can be compared to the ‘observation’. This information can be used to correct actual observational data. Other papers turned up in the course of research similarly use simulations to: test the method of determining the virial masses of groups of galaxies from observation;\textsuperscript{623} determine the minimum number of observations required to properly construct eclipsing binary light curves (including accounting for simulated observational errors);\textsuperscript{624} test a procedure on simulated catalogues for identifying and weighing clusters of galaxies in real redshift catalogues;\textsuperscript{625} and evaluate the experimental detection efficiency of gravitational lensing events, “including the addition of artificial stars to real data frames”.\textsuperscript{626}

As we simulate observations in order help with data analysis (and theory testing), we also simulate the act of observing (measuring, experimenting). This helps bring about the collapse of distance that is behind the blurring between the dark satellites in simulation and in the SDSS luminosity curve because it helps simulation construct its own self-sufficient system – scientists are not only simulating the system, but the observer in the system, and thus the act of observing. In the previous case study, simulating the act of experiment bled over into the language, which showed the use of both a metaphor and a narrative of experiment. In this case study, simulating the act of observing affects a collapse of distance between viewer and simulator: the results produced by the simulation inherit some of the immediacy of the results produced by a measurer of an experiment, the observer of a system, and take on some of the validity of that evidence.

In the previous case study, comparison between the simulated galaxy and photographs of real ones was used to help convince the reader that the underlying model was reliable, since it

\textsuperscript{621} Springel et al., 2005, p. 629.
\textsuperscript{622} Springel et al., 2005, pp. 633-634.
\textsuperscript{623} Aarseth & Saslaw, 1972.
\textsuperscript{624} Linnell & Proctor, 1970.
\textsuperscript{625} Press & Davis, 1982.
\textsuperscript{626} Alcock et al., 1996, p. 84.
produced something that looked very much like it ought to. The persuasive effect of the pictures was restricted by the text, which reassured the reader that the comparisons were not taken literally. In this case study, the comparative images in the hydrodynamic simulations are used for a similar rhetorical effect, but there is an added facet. One of the images produced by Illustris is not comparative, but simply a sample of simulated galaxies that demonstrates the range of morphologies achieved by the simulation (Figure 20); EAGLE includes something very similar in their publication.627 Illustris proudly tells us:

Galaxies in the mock UDF [Ultra Deep Field] appear strikingly similar to the observed population in terms of number density, colours, sizes and morphologies. Our model is the first hydrodynamic simulation from which a faithful deep UDF-like observation could be constructed, thanks to its combination of large volume and high resolution, along with the new numerical techniques that allow it to reproduce realistic galaxy morphologies.628

This image has been removed by the author of this thesis for copyright reasons.

Figure 20: Sample galaxies from Illustris arranged by standard morphological classification (Vogelsberger et al., 2014a, p. 178).

627 Schaye et al., 2015, p. 529, figure 2.
628 Vogelsberger et al., 2014a, p. 178.
EAGLE is similarly pleased by “the improvement in the realism of the simulated galaxies”. The key term here is ‘realism’. In taking a simulated ‘Hubble’ image of the galaxy, in ‘photographing’ various galaxy morphologies (the images are based on filters used in SDSS), Illustris and EAGLE are not only demonstrating that their underlying model reproduces accurate results, but that this model is close to reality. The closer to reality the visible aspects of the simulation, the more faith we have in the reality of the invisible, or dark, aspects (see especially Figure 18). In getting closer to realism, the hydrodynamic simulations extend the usefulness of this approach as they have the ability to create universe-like complexity. In its black-boxed capacity, this system is self-contained, producing a set of results that are its ‘data’. But this is temporary – denuded of its opacity, the simulation will become a simulation of, a theoretical representation contrasted with its target of study.

The emergence of plausibility as a legitimate criterion comes from the superposition of the simulation as constructed system on the simulation as representation. This seems like a necessary outcome for structures made of dark matter, as dark matter has not only yet to be directly confirmed but which would be difficult to detect even in principle. Perhaps, however, this is symptomatic of hydrodynamic simulations, or other such supersimulations. Stéphanie Ruphy takes a critical position on the possibility of validating large, complex simulations, suggesting there is a Duhemian problem that arises in composite simulations, where several sub-models come together. The more disparate models a simulation brings together, and the greater the simulation’s epistemic opacity, the less we can be certain of the simulation’s representational adequacy. Thus:

The stories or pictures delivered by computer simulations are plausible in the sense that they are compatible both with the data at hand and with the current state of theoretical knowledge. And they are realistic in two senses: first because their ambition is to include as many features and aspects of the system as possible, second because of the transformation of their outputs into images that ‘look’ like images built from observational or experimental data. I contend that computer simulations do produce useful knowledge about our world to the extent that they allow us to learn about what could be or could have been the case in our world, if not knowledge about what is or was actually the case in our world.

Problems of simulation validation aside, part of the reason Ruphy is hesitant to ascribe too much power to simulations is the very ability of simulations to look like genuine substitute systems – aided of course by sophisticated visualisation techniques. Obviously simulations cannot tell us, by themselves, what is actually the case in our world, despite looking like they can, as both scientists and philosophers have repeatedly reminded their readers throughout the course of this thesis (and see, again, section 4.7.2 of this chapter). It is through comparison and negotiation with many other systems of evidence that this knowledge comes about, inasmuch as scientists can ever be confident in saying what is ‘actually the case’. Simulations represent possibilities, or

---

629 Schaye et al., 2015, p. 522.
630 Ruphy, 2015, p. 144. Emphasis in original.
plausibilities, as much as models do, albeit more seductively. It is the superposition of these possibilities over or with the available evidence that enables scientists to gain some understanding of underlying mechanisms or causes. The more complex or 'real' the simulation, the more integrated this negotiation can become. Simulations have power precisely because they give what Ruphy calls modal knowledge – a range of worlds, a flexibility of theory that is then necessarily restrained and recalled by available evidence. Arguably, this has always been the case in science. With the integration of observational and simulation evidence, such possible worlds can become plausible, in a more robust sense than Ruphy implies. While these outcomes will naturally cause some philosophers and scientists to shy away, one might suggest that in the case of dark satellites or dark halos we are faced with objects that we have a compelling scientific reason to believe exist, and yet which by their nature remain resistant to understanding using purely material instrumentation. In dark matter studies, it may be that the negotiated simulation-observation evidence is the best (only?) way to make progress, though doing so may require the gradual reworking of some epistemic categories.

4.7.1 Simulation and Empirical Systems

According to Nordmann, the conflation of the model (simulated or material) with the epistemic target emerges due to the opacity of the system: the complexity of the technology is both the reason for its use in the first place (since it mimics the complexity of the real system) and for its confusion with the real system. Nordmann's observations have much in common with Dowling's black boxes, but whereas Dowling allows the re-opening of the system for explanatory purposes, Nordmann insists that:

In the technoscientific ‘animistic’ employment of computer simulations, animal models, algorithmic structures, and other substitutive uses of dynamic systems, those systems can be controlled without an understanding of their construction. This is evidenced quite simply by the fact that their scientific users do not actually make the requisite hard and software (and for the most part would not know how to make it).

The collapse of distance comes from the complexity of the system(s), and it cannot be reconstructed since the operators of the systems are ignorant of their construction. Instead, "technoscience acquires a feeling for the system's behaviour, probes its sensitivities to parameter

---

632 Nordmann here makes space for philosophers, requiring them to remind technoscientists of the fictional aspects of their objects, but, with due apology to the enterprise of philosophy of science, it is not clear how this problem is not resolved within technoscientific practice itself. If one brings external technoscience into the mix – returning the focus to the industrial aspects of knowledge generation – the act of moving the technoscientific object into the realm of its practical application necessarily recalls the representative capacities of the object. For example, in the case of the onco-mouse, while the mouse itself may function as a self-sufficient reality while it is being cured, relating that information to the cure of human tumours recalls the representation, and reconstructs the distance through physical dissimilarities. As with Dowling's black boxes, the collapse of distance is temporary.
variations, and the like, then transfers this knowledge as best one can from the isolated model systems to the integrated ‘real’ systems.”

There is certainly evidence, within this thesis alone, that Nordmann’s collapse of distance occurs when using simulation, especially when visual elements come into play. I think Nordmann is wrong, however, in seeing this collapse as irrevocable. Firstly, it seems that often the users of the simulation do construct the requisite software. The modern astrophysicist is a simulationist, a coder, perhaps even a computer builder. Some of the outcomes of the simulation are surprising, the result of complex pathways a human cannot follow – such is the knowledge generative capacity of simulation. Yet just as it is opaque, simulation can also become very transparent, familiarity bred from the act of construction (albeit not usually of the material components) as well as the act of use. Perhaps this is particular to astrophysics, as there is less of a division of labour between the construction of the (immaterial parts of) the simulation technology, and its use – astrophysicists do not tend to accept black-boxed simulation modules so readily. Yet it could also be argued that as technoscientific work destroys the distance between object and representation, simulation continually reconstructs the difference.

Part of the problem might lie in the fact that the interaction between simulation and empirical systems like experiment is under explored. Bensaude-Vincent et al, for example, seem to understand the ‘back-and-forth process with experiment’ as the only adjustment and calibration of the theory within the simulation for the purposes of developing a better model. Undoubtedly this is a very important interaction, but misses the use of simulation for data analysis, and furthermore obscures the importance of the intersection between simulation, theory, and empirical data. Annamaria Carusi, Kevin Burrage, and Blanca Rodríguez perform a closer analysis of the simulation-experiment setup in cardiac electrophysiology, which investigates the electrical activity of the heart. Exploring electrophysiological phenomena, Carusi et al demonstrate, requires the construction of model-simulation-experiment (MSE) system, in which all three aspects of the system are embedded within each other. This chapter has likewise emphasised the interdependence of model construction and simulation. There is a dialectic between simulation and experiment; as the simulation is constructed, so too is the experimental target domain, the specific aspects of cardiac electrophysiology under investigation. As the MSE system is used, all parts of the system are adjusted subject to knowledge from other parts of the system:

The need for using different techniques to obtain information at different levels comes from the need to address the limitations of each of the experimental data sets used in the definition of the MSE system at each level of integration.

634 Sundberg, 2010a.
635 The term ‘empirical systems’ refers to scientific setups that produce data of the real world. It is only in this narrow sense that observation and experiment are treated as similar in this section – both produce real data sets that are compared to simulation data.
636 Carusi et al., 2012.
637 Carusi et al., 2012, p. H128.
Simulation was likewise used in the missing satellites problem to address certain limitations of the (observational) data sets. As the electrophysiology MSE system demonstrates, the attempts to work around such limitations may come from anywhere within the complex model-simulation-experiment setup, the mashed together acronym attempting to express the embeddedness of these components. Such a process of negotiation around limitations is furthermore temporal – the background knowledge on which the initial MSE is constructed has also emerged from the negotiations of previous MSE systems. It is, in fact, an iterative process:

It is more helpful to think of the relationship between the MSE system and its target domain as an ongoing dynamic process of iteration and adjustment. The dynamics of the iterative process are driven by advances in both experimental and computational techniques as well as investigations on their combined use in cardiac electrophysiology research. ... As new experimental data and techniques become available, inconsistencies between computational model predictions and experiments result in refinement and improved knowledge of the corresponding MSE system.\textsuperscript{638}

Exactly the same process has been identified in the case of the missing satellites, where advances in both simulation and observation technology contributed to improved knowledge both of the target domain and of the system itself. This sort of embeddedness between knowledge and the enactment of knowledge seems to be quintessentially technoscientific, though Carusi et al do not phrase their work in those terms. The simulation may construct a self-contained world in technoscientific terms, but it is difficult to understand the construction of both the simulation, its model, and even the interpretation of the results of the empirical system as isolated. In the case of the missing satellites problem, observational data was corrected by simulation, simulation was constructed in conjunction with this data (including the simulation of the very instruments of observation), and understanding the dark satellites as plausible could not have arisen without the close interplay of the two arenas of evidence.

Furthermore, it is this very embeddedness that prevents confusing representation and reality. Without the SDSS data (and for the previous chapter, without the rotation curves), the dark satellites (halos) would have remained speculative. Some empirical system is necessary, otherwise simulation has nothing with which to negotiate, and so progress our knowledge. More broadly, we can understand this as being a virtue of simulation's non-materiality. A simulation's objects are non-material, and are constructed, even if that construction is obscure. It is precisely because the black box of simulation can be opened that enables the artful (re)construction of the immediacy: in virtue of simulations very immateriality can we know its building blocks (even if we cannot from this understand the system's operations as a whole) and so recover "agreement in terms of a coordination that does not require physical similarity or a likeness between the representation and what it represents".\textsuperscript{639} That is, comparison between a simulation and a material system recovers the theoretical nature of the relations between the representation and

\textsuperscript{638} Carusi et al., 2012, p. H151.
\textsuperscript{639} Nordmann, 2006, pp. 10-11.
its target, and so the distance between the two. The STM simulation is compared to the STM experiment – immateriality to materiality.

Of course in the case of astrophysics – and other disciplines that don't have material systems as obvious as experiment – things are slightly less convenient. The simulation produces almost-real data and images when it acts as a contained system, and I have argued that through simulation the theoretical model can be understood as a legitimate epistemic target which produces, without inverted commas, data. This black-boxing is temporary, and the act of comparison with an empirical system recalls the fictional nature of the simulation – very strongly in the case of the STM system, as the comparison is between material and non-material. And it is through simulation's flexibility – in looking like theory, or experiment, or observation – that this situation can arise in the first place, giving us the ability to first black-box a system, and then open it up. Yet observation does not provide the same material recalcitrance as experiment does, even though it is necessary for empirical grounding. Hence the ease of overlap between simulated and observational data, and hence the negotiated existence criterion of plausibility found in the missing satellites problem case study.

The interplay between simulation, theory, and an empirical system like observation or experiment are complex and deserve more attention that can be given here. For the purposes of this case study, it is sufficient to restrict ourselves to the observation that are at least three interactions that are relevant for knowledge generation in cases like missing satellites problem. The first is the interaction (i) between empirical data and theory: the classical narrative of 'science', important since it is often the discourse reconstructed in scientific work. The portrayal of the missing satellites problem as a discrepancy between theoretical prediction and observation, and the subsequent division between astrophysical and cosmological solutions, comes from this approach. To be 'technoscience', this interaction should be embedded within the apparatus that produces the data – the observation instrumentation. In astrophysics, though theory is necessarily constrained by empirical data, there seems to be little interaction with the activities of observing and theory crafting. Typically, one does not see astrophysics as a technoscience because of this lack of obvious interaction. It seems, however, that the interaction between data and theory should also be embedded within the apparatus that produces the theory, or rather, that draws the empirical content out of the theory. This is because this latter apparatus – simulation – is also a technology. Through simulation technology, the theory does interact with observation technology, and vice versa, as the dialectic around the luminosity curve shows.

There are therefore (at least) two more interactions that are important: (ii) between simulation and some empirical system – experiment for cardiac electrophysiology and nanotechnology, observation for astrophysics – and (iii) between simulation and theory (or 'model'). This is really more of a tri-way intersection: between simulation and observation is the technoscientific

\[640 \text{In the sense discussed so far: the construction of simulation necessitates the co-construction of the theoretical model, which is only manifested and used through simulation though it may be expressed external to it.}\]
collapse of distance, as simulation turns theory into a technology. Through both the luminosity curve and the visualisations of the hydrodynamic simulations, theoretical data and empirical data are melded with each other. In interacting with theory in this manner, simulation constructs the model as an epistemic target, but in being virtual also recalls the fictional nature of the model when compared with and integrated within, the results of the empirical system.

4.7.2 Fear of Seduction (Again)

Nordmann's collapse of distance doesn't wander into the realm of seductiveness if only because with material systems one is not dealing with concerns about representation and reality per se. Though the previous section argued that comparison with observation recalls the fictional nature of the model, the virtual nature of simulation still leads to worries that scientists might be seduced, in spite of common sense and the real world. Like in previous chapters, in the case of the missing satellites these worries hang in the background, this time through the guise of some sort of technological determinism. High resolution simulations were a necessary condition for the resolution of the missing satellites problem. Hydrodynamic simulations seem to have cemented what a combination of high resolution dark matter simulations and semi-analytic models was already strongly suggesting. Other recent 'problems' in astrophysics are suggestive of a similar technologically reliant conclusion. The overmerging problem, for example, was a direct result of poor resolution in simulation – it was in fact solved by increased computing power and code efficiency. The missing satellites problem only became a problem because of better simulations, and from a certain perspective (like that implied by Font et al's position), it was also solved through the use of more complex simulations. In a separate but related problem, it had also proved difficult to obtain realistic disc galaxies alongside ellipticals – until hydrodynamic simulations like Illustris.

The good distribution of morphologies that these new simulations exhibited showed that “previous futile attempts to achieve [an accurate model] were not due to an inherent flaw of the $\Lambda$CDM paradigm, but rather due to limitations of numerical algorithms and physical modelling.”

The authors behind EAGLE, however, seem oddly resistant to this interpretation:

> The improvement in the resolution afforded by increases in computing power and code efficiency has also been important, but perhaps mostly because higher resolution has helped to make the implemented feedback more efficient by reducing spurious, numerical radiative losses. Improvements in the numerical techniques to solve the hydrodynamics have also been made ... and may even be critical for particular applications ..., but overall their effect appears to be small compared to reasonable variations in subgrid models for feedback processes.

---

641 Vogelsberger et al., 2014a, pp. 177-178.
642 Schaye et al., 2015, p. 522.
While improvement in technology is clearly incredibly important, the authors of EAGLE felt it was actually better theoretical models for 'subgrid' processes that were behind the increase in the realism of galactic models. Essentially this assertion seeks to deny technological determinism about the recent progress in astrophysics, (rightly) pointing out the large improvements in theoretical understanding (Illustris does, to give them their due, also see progress in 'physical modelling'). Of course, as this chapter has argued, it is a combination of theoretical and technical progress that has solved problems like that of the missing satellites. The EAGLE authors are likely acutely aware of the importance of technological improvement; in fact, there seems to be a lot of evidence that astrophysicists in general are far from blind to the importance of technology in their predominantly theoretical discipline. It is this awareness that leads the EAGLE physicists to urge their audience not to forget the importance of theory crafting. Attributing the bulk of the progress in problem solving to improvements in computational technology leads to a less critical stance on the theory, as one assumes the reason behind certain problems is not a faulty understanding, but the lack of power or resolution in the simulation. One can understand these concerns, as though there were worries that the overmerging problem might be caused by a bad assumption, the argument that there was simply insufficient resolution in the simulations won. In the missing satellites case, too, it is the existing, conservative theory (ΛCDM) that emerges victorious.

Yet achieving this solution was contingent on having better simulations – you may have better subgrid models, but you cannot actually model them (in the context of the galactic processes) without higher resolution simulations. In fact the authors of EAGLE are less concerned with technological advances being given too much credit for knowledge progression than they are for what this implies about the power of simulation. EAGLE dedicated a section of their paper to the ‘implications of the critical role of subgrid models for feedback’ that essentially warned against the uncritical interpretation of the results of simulations that employ these advanced subgrid models: “the risk of misinterpretation is real, because it can be shown that many simulations underestimate the effectiveness of feedback due to excessive radiative losses ..., which themselves are caused by a lack of resolution and insufficiently realistic modelling”.643 While more powerful than ever before, hydrodynamic simulations still do not achieve the ideal resolution across a large dynamic range and volume. When combined with a better but still comparatively poor understanding of the subgrid processes, this means we must remain aware that “the ab initio predictive power of the simulations is currently limited when it comes to the properties of galaxies.”644

On one hand, better simulations with higher resolutions and more processing power enable us to create more realistic models. On the other hand, EAGLE insists, it is not the simulation that grants us this epistemologically beguiling realism, but the use of better subgrid models. Furthermore, even if we use simulation to instantiate these ‘more realistic’ models, the simulation technology still has limitations that have non-negligible effects on the theory itself. Despite the fact that it seems like higher resolution simulations have solved all our problems, EAGLE cautions that we

643 Schaye et al., 2015, p. 523.
644 Schaye et al., 2015, p. 523.
must not assume that they will do so in all or future cases. The same worries about the uncritical use of simulation as a technique were also seen when simulation was beginning to be used widely in the 1970s and 1980s, as Chapter 2 details, and again in Chapter 3 (section 3.9). The fear of seduction arises again in this chapter because simulation has an even greater ability to create a substitute real, incorporating ever more complexities until it resembles the noisy targets of science, and then producing data that can be integrated seemingly seamlessly with the real thing.

One might in fact suggest that those like Nordmann who see visualisation at the heart of the collapse of distance, have been themselves seduced. By this I mean that it is no doubt the case that the visualisations of simulations are incredibly persuasive, and potentially seductive and misleading. However, it is not in the images (or at least, not only in the images) that simulation and empirical data blend – it is in the marrying of datas. The luminosity curve is a very good example of this: as an image, the ‘observed’ curve tells us that there are dark satellites. However this image is built from data – numbers – that have been observed, processed by simulation, or simply straight up simulated. It may be that visualisation seduces, but the invisibility of numbers suggests that if we have something to fear – and fear is useful, even if it is never realised – it in the silent alliance of real and virtual datas. Sundberg describes how climate scientists also blur simulated and observational data with what is referred to as ‘re-analysis’ data. In much the same way that the luminosity curve was completed by Tollerud et al, re-analysis returns a spatially and temporally coherent dataset using a combination of observation and theory – “re-analysis and data assimilation is commonly referred to as a ‘synthesis’ of models and measurements.” Once again, simulation causes scientists to tread a fine line:

On the one hand, the heavy use of observations for comparative purposes could be referred to as a maintenance of the ‘reality principle’ … On the other hand, production of re-analysis indicates the precession of the model – simulacra – because it involves explicit and overt adaption of the observations to the models.

Perhaps it is because of the cautionary words that such fears of seduction are largely unrealised in practice: while best fit models and compatibility with observations has always been important to astrophysicists, they do not portray their simulations as true simulacra, or even overly faithful (mimetic) representations. The conclusions of simulation papers, like most scientific articles, are full of caveats about simplifying assumptions and incomplete models.

645 Seeing this brief regress to earlier themes of simulation literature also suggests that the hydrodynamic simulations are a further instance of sustaining technology, in the same sense that supersimulations in the early 2000s were (see Chapter 2). This might explain the entirely gratuitous rhetoric at the beginning of the EAGLE paper, which tells us “simulations enable astronomers to ’turn the knobs’ much as experimental physicists are able to in the laboratory” (Schaye et al., 2015, p. 521). Where supersimulations were sold in terms of their increased power, it might be that hydrodynamic simulations are sold in terms of their increased reality, with several EAGLE related papers emphasising the improvement in modelling capabilities of hydrodynamic over dark matter only simulations, allowing more detailed and more accurate comparisons with observation.

646 Sundberg, 2010a, p. 279.

647 Sundberg, 2010a, p. 279.
Nordmann writes: “where models used to mediate between theory and reality and thus highlighted the distance between theory and reality, theory now offers algorithmic building blocks to construct a substitute reality that is hardly less complex than its original.” While simulation does have the capacity to construct realities with seductive complexity, there is no indication within the case studies of this thesis that astrophysicists have actually been seduced. Furthermore, the interaction between experiment technology and simulation as technology of theory can be understood as playing the same role as models in ‘mediating’ between theory and reality. Through the aim of bringing the two systems of the experiment and in simulation into sufficient likeness (one of the major ways in which simulation is used to theory craft), their very differences are highlighted. Even, then, when scientists use an instrument that they themselves have not constructed, and whose workings are therefore opaque to them, the act of matching instrumental results with simulated results constructs an immediacy that is ‘artful’ through the conscious recognition one is not like the other. The same can be said for an area like astrophysics that must substitute observation for experiment (in terms of empiricism).

Of course at the same time the complexity simulation effects enables theory to produce ‘observations’ and results in its own right, as a (fictional) epistemic target. Without the interaction between simulation and observation technology being defined by materiality, it becomes much harder to restore the difference between model and modelled. Whereas in interaction with experiment it will always be apparent which system is actual, observational technology cannot provide the same level of material recalcitrance for the astrophysicist. In both astrophysical case studies, observational data is mostly confronted with theory through the medium of simulation, which simulates not only the system and the observer, but also the theory. The overlap between the use of simulation as a technology of theory and as simulated observation also blurs the line between the objects created in simulation-theory and those ‘seen’ in simulation-observation. Thus the dark satellites of the simulated universe find their expression in the luminosity curve of the observed one. The difference is not entirely effaced – as EAGLE reminds it readers, the simulation models are still representations – but this overlap does have the effect of increasing the value of ‘plausibility’ as a condition for existence. As our data is negotiated, so are our epistemic categories.

4.8 Conclusion
In 1999, simulations revealed there was a severe discrepancy between the amount of substructure predicted in the Local Group, and the number of satellite galaxies actually observed. Suggestions for resolving this discrepancy were either cosmological – changing the CDM model in some way – or astrophysical – using existing physical mechanisms that acted on a galactic scale. With the release of the SDSS data and the discovery of around twenty new ultra-faint satellites,

648 Nordmann, 2011, p. 28.
astrophysical solutions began to gain the ascendancy.\textsuperscript{649} By modern day, it seems that a consensus has been reached that a combination of astrophysical mechanisms acts to suppress star formation in most but not all the simulated dark halos. Those halos that are left luminous by the altered model match the luminosity function derived from the SDSS data quite well. The missing satellites problem has been ‘solved’, leaving astrophysicists with a much improved understanding of galaxy formation and a growing conviction of the importance of baryonic processes. This outcome was achieved through a convoluted process of negotiation between simulation ‘evidence’ and observational evidence, both limited and allowed for by improvements in instrumental and computational technologies throughout the episode. The negotiation between simulation and observation occurred with the technology in order to progress theoretical understanding, and to answer a problem emerging from the theoretical model. Along the way, understandings of what counts as suitable evidence have also been negotiated, as prescriptions of plausibility are given greater weight in circumstances where observation has no voice.

It should be apparent that this negotiation process is very pragmatic. In order to solve problems, astrophysicists seek ways to negotiate around the technical limits of simulation and observation, around gaps in theory and data alike. In order to obtain progress, however, something needs to be compromised. The missing satellites were never completely found – SDSS unearthed only 20 more faint dwarf galaxies; this is not even enough to satisfy the order of magnitude found in the 1999 simulations! Yet the problem has been ‘solved’ through a negotiation around not only technical limits, but what constitutes an adequate solution. It is sufficient that dark satellites are plausible, and that there exists a plausible combination of mechanisms (that actually do exist) to keep the small halos at low luminosity. Astrophysicists not only negotiated a solution to the missing satellites problem, but they also negotiated what counted as sufficient evidence. Somehow the satellites have been reclassified from ‘missing’ to ‘dark’, without ever being found. This may be an outcome of the power imbalance in the technical negotiation between simulation and observation – simulation, on a purely computing basis, progresses at a pace at least comparable to Moore’s Law, while observation technology is comparatively slow, new data trickling in from smaller instruments while large surveys like SDSS only provide a flood of information every decade or so. It’s unsurprising that simulation results can be made to work as stopgap in terms of justification for a theory. Astrophysicists then appeal to ‘future observations’ to give a solid observational grounding, though in the case of the dark satellites it seems that these appeals, while no doubt sincere, are somewhat empty. The norm of modern astrophysics may

\textsuperscript{649} How and why precisely this happened has not been explored in this chapter. On the face of it, there is no clear-cut empirical reason why cosmological models may not still turn out to provide the overall solution – as there are papers claiming the solution of the missing satellites problem within the CDM paradigm, there are also modern solutions involving decaying dark matter (e.g. Wang et al., 2014), or combination cold and warm dark matter (CWDM) models (e.g. Macciò, Ruchayskiy, Boyarsky, & Munoz-Cuartas, 2013). The choice of astrophysical solution may be attributed to the many successes of the CDM, and the disinclination to change the model without strong proof to the contrary. It may also be too early to be making such strong statements, as the results from large hydrodynamic simulations have only just started being published at the time of writing.
well be to operate in this negotiating framework, suspended in the potential of the future but working pragmatically with what is available.

The activities of simulation as a technology of theory, in a dialectic with observation technology, and mediating between observational data and its theoretical representation, are embedded within each and should not even be understood as particularly distinct. What is important here from the technoscientific perspective is that the simulation cannot be removed from the knowledge it produces, perhaps not even from the data it helps correct. It is not, however, a case of out with the old and in with the new: this chapter has been less concerned with labelling astrophysics as a technoscience than with using the technoscientific framework to draw out aspects of modern scientific practice, and how these relate to more traditional constructions of science. In particular, it helps that astrophysics is outside of the material, industrial, and practical comfort zone that technoscience typically finds a home in.

Just as the experimental turn in history and philosophy of science revealed an over-emphasis on theory, the discussions in technoscience act as a ‘continuation’ of this reconfiguration by recalling the importance of technology. Benefiting from these frameworks is both the historical perspective – as scholars reconfigure history of science to include the artisans as much as the savants\(^{650}\) – and the philosophical analysis – no longer can we safely consider theories and data as disembedded from their conditions of production. Temporarily pulling technoscience away from experiment highlights rather than obscures these methodological benefits, particularly the latter. In focusing on astrophysics and the use of simulation as a technology of theory we are not reminded of the importance of theory – which would be somewhat ridiculous, as it recalls itself easily enough in historical, philosophical, and scientific narratives – but rather we are reminded that a good part of science is concerned with explanatory goals, and that these goals are furthered in new ways as new technologies appear.

\(^{650}\) Klein, 2005b.
5 Generating Knowledge with Simulation

5.1 Introduction
The central questions of this thesis have been epistemological: how do simulations generate knowledge? How do they generate new knowledge? How are simulations (re)presented as knowledge-generating tools by scientists, and what role does this play in legitimating simulation knowledge? The preceding case studies have attempted to shed light on these questions, but also made it apparent that there is unlikely to be one definitive answer that applies to all simulations in all disciplines. This thesis has of necessity been selective and has not focused on: simulating an experiment or instrument before building it; simulating an actual experiment (with the brief exception of section 4.7.1); fictional simulations not intended to be representative (but which nevertheless can be epistemologically generative); what Morgan calls semi-material experiments, or simulations that use material inputs; simulations of one physical process by another physical process (which may involve computer simulations); the use of computers to process large amounts of empirical data ('big data'); and many other simulations used for different aims and working with different theoretical structures than those in astrophysics. This returns to the point made in the initial philosophical analysis (section 1.5) that it would be highly difficult to speak of an 'epistemology of simulation' and manage to incorporate the wide variety of ways in which simulation is used.

The fact that simulation is a plastic and highly adaptable technique has been a major factor in simulation's ubiquitous acceptance and application. Part of this flexibility, and perhaps the reason simulation has found its place in so many varying disciplines, is the methodological ambiguity that makes the technique both interesting and confounding for philosophers. Looking at the *Simulation* journal revealed that having a foot in both the experimental and theoretical modelling camps lent simulation enough stability to be accepted as a valid technique for generating knowledge. Simulation used techniques of validation and verification from the modelling tradition, and from experiment borrowed a methodology centred around manipulation of the system. In the bar instability case study, it was by flipping between experiment and theory that simulation was able to ‘confirm’ and ‘test’ predictions in both a purely theoretical environment and in conversation with observational evidence. Simulation adopted the role of experiment when it was useful to distinguish between observer and observed, and between technique and epistemic target, and this had a meaningful effect on some limited cases of epistemological justification (like Toomre’s criterion). The case of the missing satellites took this one step further,

---

651 See Winsberg, 2010, chapter 5; Contessa, 2010; Wise, 2011.
652 Morgan, 2003. Morgan’s example of a semi-material experiment is where a bone is scanned, a model of the bone constructed in the computer, and then that is used to construct an idealised situation that ‘tests’ the bone. This is what Morgan calls ‘virtually an experiment’, as opposed to full simulations of the sort considered in this thesis, which are ‘virtual experiments’.
653 This is not to say that one cannot identify commonalities across the disciplines – after all, they are all using something called ‘simulation’ – but the point is that this is a highly complex task, and not one to which this thesis aspires.
showing how simulation effected both the construction of a world and the construction of the representation of a (our) world. In constructing an experiment-like arena, simulation allowed theories (or models) to produce empirical content that could be robustly compared with observational data. Through the previous chapters, it has been demonstrated that the dual methodology that simulation showcases, at once *sui generis* and aggregate, is highly fruitful for producing new knowledge where, from a classical standpoint, there is only old knowledge, or none at all. The first part of this chapter (section 5.2) will revisit these themes and describe the main outcomes of the thesis.

Section 5.3 of this chapter draws attention to the role that simulation images played in the presentation of simulation results across the three chapters. This section also pays particular attention to the graphs produced by simulation for comparison with observation (rotation curves) or in conjunction with observation (luminosity curves). I will then briefly return to the issue of philosophical novelty (section 5.4) in order to suggest some general conclusions about the practice of simulation in science. The first part of this section points out that simulation is particularly well-suited to the historical style of reasoning, and the second part – in looking at how researchers have interpreted simulation as a ‘culture’ – highlights some common aspects we see in modern science and simulation, particularly simulation-dominated disciplines.

A short excursus (section 5.5) discusses the treatment of simulation as virtual and disembedded versus an understanding that makes simulation material and contextualised – this is of importance both from a historiographical standpoint (for this thesis and HPS simulation literature in general) and an epistemological one. In sympathy with the general tone of the rest of the thesis, this penultimate section argues that the virtual is being introduced into science in a very epistemologically real way. Finally, the conclusion to this chapter (section 5.6) provides an answer to the motivating question of the thesis: how can the virtual teach us about the real?

## 5.2 The Empirical Content of Theory

Simulation’s place on the methodological map is a point of contention in the HPS literature, and even understanding simulation as somehow both experiment and theory raises more questions than it answers. Part of the aim of the preceding chapters has been to clarify where simulation might stand in relation to the more traditional forms of knowledge generation, and furthermore how such positioning affects simulation’s ability to generate knowledge. It does seem that simulation allows us to ‘experiment on theory’, in the sense that with simulation we can construct worlds and objects out of theory, and manipulate and intervene with those targets. In doing so we can test various hypotheses and compare results with observation – as astrophysicists did when curve-fitting the rotation curves of dark halos to that measured by telescopes. Yet simulation’s role is not restricted to acting as a ‘bridge’ or ‘mediator’ between theory and data. Simulation adopts and transforms our existing ways of doing science.

One of the interesting observations that has been made along the way is that while simulation is not experiment *per se*, it does seem to be one way through which we can learn about the activity
of modern experimentation. A very different beast to its more traditional predecessor, modern experiment can be very large, very costly, and very ambiguous; and it also often involves, in one role or another, simulation. Rather than using experiment to attempt to explain simulation, it may be that simulation can provide us with new understandings of experiment. The bar instability case showed that the experimental language used by the scientists to describe simulation was not mere rhetoric, but had a definite effect on how scientific knowledge was generated and justified. Other examples are the scanning tunnelling microscope simulation (section 4.7), where the simulation is simultaneous with the experiment; and Morrison’s case study of the Higgs boson, which she concludes by suggesting that “the use of simulation at the LHC [Large Hadron Collider] casts doubt on the very distinction between experiment and simulation; the latter is simply an integral part of the former.”

Merz makes very similar observations about the interplay of particle physics experiments and simulations; her observations “raise doubt about the possibility of discussing simulation as a distinct and decoupled epistemic activity.” When compared with simulation, experiment tends to be broadly portrayed as an intervention with a material target, but when simulation is included in an experiment, this portrayal is certainly an oversimplification. As Morrison’s work suggests, it may be that the key feature of experiment is not, after all, its materiality, but the ability to co-construct scientific representation and object. Though beyond the scope of this work, such interactions lend weight to understanding simulation objects as legitimate epistemic targets (or things), this time embedded in the wider networks of experimental systems.

Putting this potential path of research aside, it is one of the main outcomes of the preceding chapters that the question of whether simulation is ‘experiment’ or is ‘theory’ is almost moot in the face of the more interesting observation that it is the representation of simulation as one or the other that takes precedence in text and in practice. A variety of narratives are used in the case studies to place simulation in one or more traditional roles – simulation can ‘confirm’ or ‘test’ theoretical predictions, but it is also involved in making those predictions. Both case studies emphasised that in moving between the roles of theory and experiment, simulation displayed an epistemological flexibility that was crucial to the production and legitimisation of knowledge. Not only does such flexibility enable simulation to marry the empirical content of the theory with real data, but it also connects simulation to existing forms of knowledge generation in looking like, as simulation does, both theory and experiment.

An important facet, then, of this thesis’ contribution to the experiment-theory discussion, is that the practice of simulation shows the continuity and maintenance of existing ways of making, confirming, and presenting knowledge. The language used in the scientific literature still borrows from classic ideas of the role of experiment and theory, even though it is often simulation that is playing both roles. In addition, the case studies also showed that simulation takes these existing ways of doing and reconfigures and intertwines them; a prime example is the blended empiricism of the luminosity curve. Simulation merges the empirical content of the theory together with the

654 Morrison, 2015, p. 316.
655 Merz, 2006, p. 168.
empirical content of the observational data, resulting in the re-construction of our understandings of what constitutes sufficient evidence. Theorising becomes theory crafting, involving an iterative and dynamic dialectic between available evidence and multiple simulation runs. We understand simulating as model construction, drawing together multiple sources of knowledge into a coherent if complex whole. Experiment is interpreted – through simulation – in terms of its rhetoric, roles, and manipulability. Observation is both mediated and simulated. Simulation remains connected to but also acts to transform traditional epistemic categories.

One of these reconfigured categories, this thesis has suggested, may well be ‘empirical’. As other HPS scholars have argued, simulation is at least epistemologically on par with experiment (see section 1.2). In this thesis, that argument has taken on the form of understanding simulation as a technique through which theory can become an epistemic target, and a model can have an empirical content – not in the limiting way that Winsberg uses the term, nor in the mediated way that Morrison describes, but in the more robust sense that simulation data is deemed sufficiently empirical that it can be integrated with actual observational and experimental data.

The astrophysical case studies described how this superposition of the real and the virtual occurs in practice. In order for the rigorous comparison of data to be possible, simulation first has to build a self-contained virtual world. Hayles’ concepts of the Platonic backhand and forehand can assist us in describing how this happens:

The Platonic backhand works by inferring from the world’s noisy multiplicity a simplified abstraction. So far so good: this is what theorizing should do. The problem comes when the move circles around to constitute the abstraction as the originary form from which the world’s multiplicity derives. Then complexity appears as a ‘fuzzing up’ of an essential reality rather than as a manifestation of the world’s holistic nature.

Often abstraction is quite necessary to understand complex phenomena, by removing unnecessary details that obscure the mechanism of interest (which will naturally change as the question of interest changes). Of course, the Platonic aspect comes in when the messy details are entirely discarded in favour of simplicity or elegance. The second move Hayles identifies – the Platonic forehand – will be more familiar to readers of this thesis as it requires the assistance of powerful computers. The Platonic forehand “starts from simplified abstractions and, using simulation techniques such as genetic algorithms, evolves a multiplicity sufficiently complex that it can be seen as a world of its own.” The enactment of the Platonic forehand can be seen very

---

658 Batterman, 2009. This trick, Hayles points out, has been around since the Greeks, though Batterman describes idealisation to reveal essential features as ‘non-traditional’ modelling. According to Batterman, traditional modelling requires the model to conform more closely to reality – which are the aims of the very modern astrophysicists of Chapter 4.
strongly in the missing satellites problem (especially section 4.5.1). It is when the backhand and the forehand work together that a virtual, contained world can be created, as Hayles describes:

The two moves thus make their play in opposite directions. The backhand goes from noisy multiplicity to reductive simplicity, whereas the forehand swings from simplicity to multiplicity. They share a common ideology – privileging the abstract as the Real and downplaying the importance of material instantiation. When they work together, they lay the groundwork for a new variation on an ancient game, in which disembodied information becomes the ultimate Platonic Form.\(^{660}\)

In a way, the outcome of the missing satellites problem can be described as the working together of the Platonic backhand and forehand. First we describe the Universe in simplified terms and ideas, reduced so as to be mathematically workable and humanly understandable. Though simplified, these models capture the relevant aspects of the phenomena. Then we apply simulation, to turn simplicity into complexity. Hydrodynamic simulations are powerful enough to construct a sufficiently complex world – one that can be ‘experimented’ upon and with – making universe-like complexity out of the CDM model, several semi-isolated subgrid models, and some initial conditions.

An interesting point to make here is that simulation is necessary to construct this complexity. Modern astrophysics can rarely boast ‘a model’ or ‘a theory’ of a target or phenomenon; instead explanation is achieved through the coming together of many models. Galaxy formation, for example, is governed by the CDM model, but also requires the addition of several subgrid models to produce a sufficient likeness of the Universe. Introducing dark halos into the galactic model required that work be done variously on black holes, the persistence of spiral structure, disc stability, and dark matter. In addition, astrophysics is known for its liberal implementation of other areas of physics like optics, hydrodynamics, and quantum mechanics; chemistry also plays a large role. Simulation is a technique very well suited to facilitate this coming together of disparate models into a coherent whole, especially when the information available may not even qualify as a full model or theory. This may take the form of post-hoc rules, like Toomre’s criterion, which emerge as a result of simulating the system. Guillemot, in the context of climate science, likewise notes the capacity of simulation to “integrate heterogeneous theories, milieus and elements, to span scales of time, space and complexity, to combine the singular and the contingent”.\(^{661}\) These fragmentary pieces of knowledge are cohered through the medium of simulation into a more understandable whole. Thus simulation allows for both fragmentation and construction – for both simplification and complexity. In constructing systems complex enough to be ‘realistic’ – like the hydrodynamic simulations that model both baryonic and dark matter – simulation makes theory a legitimate epistemic target, open to intervention, manipulation, and measuring. At the same time, the constitutive parts are also flexible enough to enable the theory crafting and negotiation with observation seen in the case studies.


What both scientists and philosophers fear, however, is taking this sort of superposition to the extreme, losing the ability to distinguish between real and virtual and therefore generating unsubstantiated and speculative theories and objects.\footnote{It seems that of these sorts of worries may be a characteristic of modern science, of which simulation is more of a symptom than a cause. Similar fears and debates around the legitimacy of new ways of doing science (especially physics) can be found in: discussions around the multiverse (Ellis, 2014), and string theory (Ritson & Camilleri, 2015), in cosmology (Pearce, 2015, especially cf. chapters 3 and 6), and most certainly in dark matter studies. Of course, these debates can also be connected with past debates in the history of science but we can perhaps allow that in simulation they at least take on a new form.} The *Simulation* journal (and complementary sociological studies like Turkle’s) showed the onset of these fears in the 1970s and 1980s. Dark matter studies are in some sense the exemplar of this problem; the crucial question of dark matter’s existence – of its reality – is one that can only be answered through a medium of virtuality – of simulation. The persuasive pictures of early spiral galaxy simulations, and the amazing realism of the hydrodynamic simulations, were again accompanied by warnings against seduction. An important factor in this, however, is that the fear of seduction, while stubbornly persistent and ever-present, is never actually realised. The previous chapters have emphasised that while simulation objects can be legitimate epistemic targets in their own right, observational evidence, however meagre, played a pivotal role. In the instance of galactic halos, this was mainly the rotation curves; for the missing satellites, it was the SDSS data. As we saw by comparing Toomre’s criterion with Ostriker and Peeble’s conjecture, simulation can only ‘confirm’ predictions when it has enough empirical force to do so.

To put it another way, the data that simulation produces is only legitimate because it maintains its virtuality. The comparison of the real system with the virtual is what both teaches us about the system – through the processes of superposition and negotiation – and what recalls the fictional nature of the model – through the act of comparison. This is when we move out of the understanding of simulation as an isolated system, in which the theory constructs an epistemic target, and move into the understanding of a simulation of, that is, of learning about a certain simuland. At the same time, though observations and empirical data act as necessary constraints, simulations are needed to fill in the gaps and ambiguities – both forms of evidence (simulation and observation) are required to answer theoretical questions and solve problems.\footnote{Cf. Guillemot, 2010, p. 244.} Through this duality, virtual objects are given the same signification as the real scientific objects, but not the authority of reality.\footnote{Cf. Cameron, 2007, who discusses the comparison of real and virtual objects in archaeological simulations.} Simulation can therefore only make claims to plausibility, not existence.

By marrying scientific techniques in this manner, simulation knowledge necessary takes on a more chimeric form. Though the hydrodynamic simulations did not reify the dark missing satellites per se, the superposition of the simulated observations of the simulated galaxy cluster over real observations of the real Local Group resulted in the acknowledgement that the dark satellites were highly plausible. Plausibility became a robust way of understanding the existence of an object – the satellites, or the halos (indeed, dark matter in general) are not confirmed by direct observation, and yet they have a compelling presence in the realistic simulated world, and
thus in scientific explanation. The two case studies particularly have worked to demonstrate how such fragments of virtuality in our empiricism have resolved problems or assisted in the development of new theories and explanations. The fear of seduction acts to limit the reification of simulation objects – the mere possibility that we might fall into the trap of speculative science helps to prevent it. At the same time, the growing legitimacy of the simulation objects as epistemic targets works to convince us of their existence, with the appropriate empirical caveats. Plausibility as a criterion for existence treads this line between construction and reality. It is a negotiated criterion based on balancing the fear of seduction with the increasing use of virtual tools to learn about a real world.

As Turkle observes, “twenty years ago, professionals in science and design flirted with simulation even as they were suspicious of it. Today, they are wary but wed to it.” That both the fear and the resistance co-exist is reflective of the actual use of simulation in scientific practice. Simulation has become a necessary tool, and ‘direct’ or unambiguous empirical evidence is no longer abundant, particularly for many areas of modern physics. In order to progress knowledge, we must find new ways of producing it. Hence it is not about maintaining some clear line between theory and empiricism. The very use of simulation in science tells us it is not a case of real versus virtual, either in practice or in epistemics, but about the extension of our epistemological capabilities using the virtual. Simulation, as a technology of theory, allows theoretical representations to produce both data and observations – with the observer and occasionally the observer’s instrument placed in the simulated system itself. Simulation extends both theory and experiment: the marriage of existing scientific traditions and techniques with the powerful potential of virtual spaces.

5.3 Visualisation and Simulated Observations

One of the continuing background themes that threads through the case studies is the use of computer images and films as an important part of both conducting simulation and presenting simulation results. Though included in the analysis of the case studies, the visualisations of simulation have not been given greater focus because studying their use in astrophysics alone could probably support an entire other thesis, backdropped by the expansive visual studies literature both within and without HPS. In the context of this thesis, the use of simulation images is placed in three limited contexts.

The first is rhetoric, seen most strongly in the bar instability case study, where still and moving images of simulated galaxies exhibiting spiral structure were used to help validate both the simulations and their results. This visual rhetoric is certainly not a tactic peculiar to simulation or computer images, but this thesis has attempted highlight how the construction of an image using simulation can function as part of the argument both for the scientist – through ‘virtual

---

666 Hayles makes this point with respect to the man-and-machine connection (Hayles, 1999, pp. 290-291).
witnessing' – and for the presentation and acceptance of results. Several of these images, representing the combination of observational evidence and simulation results, played a key role in the case studies. For the dark halos, this was the rotation curve, particularly (if one might recall) the simulation-generated graph that decomposed the rotation curve into the contributions from the various components of the galactic model (Figure 13). The simulated rotation curve could be compared to the measured curve, and the similarities observed, but the decomposed rotation curve also made a powerful argument that each component of the galactic model – including the halo – was necessary to construct the right curve. The image also gave us information about why all these components were necessary; it was clear that the dense core was required for the peak in the centre (where the halo had little to contribute), and the dark halo was needed for keeping the curve flat past the luminous edge of the disc. For the missing satellites the example we have is, of course, the corrected luminosity curve (particularly Tollerud et al's version, in Figure 16) that implied the existence of dark satellites, and which was the focus and discussion of a good portion of that chapter. Both the simulated rotation curve and the bolstered luminosity curve act as persuasive pieces of evidence in favour of the plausible existence of dark objects.

Tied to their use in persuasive rhetoric is the simulation image’s ability to seduce. This has been the second way in which simulation images were cast in the case studies. The authors of the spiral galaxy images cautioned against taking the comparisons between photographs and simulated images too seriously, pointing out that they were mere stills and the spiral patterns (in the simulation) were not sufficiently long-lived to be physically real. The images produced by the modern hydrodynamic simulations are perhaps the strongest example of potential seduction, suggesting the loss of the fictional nature of the model, and the loss of our ability to tell real from false – the simulated Hubble image is nigh on identical to the real one (Figure 19). The waters are further muddied when one recalls that the Hubble photograph itself would been the subject of interpretation and processing.668 As even a simple comparison of the images of the first case study with those of the second might indicate, this increase in the potential for seduction is a property of the increasingly life-like nature of computer images; indeed, this is one area in which HPS researchers often give simulation its earlier and more sinister meaning of deception.

This leads us into the third way in which simulation images have been understood in this thesis, and that is as constructed. To an extent, most scientific images have some element of construction to them – photographs are cleaned up or coloured, graphs have obviously erroneous data removed and scales adjusted, and sketches are themselves the product of the author's interpretation.669 Through the process of trained judgment, as Daston and Galison relate, the raw image is transformed into something that is useful when interpreted in the context in which both the image maker and the audience are embedded. More to the point, however objectively this process might or might not be performed, the resultant image is still a representation; “the prefix

668 Hentschel, 2014, pp. 372-373. This also plays on the idea of the line between the scientific image and art, which Daston and Galison speculate may be swinging back around to become synthesised once again (Daston & Galison, 2007, pp. 399-412).
re- is essential: images that strive for representation present again what already is. Representative images may purify, perfect, and smooth to get at being, at ‘what is.’ But they may not create out of whole cloth, crossing over from nature into art.”

Simulated images, on the other hand, do not begin with a representation, but create one – that is, though simulation images are often intended to represent in one way or another, their creation comes about digitally, and their origins are in theoretical data. The simulated observations provided by the hydrodynamic simulations of the previous chapter (figures 17, 19 and 20) are a very powerful example of this, because not only were they constructed images of simulated galaxies (rather than retouched photographs of real galaxies, or interpreted graphs partially based on observational data), but they were simulated observations, taken by a simulated telescope pointed at a simulated universe. As the previous sections have argued that the data taken from a complex and sufficient simulation can constitute legitimate empirical data, so too perhaps can ‘photographs’ of that system be considered legitimate scientific observation (though naturally, the same caveats apply).

Though the visual aspect of simulation was underexplored in the previous chapters and must unfortunately remain so here, it does suggest some compelling avenues of further research. Particularly, we might ask that if the ‘empirical’ content of the simulation is legitimate only because it can later be compared or synergised with real data, what status do the simulated ‘observations’ have? Are they simply presented for rhetoric or aesthetic? The interplay between processed or recoloured images and simulation images, particularly with the former used as ‘empirical evidence’, also raises questions, as does the overlap between image and data (both real and simulated). Of interest in the latter case is how some of the images – particularly graphs and other visualisations of numbers – come about as the product of the synthesis of observation and simulated data. Daston and Galison mention an atlas compiled from the SDSS data, which “explicitly embraced the combination of interpretative and algorithmic procedures” in its construction. The luminosity curve of the missing satellites took this one step further in becoming a constructed image comprised of the ‘completed’ SDSS data – completed through simulation – and the simulated data. In this thesis, simulation images encouraged immersion (Chapter 2), virtual witnessing (Chapter 3), and substitute observation (Chapter 4); they also stood always in the background of fears of seduction. But it might be suggested that if one wants to worry about seduction, the danger lies much more in the data than it does in the images. Simulation can produce images that look and act like observations of a system, but they are usually used for the sake of comparison or rhetoric, to illustrate how alike the simulated and real worlds look. Simulation data, on the other hand, is often used for the interpretation or completion of observational data, or, as in the case of the missing satellites, combined with

671 Though, bearing in mind the observations by scholars about the importance of the aesthetic in scientific images (see for example Hentschel, 2014, chapter 12), ‘simply’ may be the wrong word to use.
672 Daston & Galison, 2007, p. 390. Now named the NASA-Sloan Atlas (http://www.nsatlas.org, retrieved March 30, 2016), the catalogue has been expanded to include data from several other surveys.
empirical evidence. With data, the line between virtual and real has a higher potential to be unseen.

5.4 Returning to the Issue of Philosophical Novelty
One of the greatest attractions for studying simulations in science is that their use suggests something philosophically novel is going on – this was discussed earlier in the thesis (section 1.5), where it was suggested that simulation looks different depending on the discipline. Another facet of the novelty of simulation is that the resistance of the technique to classification coupled with its ubiquity across multiple disciplines, might suggest that something has shifted in the way contemporary scientists do science, compared to their less computerised predecessors. This thesis has suggested, modestly, that simulation work is distinct, though it has close association with other forms of knowledge generation. Undoubtedly simulation and computers in general have drastically changed science (as they have changed our society) – Ann Johnson points out, in the spirit of Nicholas Jardine, that computers have changed the very questions we ask and can ask. However, since no analysis has been made here of what astrophysics looked like before simulation, we cannot satisfactorily answer any questions about what simulation has changed. Nevertheless, it seems possible to make some suggestions about what simulation science might look like from a broad perspective.

5.4.1 Styles of Reasoning
Sam Schweber and Matthias Wächter have suggested, for example, that simulations might constitute a new Hacking-type style of reasoning. According to them, the simulation style of reasoning makes a movement from a reductionist quantum mechanical conception, describing everything in terms of fundamental laws, to a constructionist approach that involves building the world from those foundational theories. Astrophysical simulations do indeed emphasise the construction of the phenomena in question, and this construction (including the construction of the model) is a key factor in developing an explanation or theory. In addition, as the analysis of the Simulation papers might have suggested to those familiar with Hacking’s work on styles, the establishment of simulation might have something in common with the manner in which styles proper become autonomous, and then once autonomous, become self-authenticating.

673 Johnson, 2006b, pp. 25-27.
675 See Hacking, 1992, though there is a wide selection of literature discussing ‘styles of reasoning’ in particular (an example is the special issue on styles of reasoning in Studies in History and Philosophy of Science, 2012, Vol. 43), and even more on ‘styles’ in general. Personally, I would hesitate to describe simulation as a ‘style’ – it seems to me simulation is more of a technique that is used to employ several styles, mostly notably (in astrophysics at least) those of the statistical and historical styles. If there is a new ‘style’ emerging in modern physics it may be encouraged by or particularly well suited to simulation. However, given that among its novelty, simulation also demonstrates continuity with previous traditions, ways of thinking, and styles of reasoning, the question of ‘novelty’ is a thorny one, as is the question of what exactly one means by ‘style’.
it is beyond the scope of this thesis to go into this matter, it could be suggested that simulation encourages the use of certain styles, particularly the historical style, in which the history of an object is used to explain its features.

At the end of the bar instability case study, it was suggested that the galactic model acquired a historical element in terms of its explanation – the bar instability appears during the course of the disc's evolution, and is dissipated by the growth of the central dense core; it therefore forms part of the galaxy's history and is not simply an artefactual phenomenon. Of course, the entire story of the bar instability involved models of galaxies that evolved in time and thus showed or didn't show the formation of a bar instability. Almost all the other simulations discussed in this thesis are evolutions of systems or objects, with the ability to go backward and forward in time being one of simulation's chief attractions to scientists. Wise, looking at the use of partial differential equations (PDEs) in physics, argues that without such narrative elements, we cannot relate our mathematics or models to the real world. Though typically associated with deductive narratives, "the simulation generates an evolutionary explanatory narrative to replace the deductive narratives traditionally associated with PDE's".676

It is in fact very difficult to think of an example of a static simulation in astrophysics. One simulation, exploring the histories and present-day properties of the Milky Way's subhalos, found that individual halos can have drastically different histories: some of the subhalos may have lost over 99% of their mass, some none at all; some subhalos may have had their peak velocity reduced, or it may have gone entirely unchanged. The simulation also found that not only are subhalo and host halo evolution linked, but due to the fact that the subhalos move around a lot, they can affect the formation of systems that are not, at present day, within the host halo's virial radius: "in other words, the assembly history of CDM halos must depend on their environment."677 Hence, knowing only the present day properties of a halo "is not sufficient to infer its accretion history or halo occupation distribution in a statistically correct way ... suggesting that simulations should be employed whenever structure formation needs to be followed accurately."678

There is evidence that not only is historical explanation widely employed in astrophysics, but that it is one of the best ways in which we can understand astrophysical objects. Wise suggests "explanations of the behaviour of complex systems may always require a turn to historical narratives".679 The modern model of a galaxy serves to 'explain' a galaxy, but it does so more through historical explanation than a description of physical rules that govern a galaxy's behaviour. It is not the static picture of a stable spiral galaxy that serves as an explanation, but the inference to the dynamic simulation that depicts the persistence of non-barred spiral arms over several rotations, and that furthermore explains these spirals (and other components) in terms of the formation of the galaxy itself. Beyond structure formation and galaxy studies, Jacob

676 Wise, 2011, p. 357.
677 Diemand, Kuhlen, & Madau, 2007b, p. 873.
678 Diemand et al., 2007b, p. 874.
Pearce has demonstrated how the (itself evolving) historical style of explanation has become central to cosmological inquiry as a whole. It seems that as a technique simulation is very well suited to and may even encourage historical explanations; simulation’s introduction into cosmology and astrophysics and the effect it has had on their modes of explanation suggests an interesting avenue of further research.

5.4.2 Cultures
Perhaps it is more useful to this thesis not to suggest in the dangerous manner of grand narratives that simulation constitutes a revolution, Hacking-style or otherwise, but rather to understand simulation as part of a shift in perspective, towards a different sort of culture. Johnson and Lenhard see simulation as encouraging a culture of prediction, “a reorientation in the practices of scientists and engineers” that has amplified the character and role of prediction in science. They argue that while mathematical models aim at transparent explanation, the opacity of computational models suggest a refocussing towards performance (in terms of predictive capacity), a sentiment which seems to have something in common with some technoscientific writing. The opacity of computer models encourages experimentation and exploratory work, but at the same time “epistemic opacity and the lack of ontological referents emerge from the layering of models and the distance from the real world.” Though this may often cause individual models to have a limited life span, Johnson and Lenhard see the speed, ease, and versatility of computational modelling as ample compensation in this new mode of operating:

This amounts to a change in scientific culture: scientific research proceeds by experimenting on computational models and assessing visual outputs, replacing efforts to write the one computational model that will yield all the right results. The new strategy is exploratory, evolutionary, adaptive, and provisional.

Certainly the conclusions of this thesis have suggested a similar outcome. However, that such a culture of exploration, evolution, and adaptation (or perhaps, negotiation) necessarily places an emphasis on prediction does not seem to be an obvious conclusion. Johnson and Lenhard appear to see predictive power as the pay-off for undertaking provisional modelling, which they portray as “a culture that – openly or implicitly – accepts missing sustainability”. While this conclusion seems valid for some areas where simulation is employed (namely climate science, where prediction is centrally important for not only scientific but also political reasons – and where models have limited life spans), astrophysics uses simulation’s exploratory mode differently.

Sundberg, in undertaking a sociological study of different groups using simulation, is in a position to compare astrophysicists and meteorologists. She describes two cultures that both disciplines

---

680 Pearce, 2015.
682 Johnson & Lenhard, 2011, p. 194.
participate in: the culture of simulation is characterised as postmodern, fluid, decentred, and opaque, in which the search for mechanisms and depth is futile, and in which visualisations play a very important role. Sundberg here means ‘simulation’ in the sense of ‘simulacrum’ – that is, more connected to the postmodern idea of simulation as a substitute real than computer simulation per se, though in science the latter plays an important role in the former. Sundberg contrasts this with the culture of calculation, which is modernist, focusing on linearity, logic, and depth. Importantly, Sundberg does not intend to polarise these two cultures or pit them against each other – she notes that both astrophysics and meteorology have elements of both the culture of simulation and that of calculation. Rather, her aim is to explore how the two outlooks exist side-by-side and manifest in different simulation activities and situations. For example, the black box use of codes – a trend towards postmodern superficiality – is more accepted in meteorology than it is in astrophysics, where the latter prefers to know the sorts of physics and algorithms that goes into a simulation, or to build the simulation themselves. On the topic of visualisations, Sundberg’s astrophysicist actors – as we have likewise seen in this thesis – use visualisations to great effect while also cautioning against an uncritical interpretation. Hence:

Astrophysics presentations appear more ‘modern’ than meteorological counterparts regarding attempts to unpack simulations and codes into the underlying equations. At the same time, the use of and discourse on animations signal inclination towards ‘postmodernity’ in astrophysics, but not necessarily seduction.685

The ‘culture of simulation’ that astrophysics participates in is not overarching. The previous sections of this chapter have underscored this same aspect of simulation, how “modern as well as postmodern features are exhibited in the situated character of numerical simulation practices.”686 As astrophysicists use simulations as simulacrum, they also present the results within classical narratives. Modelling, comparing results with observation, and crafting theories are all continuous with existing ways of doing science, though they may be enhanced and reconfigured by simulation. We can also see, through Sundberg’s work and comparing this thesis’ own conclusions to that of Johnson and Lenhard’s, how it can be dangerous to talk of a single epistemology of simulation. At the same time, it is this very plurality of practice that gives simulation its potency as a technique, opening up new ways of doing science as well as adapting to existing ways of doing science.

‘Culture’ seems like a better way to describe simulation in science than ‘style of reasoning’ if only because ‘culture’ is a broader and more encompassing concept. Regardless of the label, several authors, and this thesis, have identified aspects of practice that doing science with simulation seems to encourage: an exploratory (or ‘playful’, tinkering) mode; the coming together of ad hoc elements into a fragmentary whole; adaptiveness or negotiation between the evidence and the model (importantly, a dialectic); an emphasis on visualisation; a greater capacity for the

685 Sundberg, 2010a, p. 278.
686 Sundberg, 2010a, p. 280.
evolutionary or historical method of explanation; and an opacity that while exploited is not permanent, blending with the more traditional culture of calculation.

5.5 Excursus: The (De)contextualisation of Simulation

As we saw in Chapter 1, a good deal of philosophical literature takes for granted that simulation is theoretical practice. In fact, simulation is often differentiated from experiment on the grounds that simulation is a virtual, abstract process (whereas experiment is material practice). This thesis has encouraged this interpretation – the simulated galaxy is virtual, astrophysicists manipulate virtual objects, and the results of a simulation form the empirical content of theory. Despite sometimes looking like experiment, simulation misses that crucial immediacy with the target system that we associate (perhaps falsely) with experiment – this is more a methodological point than an epistemic one. It takes philosophical work to argue for simulation having an important material presence, whether via the computer or something else.

But simulation is, in an obvious way, a material practice that is contextualised and embedded in the conditions of its production. As was discussed in both case studies, the computer places very real limitations on the theoretical models that simulations implement. In the Simulation journal, the physicality of the computer was an important aspect of early simulation studies; the activity of simulating – moving around components, typing, reading off a graphical display – was crucial for developing rapport. The ordinary back-and-forth, messy, and non-linear scientific process doesn’t change just because the system of investigation is virtual. Scientists have geographical and institutional limitations, not only in the sorts of computers they have access to, but in the expertise that is available to them. There is the anthropology of the thing: many scientists use codes developed by other people, and large groups of scientists are required for large simulations, which are elements of scientific practice that are directly relevant to epistemic questions involving justification and background assumptions. The solutions for the bar instability and missing satellites problems were contingent, the result of the localised conditions of production under which simulations operated. It was, for example, Jeremiah Ostriker's previous work on Maclaurin spheroids that suggested to him that the halo might be a solution to the bar instability.

This is essentially a historical point; it seems self-evident given that we are working off the back of the many studies on localised knowledge production done by HPS scholars. Yet despite acknowledging this, in presentation and representation simulation is inarguably virtual. Any contextuality is overridden by the preoccupation with the abstract simulation model, virtual computer images, or results in the form of disembodied data. The fear of seduction hinges on this. The process of simulating on a physical computer is elided in the presentation of simulation results, where the specifics of the computer or the facility that houses it are rarely referenced anymore, particularly for the more ‘normal’ simulations (those not run with supercomputers). Instead, as the case studies have shown, a variety of narratives is used in which simulation is placed in reconstructed roles – but just as those narratives are necessary to place simulation among the repertoire of scientific techniques, they also act to disambiguate simulation from its
conditions of production. Merz, in looking at the use of simulation alongside material experiment in particle physics, highlights this problem:

The account of the importance and role of simulation throughout different experimental phases may seduce the reader into believing that simulation studies might fully substitute for material exploration until the moment that the ‘real’ collider and detectors are constructed. The account presents the fiction of a (materially) disembedded practice. The simulation laboratory seems to be completely decoupled from the material world of ‘traditional’ experimentation.  

In Merz’s study, simulation is an integral part of the experimental process – which, more than any other form of knowledge production in this thesis has been repeatedly interpreted as material, particularly when compared to simulation. As a part of the particle physics experimental system, Merz’s simulations are embedded within the same localised practices as that of the material laboratories, yet they are presented as if they are not.

The same narrative structures have been seen in this thesis. In Chapter 2, it was noted that after the digital versus analogue debates in *Simulation*, the physicality of the computer faded into the background. The two other main themes identified in the journal supported the idea of simulation as virtual practice: the theoretical, abstract model became the foremost theme from the 1970s onwards, and the man-and-machine theme showed that relating simulation to experiment relied on manipulation rather than materiality. In the halo story, narrative structures placed simulation in the roles of both theory and experiment – this methodological flexibility is not possible if simulation is presented as contextualised and localised. The missing satellites chapter talked about technologies, and underscored that in order to uncover a solution to the missing satellites problem, simulation data became blended with observational data. In order for this to happen, the two datas needed to be of the same type – the luminosity curve could not be produced by combining real observations and theoretical predictions. It could, however, be produced by negotiating between the virtual data that comes from a simulated observation, and observational data that is interpreted through simulation. It might in fact be suggested that the solutions to the case studies were contingent on simulation being understood as a virtual, decontextualised process, because the flexibility that interpretation entails is crucial to generating knowledge using simulation.

In a sense the choice to suspend the historical or localised context of simulating was deliberate, given that the focus of this thesis was on the use of simulation as a technique for generating ‘data-like’ knowledge rather than understanding it as an enacted activity. The choice of research matter also had an effect, as published scientific papers likewise tend to ignore the actual practice of making science – realising this is what triggered the experimental turn in philosophy of science. Unlike experiment, however, simulation’s context is far more difficult to recall, due in part to what Hayles calls a condition of virtuality. The disembedding of simulation knowledge seems to be inherited from a wider trend to do the same to computational data in general. Virtuality, as Hayles

---

(in the late 1990s) describes, “implies we participate in the cultural perception that information and materiality are conceptually distinct and that information is in some sense more essential, more important, and more fundamental than materiality.”

It is by virtue of its virtuality that simulation can perform the Platonic backhand and forehand, and create substitute worlds that produce legitimate data.

However the affectation of virtuality is just that – Hayles emphasises that portraying information as distinct from materiality is a cultural perception, a result of the “contemporary pressure toward dematerialisation”. Likewise in media studies, where digital interaction and material objects collide frequently, researchers have come to a similar conclusion: the concern with materiality is seen as a reaction to the myth of the immaterial, rather than an actual dematerialisation.

This presents a methodological challenge to HPS scholars looking at simulation. It is generally the role of historians of science (and I might suggest, of central importance to philosophers of science) to restore the contextuality of knowledge production, but it is not always so easy. We may have in our favour the sociological and experimental turns, but Hayles observes that “although researchers in the physical and human sciences acknowledged the importance of materiality in different ways, they nevertheless collaborated in creating the postmodern ideology that the body’s materiality is secondary to the logical or semiotic structures it encodes.” Dark matter studies in particular twists together the virtual simulation object and the material scientific object, because the evidence that convinces us dark structures exist is interpreted through simulation, and it is through simulation that dark matter is embedded in the theoretical framework of modern astrophysics.

The point I’m trying to make here has several layers. The first is that simulation (simulating) is a contextualised and embedded process. The second is that simulations (and simulation results) are presented as if they were wholly virtual, and their conditions of production are hidden; this we know from work done by critical theorists, STS researchers, and their sister disciplines. Both of these points have been apparent, albeit not central, in this thesis. The third point, which I feel has been underemphasised in the HPS simulation literature, is that even if simulation’s decontextualisation is false, the fact that it occurs and it has effects on the scientific knowledge produced using simulation means it must be given due attention – as Hayles and others might try to restore context, we should also pay attention to simulation within the condition of virtuality. It is in fact a requirement of the superposition of datas – and thus their blending – that the results of the simulation can be disembedded, even as temporary fiction.

Furthermore, as this thesis has also tried to emphasise, it is in the interplay from immaterial to material – theory to experiment, virtual to real, abstract to target – that the epistemology lies. It
is not that simulation is wholly virtual, it is that simulation's contextuality is engaged in a continual back-and-forth with the myth of its decontextualisation, and that that virtuality – even false – bleeds over into the knowledge and objects generated with simulation. Merz points out that "to adequately conceptualize simulation one has to address not only the disembedded nature of simulation but also the *dynamic interplay of the disembedding and reembedding* moves that determine how and to which epistemic means simulation is employed."\(^{692}\) The turn to practice in HPS occurs simultaneously with the increasing use of computers in science; this thesis has suggested that we take both the material culture of science, and the immaterial culture of computer simulations together, in order to enhance each other.\(^{693}\) The crossing of the real and the virtual became a theme in this thesis, with the emphasis on the superposition of real and virtual data, the blending of observation and theory. If we intend to accept the simulation technique as producing legitimate scientific knowledge, especially in simulation-dominated sciences like astrophysics, we must find space for the virtual in our epistemologies.

### 5.6 Conclusion

By blending empiricism and virtuality in new ways, simulation extends our capacity for knowledge generation. Through the case studies, this thesis has sought to closely illustrate simulation in practice, and so demonstrate the complexities of producing knowledge despite shaky empirical grounds, and the resultant deep intertwining of simulation and other forms of knowledge generation. The answer, then, to the question of 'how can the virtual teach us about the real?' is: through a superposition of simulated data over observational data. This is an expansion on an idea that Turkle describes using the metaphor of Plato's cave, paraphrasing one of her participants:

> When he [the scientist] places simulation alongside the real, it is to throw the real into sharper relief; simulation's errors sharpen his view of where the real resides. But, like the inhabitants of Plato's cave, [the scientist] knows reality through the shadows it casts.\(^{694}\)

Perhaps we might say that the scientist *only* knows reality through the shadows it casts. Comparison between simulation and observational data is not a surprising method – and certainly not a new one – but what this thesis has emphasised are the cases where simple comparison of theory and evidence is not particularly useful. For astrophysics, and particularly for dark matter studies, the evidence is often ambiguous and scarce. The case studies have shown that simulation has made epistemological progress possible by expanding the capacity for comparison. In order to compare whole worlds, to superpose one upon the other so as to see or fill in the gaps in our knowledge, simulation needs to be powerful enough, and positioned appropriately, to construct, first, an isolated and sufficient world. Within this positioning and these simulated worlds, the virtual epistemic target made of theory and models can be (in a

---

\(^{693}\) See also Pias, 2011 and Cartwright, 2015.  
\(^{694}\) Turkle, 2009, p. 81.
limited manner) a legitimate source of empirical data. Importantly, however, this isolation is temporary; perhaps more accurately, it is a false but deliberate reconstruction. For any comparison or superposition to be possible, and in order to learn about the actual targets of our investigations (and to prevent complete seduction), the recalcitrance of observational or experimental data, however singular or ambiguous, is required.

In pulling together many theories and hypotheses, simulation turns fragmentation into coherence, developing not an ad-hoc science, but a post-hoc one, where theory crafting and negotiation with the empirical evidence is the normal mode of scientific practice. Through the superposition of simulated and real, knowledge emerges – a stable galactic model that maintains spiral structure and has the right rotation curve; a model of structure formation that includes baryonic processes and the creation of dark substructure. Hence, and but, this superposition necessarily blurs demarcations. ‘Empirical evidence’, it has been suggested, might also emerge from theory, in conjunction with observation. Since the solutions to the problems posed by the case studies are neither purely empirical nor purely fictional, plausibility has been shown to be a legitimate criterion for existence. Through seeking to understand the world we reconstruct it, and simulations are a powerful tool with which we may increase the potential space of scientific objects into the realm of the virtual.
References


Electronic Associates Inc. (1966). We didn’t want a computer that could do everything. We wanted system a scientist could do everything with. *Simulation*, 6(6), i-ii.


McKenna, N. (1972). We're professionals-- so why keep it a secret? *Simulation, 18*(6), vii-viii.


Ören, T. I. (1977). Simulation - as it has been, is, and should be. Simulation, 29(5), 182-183.


Palme, J. (1977). Moving pictures show simulation to user. Simulation, 29(6), 204-209.


Minerva Access is the Institutional Repository of The University of Melbourne

Author/s: Wilson, Katia Stephanie

Title: Astrophysics in simulacrum: the epistemological role of computer simulations in dark matter studies

Date: 2016

Persistent Link: http://hdl.handle.net/11343/124159

File Description: Redacted thesis - Astrophysics in simulacrum: the epistemological role of computer simulations in dark matter studies

Terms and Conditions: Terms and Conditions: Copyright in works deposited in Minerva Access is retained by the copyright owner. The work may not be altered without permission from the copyright owner. Readers may only download, print and save electronic copies of whole works for their own personal non-commercial use. Any use that exceeds these limits requires permission from the copyright owner. Attribution is essential when quoting or paraphrasing from these works.