MODELS IN SCIENTIFIC PRACTICE

by

Yoichi Ishida

B.A. in Philosophy, University of Nevada, Reno, 2005

M.A. in Philosophy, University of Nevada, Reno, 2007

M.A. in History and Philosophy of Science,

University of Pittsburgh, 2011

Submitted to the Graduate Faculty of

the Kenneth P. Dietrich School of Arts and Sciences in partial

fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2014

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Yoichi Ishida

It was defended on

July 31, 2014

and approved by

Sandra D. Mitchell, Professor, History and Philosophy of Science

James G. Lennox, Professor, History and Philosophy of Science

Kenneth F. Schaffner, Distinguished University Professor, History and Philosophy of Science

P. Kyle Stanford, Professor, Logic and Philosophy of Science, University of California, Irvine

Dissertation Advisors: Sandra D. Mitchell, Professor, History and Philosophy of Science,

James G. Lennox, Professor, History and Philosophy of Science

MODELS IN SCIENTIFIC PRACTICE

Yoichi Ishida, PhD

University of Pittsburgh, 2014

This dissertation presents an account of the practice of modeling in science in which scientists' perceptual and bodily interactions with external representations take center stage. I argue that modeling is primarily a practice of constructing, manipulating, and analyzing external representations in service of cognitive and epistemic aims of research, and show that this account better captures important aspects of the practice of modeling than accounts currently popular in philosophy of science.

Philosophical accounts of the practice of modeling classify models according to the categories of abstract and concrete entities developed in metaphysics. I argue that this type of account obscures the practice of modeling. In particular, using the analysis of the Lotka-Volterra model as an example, I argue that understanding mathematical models as abstract entities—non-spatiotemporally located, imperceptible entities—obscures the fact that the analysis of the Lotka-Volterra model relies primarily on visual perception of external representations, especially hand- or computer-generated graphs. Instead, I suggest that we apply the concepts of internal and external representations, developed in cognitive science, to models, including mathematical models.

I then present two case studies that illustrate different aspects of modeling, understood as a practice of constructing, manipulating, and analyzing external representations. First, using Sewall Wright's long-term research on isolation by distance, I articulate the relationship between the uses of a model, the particular aims of research, and the criteria of success relevant to a given use of the model. I argue that uses of the same model can shift over the course of scientists' research in response to shifts in aim and that criteria of success for one use of a model can be different from those for another use of the same model. Second, I argue that in successful scientific research, a

scientist uses a model according to the methodological principles of realism and instrumentalism despite the tension that they create among the scientist's uses of the model over time. This thesis is supported by a detailed analysis of successful scientific research done by Seymour Benzer in the 1950s and 60s.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

This dissertation is a result of different lines of research on scientific modeling that I have engaged in for the last several years. For their invaluable feedback at every stage of this project, I thank my co-directors, Sandy Mitchell and Jim Lennox. I'm also grateful to my committee members, Ken Schaffner and Kyle Stanford for their feedback. I especially thank Kyle for hosting me during my research visit to Department of Logic and Philosophy of Science at University of California, Irvine in the summer of 2012 and engaging in numerous discussions on many ideas presented in this dissertation. I also thank Kyle for introducing me to Howard Stein's paper in 2009 and having discussed instrumentalism with me ever since.

I'm deeply grateful to Elizabeth O'Neill and Joe McCaffrey for having a number of discussions on virtually every topic in this dissertation. Elizabeth read most parts of the dissertation in various stages and provided important comments and criticisms. Joe and I did a reading group on Edwin Hutchins' *Cognition in the Wild*, and we had many discussions about distributed cognition and the practice of modeling.

Many people read my work or discussed ideas with me at various occasions. I thank Marshall Abrams, Marina Baldissera Pacchetti, Nora Boyd, Dick Burian, Thomas Cunningham, Peter Gildenhuys, Leah Henderson, Taku Iwatsuki, Ryota Morimoto, Nancy Nersessian, Tom Nickles, Aaron Novick, Bob Olby, Laura Perini, Collin Rice, Alirio Rosales, Senji Tanaka, and Derek Turner.

Michael Grabe kindly invited me to work with him on computational modeling of pH regulation in lysosomes. Doing actual scientific modeling helped me stay close to the everyday practice of science while I developed my general views about models and modeling. I am grateful to Michael for his generosity and mentorship, and to Smita Nayak and Joe Mindell for their collaboration.

I presented parts of Chapter 3 at the 2009 and 2011 meetings of the International Society for the Historical, Philosophical, and Social Studies of Biology and part of Chapter 4 at the 2013 meeting. I wish to thank the audience for discussions.

Charles Greifenstein of the American Philosophical Society Library helped my research on Sewall Wright Papers. Loma Karklins, Archivist at the Caltech Archives, helped my research on Seymour Benzer Papers, and Charlotte Erwin, the head of the Caltech Archives, granted me access to the Benzer papers, which were still being processed at the time of my visit. I thank all these people for their expert help.

I thank the American Philosophical Society for permission to use materials from the Sewall Wright Papers and the Caltech Archives for permission to use materials from the Seymour Benzer Papers. I also thank Cambridge University Press, University of Chicago Press, Scientific American, the Genetics Society of America, and the American Society of Naturalists for permission to reproduce materials under copyright.

Chapter 3 is based on research supported by the Wesley C. Salmon Fund, University of Pittsburgh, and Chapter 4 by a Doctoral Dissertation Research Improvement Grant from National Science Foundation (NSF #1230201). I'm grateful for the financial assistance from these institutions.

In the past year, I worked as a Program Assistant at the Center for Philosophy of Science at the University of Pittsburgh and did the final stage of writing there. I want to thank the Center stuff, John Norton, Karen Kovalchick, Joyce McDonald, and Cheryl Greer for making the Center a great place to work. I'm also grateful to the HPS Department stuff, Rita Levine, Joann McIntyre, and Natalie Schweninger, for providing necessary administrative support.

This dissertation is dedicated to my mother, who trusted and supported my decision to leave my home country and pursue philosophy.

1.0 PERSPECTIVES ON MODELS

1.1 INTRODUCTION

Nelson Goodman once observed that "[f]ew terms are used in popular and scientific discourse more promiscuously than 'model"' (Goodman 1976, 171). His observation still holds true: Recent historical works on models suggest that the meaning of the term 'model' in scientific discourse significantly changed over the course of the last three centuries (Griesemer 2004, 437–439; de Chadarevian and Hopwood 2004). This fluidity of the meaning of the term 'model' raises a basic question that any scholar wanting to study the practice of modeling in science needs to confront: What is a model?

This chapter explains some of the most influential answers to this question. Two of them underly the formal and informal versions of the semantic view of theories, while others are responses to the semantic view. I shall then argue that most of these answers employ concepts of abstract and concrete entities developed in metaphysics. These answers represent what I call the *metaphysics based* perspective on models and modeling. I then outline an alternative perspective that adopts concepts developed in scientific studies of representations, especially studies in cognitive science. Finally, I contrast this perspective with the perspective characteristic of those who regard models as agents playing various roles in science.

1.2 THE SEMANTIC VIEW OF THEORIES

Since the 1950s, models in science have gained increasing and renewed attention in philosophy of science (Bailer-Jones 1999). One of the impetuses for this development was the semantic view of theories. Although the semantic view was originally concerned with formal, mathematical analysis of the structure of scientific theories, some philosophers have developed a more informal version of the semantic view (Odenbaugh 2008, 510). The semantic view has provided the background for some influential accounts of the practice of modeling today. Here I summarize this background in order to highlight how my own account of modeling differs from the accounts based on the semantic view.

1.2.1 A Formal Version

Applying Alfred Tarski's (1953, 11) concept of model in mathematical logic, Patrick Suppes has argued that a scientific model is a set-theoretical (i.e., mathematical) entity in which all theorems of a theory are satisfied (Suppes 1960, 289–290). This idea underlies the set theoretic version of the semantic view of theories (Suppes 1957, Ch. 12; 1960; 1967; da Costa and French 1990; French and Ladyman 1999). The other formal version of the semantic view is known as the state space approach (van Fraassen 1970, 1972; Lloyd 1984). Differences among the formal versions of the semantic view are irrelevant to my present purpose, and here I use the state space view as an example.

In presenting his state space approach, Bas van Fraassen states the shared commitment of both the set theoretic and state space approaches:

> Like Suppes, I shall take it that (the 'pure' part of) a theory defines the kind of system to which it applies; empirical assertions would take the form that a given empirical system belongs to such a kind (or, more precisely, that one of the mathematical structures specified by the theory provides an adequate model for the empirical system). (van Fraassen 1972, 311)

Applying the model theoretic concept of model, the system to which a theory applies is called a *model* of the theory. According to van Fraassen's state space approach, a physical theory defines a system (i.e., its model) by specifying two things. First, it specifies the set of states that the system can take. To do this formally, van Fraassen says, "what we specify is a collection of mathematical entities (numbers, vectors, functions) to be used to represent these states" (van Fraassen 1972, 311). That is, theoreticians use numbers, vectors, functions, and so on to specify the states of the system.

The collection of these "mathematical entities" is called the state space of the system (van Fraassen 1972, 311). For example, a theory may define a system of gas by specifying its state in terms of its temperature t, volume v, and pressure p. The collection of triples of real numbers $< t, v, p >$ can be used to represent the states of this system, and this collection of real numbers is the state space of this system (van Fraassen 1972, 311). Second, according to van Fraassen, a physical theory defines physical magnitudes (e.g., kinetic energy), which are represented by functions defined on the state space (van Fraassen 1972, 311–312). Once its models are defined in this way, a physical theory may also provide laws of succession that describe the ways in which its model can change its state over time.

To the extent that it characterizes what models are and how they are defined, the formal version of the semantic view suggests a particular way to understand the practice of modeling. On this approach, to construct a model (of a theory) is to define its state space and physical magnitudes. Scientists do this by using mathematical objects, such as numbers, vectors, and functions. Thus, on the semantic view, mathematics is used to define or specify a model (van Fraassen 1970, 337).

1.2.2 An Informal Version

An informal version of the semantic view largely shares the basic commitment of the semantic view that we saw above in a passage from van Fraassen. For example, Ronald Giere, who is a major defender of an informal version of the semantic view, says:

> My preferred suggestion [as an account of theory structure], then, is that we understand a theory as comprising two elements: (1) a population of models, and (2) various hypotheses linking those models with systems in the real world. (Giere 1988, 85)

This statement is nearly identical to van Fraassen's, but unlike the formal version of the semantic view, Giere does not understand "models" strictly in model theoretic terms. Rather, he takes models to be abstract entities that have properties ascribed by scientists' descriptions of it (Giere 1988, 78).

Giere has developed his informal version of the semantic view into a general account of the practice of modeling. He understands the modeling practice hierarchically as shown in Figure 1.1. Let us focus just on the top half of Figure 1.1. Giere thinks that mature scientific theories contain

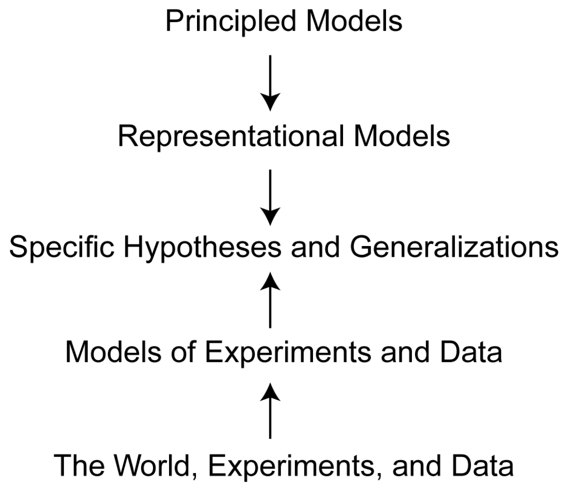


Figure 1.1: Giere's account of hierarchical relationships among models and the world. The arrows indicate the generational relationships. For example, representational models are generated from principled models, which are characterized by statements. Redrawn from Giere (2010, 270).

principles or laws, which are statements or equations, which describe models. To use his example, Newton's laws of motion are principles of classical mechanics that characterize "a class of highly abstract models (principled models)" (Giere 2010, 270). Principles are "automatically true of the principled models" (Giere 2010, 270). According to Giere, scientists generate representational models by "adding conditions and constraints to the principled models" (Giere 2010, 270). Thus, by adding Newton's gravitational law to the principled models, scientists generate representational models of the interactions of two bodies in space. In characterizing representational models, Giere insists that they are abstract objects not to be identified with any particular description of them (Giere 2004, 747, n.7; Giere 2010, 273) Representational models, he says, are "abstract objects constructed in conformity with appropriate general principles and specific conditions. One might think of them as artful specifications of the very abstract models defined by the principles" (Giere 2004, 747).

For Giere, representational models are representational because they are what scientists can use to represent parts of the world. In his view, scientists can use models to represent parts of the world because they can interpret parts of a principled model in physical terms and identify parts of a representational model constructed from the principled model with parts of the real world (Giere 2010, 271).[1] The empirical content is given to the representational model through what Giere calls a specific "hypothesis," which is a claim that "a fully interpreted and specified model fits a particular real system more or less well" (Giere 2010, 271). To return to the Newtonian model example, the representational model of a Newtonian two-body system gets empirical content when one claims that the model fits the Earth-Moon system or any other real-world two-body systems. Such hypotheses about representational models are tested by comparing representational models with models of data rather than directly with data.

Giere says that his account "capture[s] a significant part of scientific practice. And most of what scientists and other theorists of science want to say about this practice can be accommodated within my framework" (Giere 2010, 272). But we might disagree with him in various ways. For example, we might find that his hierarchical view is too simple. In this vein, Marcel Boumans argues that the actual process of modeling is not a linear progression from a body of theory (Giere's laws and principles) to models (Giere's principled models and representational models). Rather, he says, the actual process is one of integration of various elements—theoretical ideas, mathematical concepts, analogies, empirical data, etc—into a representational model (Boumans 1999, 91–94). A different kind of criticism, the one I pursue in this dissertation, is that the basic picture of the practice of modeling suggested by the semantic view is mistaken. Like the formal version, the informal version of the semantic view presents a particular picture of modeling: Models are not linguistic entities, and scientists use sentences, equations, graphs, and other representational devices to describe the properties of the models. Thus, the semantic view puts external representations to the periphery of the practice of modeling. After all, external representations are *not* models. As I show in Chapter 2, this picture of modeling obscures important aspects of the practice of modeling.

[1] As for interpretation, Giere insists that scientists do not begin with uninterpreted models and then provide interpretations. Rather, principles already provide physical interpretations of what will become parts of principled models (Giere 2010, 271).

1.3 RESPONSES TO THE CONCEPT OF MODEL IN THE SEMANTIC VIEW

James Griesemer (1990, 1991) has argued that the concept of model in the semantic view is inadequate because there are models used in theorizing that are material objects.[2] He draws our attention to remnant models, which are material models "*made from* the very individuals modeled" (Griesemer 1990, 8). A specimen in a natural history museum, for example, is made from remains of an organism it represents. According to Griesemer, natural history specimens are "physical objects which, for specific scientific purposes, are taken to represent the whole, living individuals of which they were once part. As such, specimens are remnant models of their wholes" (Griesemer 1990, 8). Using the work of Joseph Grinnell as a case study, Griesemer argues that in addition to being representations of nature, remnant models can also serve as basis for theory construction, because remnant models are robust against theory change, such as change in species concepts. In other words, remnant models can be studied from multiple theoretical perspectives (Griesemer 1991, 80).

One response to Grisemer's criticism is that it is possible to represent the representational function of material models in terms of the model theoretic concept of model. Thus, referring to the Watson-Crick model of DNA and Griesemer's example of remnant models, Steven French and James Ladyman say:

> Of course, at one, superficial, level, these *are* simply what they are: bits of wire and tinplate, or brightly coloured balls held together by plastic rods. The obvious, but important, move is to ask what their *function* is and this is equally clear: it is to represent. The famous Crick and Watson model represents, not just a particular example of DNA but *all* DNA (of that kind); likewise, the function of the coloured-balls-and-plastic-rods model of benzene, gathering dust in the school laboratory, is to represent, not a particular molecule, but *all* molecules of benzene. And this function . . . can easily be captured by the semantic view. (French and Ladyman 1999, 109)

I think French and Ladyman missed Griesemer's point in a crucial way: Griesemer argues that material models can serve scientific theorizing in different ways than abstract entities of the se-

[2] As Griesemer notes (1990, 7), this criticism echoes the points made by Max Black (1962, Ch. 13), Mary Hesse (1966), and Peter Achinstein (1968, 209–211).

mantic view (Griesemer 1991, 80–81). This difference occurs because of the material features of models, and French and Ladyman cannot capture this difference if they reinterpret material models as representing set-theoretic structures.

Another response to Griesemer's criticism is to accept that not all models are abstract entities.[3] This route has been taken by the defenders of the informal version of the semantic view (Teller 2001, Giere 2006, Weisberg 2013). But according to Giere (2004, 747) and Teller (2001, 397–398), most models in mature science are abstract objects. For example, Giere says:

> At first sight, the things that are commonly called models seem to form a quite heterogeneous class including physical models, scale models, analogue models, and mathematical models, just to name a few. Thus we have Watson's original tin and cardboard model of DNA, Rutherford's solar system model of atoms, the Bohr model of the atom, and the de Sitter model of spacetime. There are also equilibrium models in economics and drift models in evolutionary biology. I think it is possible to understand models in a way that usefully encompasses much of this heterogeneity. . . . [M]odels in advanced sciences such as physics and biology should be abstract objects constructed in conformity with appropriate general principles and specific conditions. (Giere 2004, 746–747)

Giere accepts that models can be abstract entities or concrete entities, but he insists that many or most models in advanced sciences are abstract entities.

It is now commonplace among philosophers to categorize models into abstract entities and concrete entities, regardless of whether these philosophers also accept the semantic view of theories. This classification then serves as a starting point for their analysis of the practice of modeling. Here are some representative quotes from recent publications:

> Models are abstract structures or physical structures that can potentially represent real-world phenomena. Many different things can serve as models including physically constructed scale models, model organisms, and mathematical objects such as sets of trajectories through a state-space. (Weisberg 2007, 216–217; see also Weisberg 2013, 7)

[3] For the argument that this response leads to a deflationary version of the semantic view that does not pretend to capture all forms of scientific theorizing, see Downes (1992).

> Models can range from being objects, such as a toy airplane, to being theoretical, abstract entities, such as the Standard Model of the structure of matter and its fundamental particles. (Bailer-Jones 2009, 2)

> The words "scientific model" refer to such a variety of entities that it is difficult to say anything that would be true about *all* kinds of scientific models. For instance, not all of them are abstract entities, because the wood models of molecules, and their contemporary surrogates, namely, three-dimensional computer-generated images, are concrete models, the interest of which being that they can be easily handled and looked at from different points of view. Likewise, models are not all mathematically presented, even if many of them are. (Barberousse and Ludwig 2009, 56)

> Scientific models seem to be ontologically quite diverse, including physical scale models, diagrams, and abstract (or theoretical) structures. There are several ways of dealing with this diversity. One way is to take physical models, diagrams, and so on as being ontologically unproblematic and concentrate on the more problematic abstract models. A second, more radical, solution is to regard physical models, diagrams, and so on, as resources for partially characterizing abstract models. So, all scientific models are regarded as being abstract, or at least having abstract counterparts. (Giere 2009, 249)

The authors of these quotes accept the classification of models into abstract or concrete entities. In the last quote, Giere mentions two ways that analysis of models and the practice of modeling can proceed, given this classification. The first way is pursued by, for example, by Giere and Weisberg, and the second by French and Ladyman.

Other philosophers argue that some models are *imagined concrete entities*. For example, Peter Godfrey-Smith says:

> The move that most people have been tempted to make is to say that model systems are "abstract mathematical objects" of some kind. This general outlook is familiar from the literature on the "semantic view," and it has been taken over by some writers within what I see as the alternative project of analyzing model-based science. Giere seems attracted to a view of this kind, and so is Weisberg .... My aim is not to reject this idea outright, but I will argue that it is deficient in at least some cases. It is important to the practice of model-based science, at least some of the time, that model systems can be conceived

> and treated in a more concrete way. Roughly, we might say that model systems are often treated as "imagined concrete things"—things that are imaginary or hypothetical, but which would be concrete if they were real. (Godfrey-Smith 2006b, 734–735)

Similarly, Roman Frigg says:

> What kind of things are model systems? Some, for instance wood models of a car that we put into a wind tunnel, are physical objects. But most models . . . are not. . . . The view of model systems that I advocate regards them as imagined physical systems, i.e. as hypothetical entities that, as a matter of fact, do not exist spatio-temporally but are nevertheless not purely mathematical or structural in that they would be physical things if they were real. (Frigg 2010, 253)

Godfrey-Smith and Frigg find the concept of abstract entities deficient, because abstract entities do not have physical properties. But scientists talk about ideal systems—systems that do not exist in the real world—as if they have physical properties. Thus, Godfrey-Smith and Frigg suggest that these systems, which they refer to as "model systems," are like concrete entities except that they are purely imaginary. They go on to argue that how scientists think about imagined concrete entities is analogous to how people think about imaginary events and characters in literarily fiction and that philosophical theories of fictions can illuminate the practice of scientific modeling (Godfrey-Smith 2009, Frigg 2010).

1.4 METAPHYSICS BASED PERSPECTIVE ON MODELS

What is common among the informal version of the semantic view and responses to it is that it answers the question of what a model is by applying the concepts of *metaphysics*: abstract entities and concrete entities.[4] The notion of abstract entity relevant here is the most popular one in metaphysics. It says that abstract entities are either non-spatiotemporally located or causally inefficacious or both. Paradigmatic examples of abstract entities are mathematical entities, such as numbers, functions, and sets. On the other hand, tables and chairs are paradigmatic examples

[4]Other concepts, such as abstract structures and imagined concrete entities, are based on the concepts of abstract and concrete entities.

of concrete objects, and so are inscriptions on paper, diagrams on a board, pieces of cardboard on a desk, and other material objects in scientists' environment. Unlike abstract objects, they are spatiotemporally located and causally efficacious (Rosen 2012). In other words, despite the differences in their detailed views about models and the practice of modeling, many philosophers take the *metaphysics based perspective* on models: their views about models and modeling are based on the categories of objects developed in metaphysics.

Proponents of the formal version of the semantic view do not take or avoid the metaphysics based perspective on models. For example, in advocating the application of Tarski's concept of model to scientific theories, Suppes' goal was to suggest that this concept of model and model theory are the appropriate tools for the formal analysis of scientific theories (Suppes 1960, 294–295). For Suppes, models were set-theoretic entities, and he could have said that models are abstract entities because mathematical entities, which include set-theoretic entities, are abstract entities. But he did not make this move. This may be because he knew that physical objects are also used as models in science (Suppes 1960, 291–292), but more importantly, I think he did not explicitly make a metaphysical claim because it was not necessary for his purpose. Suppes wanted to give formal analysis of scientific theories and argued that the right tool for this task is model theory, which studies mathematical structures (models in Tarski's sense). To use model theory, it is not necessary to classify mathematical structures into categories developed in metaphysics, just as we can use arithmetic without classifying numbers into categories developed in metaphysics.

But other philosophers we have seen above have a goal of understanding or analyzing the practice of modeling. This goal is different from Suppes' goal, insofar as the formal analysis of theories omits the details of the actual practice of modeling. And van Fraassen (2008, 311) explicitly says that the semantic view of theories (i.e., the formal version he advocates) does not aim to capture how modeling is actually done. Further, these philosophers have criticized the formal version of the semantic view by saying that models that scientists actually use are not like models in model theory. Thus, to say what a model is, these philosophers need to use other concepts, and since models are objects, it makes sense that they adopt general categories of objects, such as abstract and concrete entities. In doing so, as noted above, they adopt the metaphysics based perspective on models.

1.5 COGNITIVE SCIENCE BASED PERSPECTIVE ON MODELS

My goal is also to understand the practice of modeling. But instead of the metaphysics based perspective, I take what I call the cognitive science based perspective. From the metaphysics based perspective, we see models as *objects* that scientists use to represent the world and group them into metaphysical categories of objects (abstract entities, concrete entities, imagined concrete entities, etc). But we can also see models as *representations* that scientists use in service of their cognitive and epistemic goals. From this point of view, to understand models and the practice of modeling, the natural place to look for conceptual resources is the scientific study of representations, especially cognitive science. In this sense, the perspective I advocate is *cognitive science based*.

1.5.1 The Importance of External Representations

Among the conceptual resources available in cognitive science, I suggest that the concepts of internal and external representations can help illuminate the practice of modeling in science. Internal representations are representations internal to a person's mind: they are structures in the mind that bear informational content. They are synonymous with mental representations. There is a long-standing debate about the nature of mental representation: whether it is like a proposition in formal logic, a rule, a concept, an analogy, an image, or a network of neuron-like units (for an introductory review, see Thagard 2005; for a critical discussion, see Ramsey 2007). But no one would deny the importance of internal representations and operations on them (memory, inference, etc) in cognitive and epistemic activities of scientists. Thus, my focus will be on external representations. Although there is a debate about the very boundary of the mind (Clark and Chalmers 1998), it suffices here to stipulate the skin of an organism as the boundary between internal and external representations. External representations are representations external to a person's mind, that is, in her environment: they are structures in the environment that bear informational content. A person operates on them through perceptual means and hands-on manipulations.[5]

[5]This characterization of external representations is brief but sufficient for my purpose. It is a simplification of the characterization given by Zhang:

> In the present study, external representations are defined as the knowledge and structure in the environment, as physical symbols, objects, or dimensions (e.g., written symbols, beads of abacuses, dimensions of a graph, etc.), and as external rules, constraints, or relations embedded in physical configurations (e.g.,

Although the terms ‘structures’ and ‘information’ have technical senses, in saying that external representations are structures in the environment carrying informational content, I am using these terms in their colloquial sense. A structure refers to an arrangement or configuration of things in the environment. A diagram, for example, is a structure in this sense: it is a configuration of written lines, curves, and so on. Information carried by a structure is the meaning that a person can extract from the structure. To interpret a structure as carrying information, that is, to see something as an external representation, it is not necessary to be able to understand or express the meaning of the representation. For example, we can say that a cuneiform inscription is an external representation without being able to decipher it. Indeed, the fact that scholars tried to decipher it shows that they saw the inscription as an external representation before they understood its meaning.

In the past few decades, important research on how external representations and operations on them influence human cognitive activities has been done under the rubric of situated cognition (Suchman 1987), embodied mind (Lakoff and Johnson 1980, 1999; Varela et al. 1991), distributed cognition (Hutchins 1995), and extended cognition (Clark 2008).[6] This research emphasizes how real-life human cognitive processes occur as people interact with other people and representations in their environment. To put this in terms of distributed cognition, real-life cognitive processes are not entirely internal to a person’s mind but distributed across a group of people and the external environment.

No one would deny that external representations are *useful* for performing cognitive tasks and that human beings frequently use them. To do a fairly complicated algebra, for example, it is useful to write down each step; the written down steps serve as an external representation of our previous moves. It helps our memory as well as our planning of next moves. All this seems trivial. But the research in cognitive science just mentioned supports a nontrivial claim that external representations are important for human cognitive activities not merely in the sense that they are useful tools but in the sense that they are *constitutive elements* of the processes by which human beings perform cognitive tasks. If interpreted strongly, this claim implies that external representations and perceptual and physical operations on them are required for performing cognitive tasks. But

spatial relations of written digits, visual and spatial layouts of diagrams, physical constraints in abacuses, etc.). (Zhang 1997, 180)

[6]For reviews, see Hollan et al. (2000), Wilson (2002), Robbins and Aydede (2009). Norman (1993, Ch. 6) gives a particularly accessible introduction to distributed cognition.

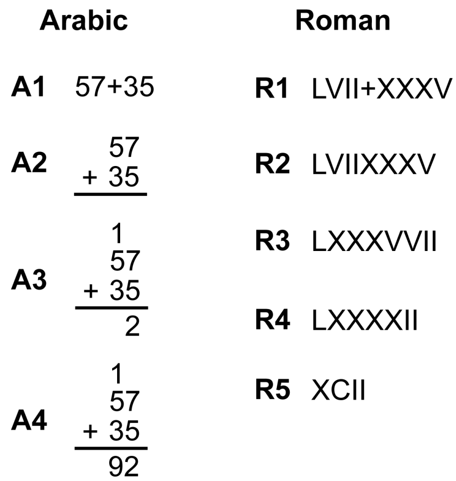


Figure 1.2: Addition in Arabic and Roman numerals. *Left*: Steps of longhand addition in Arabic numerals. *Right*: Steps of addition in Roman numerals.

this is too strong, since some cognitive tasks, such as simple addition, can be done entirely in the head.[7] Rather, the claim should be interpreted to mean that *if* human beings use particular external representations to perform a cognitive task, then those representations and perceptual and physical operations on them are constitutive elements of the processes by which human beings perform that task.[8] This is not a trivial claim because it means that when external representations are being used, they are essential to the process by which the task is done so that if different external representations are used, the cognitive process will be different.

To motivate this point, let us compare how Arabic and Roman numerals transform the processes by which we add two quantities, 57 and 35. We have two different external representations: '$57 + 35$' and 'LVII + XXXV'. To do addition with Arabic numerals, suppose we construct a representation for longhand addition (Figure 1.2). This external representation partially represent the rule of place notation: as long as we add numbers that are aligned vertically—a pattern we

[7]Even so, there is evidence that mental calculations rely on internalized forms of external representations (e.g., place-value notation in Arabic numerals) and mental simulation of the operations on external representations (e.g., calculations with an abacus) (De Cruz et al. 2010, 91–94; Dutilh Novaes 2013, 51–52).

[8]This interpretation of the sense in which external representations and operations on them are constitutive of cognitive processes is a natural extension of Catarina Dutilh Novaes' interpretation of the sense in which manipulations of mathematical notations are constitutive of mathematical reasoning (Dutilh Novaes 2013, 50). Her view is based on research on mathematical cognition from the perspective of extended cognition (see also Dutilh Novaes 2012, Ch. 5).

can easily perceive—we do not have to recall the rule of place notation. Of course, there is an exception: this external representation does not represent the rule of carry, and when we have to deal with carry, we also have to recall the rule of place notation in order to write down numbers at appropriate places. Now, with this external representation, we still have to recall the arithmetic combinations of the single digit numbers (there are forty five combinations to memorize) and apply them to the vertically aligned numbers, starting from the right most column. We then obtain $57 + 35 = 92$.

Addition in Arabic numerals requires a lot of internal operations, but with the external representation, we use pattern matching (vertical alignment), the right-to-left spatial orientation, as well as the movement of our hand from one column to next in order to create the sequence of single-digit additions needed for the task. Note that without the external representation, we will have to rely entirely on the concepts (internal representations) of ones place, tens place, and so on to organize the sequence of additions.

Roman numerals, however, call for very different operations. The internal operations involved are recalling the symbols (there are seven) and the simplification rules (one rule for each symbol). The rest are perceptual and bodily actions (Figure 1.2). Thus, to go from Step R1 to R2, we simply write down all the symbols appearing in the problem next to each other. To go from R2 to R3, we group the symbols according to perceived similarities of the symbols' shapes and then write down the symbols from left to right, following the rule that the symbol for the largest number goes to the left (Norman 1993, 66–75).

This simple case illustrates how different perceptual and bodily processes are used to perform a cognitive task, depending on the kind of external representations being used. For example, in the case of Arabic notation, the perceptual process used is that of spatial alignment of symbols, and the bodily (i.e., hand) movement used is a sequential movement from one column to next. In the Roman case, the perceptual process used is that of similarity between shapes of symbols, and the bodily movement used is rewriting similar symbols next to each other (moving symbols for larger numbers to the left).

This case also illustrates how even in the simple case of addition, paying attention to external representations is important for better understanding mathematical practice. In Chapter 2, I develop this thought further in the context of mathematical modeling, arguing that paying attention to external representations and scientists' engagement with them is important for better understanding the practice of modeling. In that chapter, I also discuss Jiajie Zhang and Donald Norman's (1994) experimental study of distributed cognition that supports the claim that external representations, if used, are constitutive elements of the processes by which human beings perform cognitive tasks.

1.5.2 Models as External Representations

If we adopt the cognitive science based perspective on models and use the concepts of internal and external representations to say what a model is, we can make three claims.

C1 Models are representations, which can be internal or external.

C2 Models are internal representations.

C3 Models are external representations.

C1 is weaker than the other two claims and can be interpreted as representing the cognitive science based perspective on models. *C2* and *C3* can be interpreted as representing possible focal points of this perspective. That is, if we accept *C1* and *C2*, we are adopting the cognitive science based perspective and specifically taking *C2* as a guiding idea. We will then focus on internal representations and operations on them in the practice of modeling. Research on model-based reasoning in science (Magnani et al. 1999, Magnani and Nersessian 2002, Nersessian 2008) is a good example of the cognitive science based perspective on models with the focus on internal representations.[9] If we accept *C1* and *C3*, we are also adopting the cognitive science based perspective but taking *C3* as a guiding idea. Rather than internal representations, our focus will be on external representations in the practice of modeling. These two focal points are complementary, and debating whether *C2* or *C3* is true would be futile.

[9] A prominent view in cognitive science is that mental representations are like models: humans think by constructing and manipulating mental models (Johnson-Laird 1980, 1983). Unfortunately, it is not always clear exactly what proponents of this view mean by model or mental model (Rips 1986; Nersessian 2008, 93). Clarifying and elaborating on the idea of mental models, Nancy Nersessian characterizes what she calls a conceptual model (a complex type of mental models) as follows:

> A model, for my present purposes, can be characterized loosely as a representation of a system with interactive parts and with representations of those interactions. …[Conceptual models] are imaginary systems designed to be structural, functional, or behavioral analogues of target phenomena. The models are dynamical in that future states can be determined through mentally simulating the model. (Nersessian 2008, 12; see also p. 93)

Drawing on research on situated cognition, distributed cognition, and embodied mind, Nersessian studies scientists' external representations to infer what their conceptual models and internal operations on them are like (e.g. Nersessian 2005, 2008).

1.5.3 Overview of the Dissertation

In this dissertation, *C3* is my guiding idea: I shall see models as external representations. Each of the following chapters defends particular theses about models and the practice of modeling, and the entire dissertation is an extended argument for the claim that shifting our perspective on models from the metaphysics based perspective to the cognitive science based perspective with the focus on external representations helps illuminate the practice of modeling in science. But this shift of perspective may not seem substantive. For, after all, I am not questioning that representations are objects and that a model is a representation of parts of the world. But the important difference is found in conceptual resources that inform each perspective: the metaphysics based perspective applies concepts from metaphysics to models, whereas the cognitive science based perspective applies concepts from cognitive science and other studies of representations to models. This difference alone does not suggest any incompatibility between the two perspectives, but I argue in Chapter 2 that the cognitive science perspective illuminates the practice of mathematical modeling, while the metaphysics based perspective obscures it.

Chapter 2 presents Ronald Giere's and Michael Weisberg's accounts of the practice of modeling. I show that their accounts are developed from the metaphysics based perspective. In their accounts, some models, such as scale models, are concrete, physical entities, and they are called concrete models. Other models, such as mathematical models, are abstract, non-physical entities, and they are called abstract models. Giere and Weisberg then suggest that scientists use interpreted equations (and other external representations) to construct, manipulate, and analyze abstract models. The idea that equations are descriptions of models rather than models themselves echoes what the semantic view says about the practice of modeling. After presenting Giere's and Weisberg's accounts, I argue that the category of abstract models obscures rather than illuminates the practice of modeling and that to better understand this practice, including mathematical modeling, we should see models as external representations. I analyze mathematical modeling from this perspective.

Given the overall picture of the practice of modeling defended in Chapter 2, Chapters 3 and 4 analyze different aspects of this practice in detail. Chapter 3 concerns the relationship among uses of a model, particular aims of research in which scientists use the model, and criteria of success relevant to a given use of the model. I argue (i) that the relationship between uses of a model and particular aims of research is dynamic in the sense that uses of a model can shift over the course of scientists' research in response to the shift in aim, and (ii) that criteria of success for one use of a model can be different from those for another use of the same model. I argue for these theses using a detailed case study of Sewall Wright's research on a specific plant population (*Linanthus parryae*). The case study reveals that the context of Wright's research radically changed over the course of research between 1941 and 1978, while he continued using the same models developed near the beginning of his research. My analysis of Wright's research shows that he put models to a variety of uses and that some of Wright's uses of models changed in response to the changing aims of his research and that criteria of success for one use of his model were different from—even irrelevant to—another use of the same model.

Chapter 4 argues that in successful scientific research, a scientist uses a model according to the methodological principles of realism and instrumentalism despite the tension that they create among the scientist's uses of the model over time. I develop this thesis by reflecting on Howard Stein's idea on realism and instrumentalism. After giving precise formulations of the realist and instrumentalist methodological principles, I argue for my thesis through a detailed analysis of successful scientific research done by Seymour Benzer in the 1950s and 60s. I then argue that epistemic realism or epistemic instrumentalism—forms of realism and instrumentalism familiar in the philosophical literature—by itself prohibits a scientist from adopting both the realist and instrumentalist methodological principles. This result poses new challenges to realists and instrumentalists, and I briefly suggest possible avenues of response that realists and instrumentalists may take.

1.6 AGENTIAL PERSPECTIVE ON MODELS

One of the lessons of the case studies in Chapters 3 and 4 is that scientists use models in many different ways to accomplish the aims of their research. This thesis is similar to one of the central theses emerging from the contributions to *Models as Mediators* (Morgan and Morrison 1999). Although I am very sympathetic to their focus on heterogeneous roles of models, the models-as-mediators group adopts what I call the *agential perspective* on models, which results in descriptions of the practice that I find wanting. Thus, to further clarify my perspective on models, I briefly contrast two perspectives.

From the agential perspective, we see models as agents playing some roles. Models are "*autonomous agents*" (Morrison and Morgan 1999, 10) that plays a variety of roles in scientific practice. Models are autonomous in the sense that they are, at least partly, independent from background theories and the world. Margaret Morrison and Mary Morgan (1999) survey the roles of various models found in the case studies in *Models as Mediators* and classify them into three broad categories: "Models in theory construction and exploration," "Models and measurement," and "Models for design and intervention" (Morrison and Morgan 1999, 18–25). For example, in the first category, Morrison and Morgan include the use of models to introduce concepts and to explore characteristics or implications of a theory in a concrete situation (1999, 18–19). In the second category, they mention the use of models as measurement devices (1999, 21), and in the third category, they include the use of models to design various technologies, including means of intervention (1999, 23–24).

In my view, models can be autonomous agents only in a metaphorical sense, because they are not in fact agents like human beings. As Daniela Bailer-Jones observes:

> The terminology of "autonomy" clearly implies that there is an agent, especially as Morrison talks about "autonomous agents," yet it is clearly not the case that a model can be anything like a human agent who "decides" to "act" in some autonomous manner. In contrast, it is people who may decide how closely a model needs to relate to the world, and how tightly empirical and theoretical constraints have to be adhered to in a particular instance. (Bailer-Jones 2009, 136)

While the agential perspective may be useful to reveal various purposes for which models are used, it tends to locate agency in a wrong place. In fact, the descriptions of scientific practice offered by those who adopt the agential perspective often tell us what *models* do rather than what scientists do with models. Morrison and Morgan frequently use the phrase "models function as ..." (e.g., Morrison and Morgan 1999, 18–24). This locution fits the agential perspective, but a consequence of using this locution is that very rarely descriptions of the practice of modeling feature human beings and their actions (see, e.g., Morrison 1998, 71–76; Cartwright 1999, 263–278).

This must be an unintended consequence of the agential perspective. But the agential perspective on models seems to lack resources to capture embodied aspects of scientific practice. For example, it does not help us to specify the perceptual and bodily processes that scientists use to engage with models in order to perform cognitive and epistemic tasks, because from this perspective, it is models that perform these tasks. But the talk about models as agents playing various roles is a metaphorical way of talking about a variety of purposes for which scientists use models, and Morrison and Morgan sometimes talk in this way (e.g., Morrison and Morgan 1999, 11). And they would not disagree with Bailer-Jones that it is in fact human beings (scientists) who perform actions with models in order to accomplish some purposes.

The agential perspective does not seem to offer a distinct advantage over the cognitive science based perspective. From either perspective, we can characterize various uses of models in science. But the agential perspective can be misleading to the extent that it misattributes agency to models and fails to describe human actions. The cognitive science based perspective does not have this disadvantage.

1.7 CONCLUSION

In this chapter I have explained some of the most influential answers to the question of what a model is. In the semantic view of theories, models are understood as mathematical entities (e.g., set-theoretical structures) or abstract entities (in the informal version of the semantic view). In response to the latter idea, philosophers have argued that some models are concrete entities or imagined concrete entities. I have argued that except for the conception of models as mathematical entities, all these views take the metaphysics based perspective on models in the sense that they apply the basic concepts of objects—abstract and concrete entities—developed in metaphysics. I have sketched the alternative perspective that applies the concepts of internal and external representations from cognitive science. My own perspective is the cognitive science based perspective with a focus on external representations: I see models as external representations that scientists interact with via perceptual and bodily processes in order to perform cognitive and epistemic tasks. I then contrasted this perspective with the agential perspective characteristic of the models as mediators group. I have argued that the agential perspective is misleading because it tends to attribute agency to models while neglecting human actions. This is not a tendency that the cognitive science perspective has.

2.0 MODELS AS EXTERNAL REPRESENTATIONS

2.1 INTRODUCTION

Mathematical modeling features equations and graphs that are interpreted as representing aspects of real-world systems. Consider, for example, a typical textbook presentation of the Lotka-Volterra model of the predator-prey system:

> The simplest imaginable model for the predator-prey system ... is to assume (1) exponential growth of the prey [whose size at time t is $X(t)$] in the absence of the predator [whose size at t is $Y(t)$], (2) a linear functional response for the predators [i.e., the rate of prey capture increases linearly with the size of the prey population], (3) a numerical response for the predators [i.e., the captured prey's influence on the rate of increase of predators] that is a constant times the linear functional response, and (4) a constant death rate among the predators. With these assumptions we have
>
> $$\frac{dX}{dt} = rX - (aX)Y \tag{2.1.1}$$
> $$\frac{dY}{dt} = b(aX)Y - dY \tag{2.1.2}$$
>
> [...] The variable, r, is the intrinsic rate of increase of the prey; (aX) is the functional response and a is the slope of the predator's functional response curve; b is the constant that relates the numerical response to the functional response; and d is the death rate of the predators. (Roughgarden 1979, 434–435)[1]

[1] For convenience, I use X and Y to denote the dependent variables. Roughgarden uses V and P for X and Y, respectively.

The presentation then goes on to show how scientists usually analyze Equations (2.1.1) and (2.1.2), what conclusions they may or may not draw about the real-world predator-prey system based on the analysis of the equations, and how they can modify the equations to accommodate different assumptions (e.g. Roughgarden 1979, 435–450).

Mathematical modeling raises a number of philosophical questions. For example, (i) What is a mathematical model? Is the Lotka-Volterra model the suitably interpreted equations or something else? (ii) What is the practice of mathematical modeling? (iii) How should we understand the representational relationship between a mathematical model and a real-world system of predators and prey? The answers to (ii) and (iii) depend on what we identify as a model, and (iii) is further complicated by the fact that some of the assumptions (e.g., the exponential growth of the prey) appear to be idealizations. Questions like these, of course, are not restricted to mathematical modeling and can arise for modeling more generally.

In this chapter, I focus on (i) and (ii). In an influential account of the practice of modeling that takes the metaphysics based perspective, (i) and (ii) are answered in the following way. Some models, such as scale models, are concrete, physical entities, and they are called *concrete models*. Other models, such as mathematical models, are abstract, non-physical entities, and they are called *abstract models*. Equations and graphs used in mathematical modeling are not abstract models because they are concrete entities. Thus, the Lotka-Volterra model is not Equations (2.1.1) and (2.1.2). It is an abstract model. This account goes on to answer (ii) by suggesting that scientists use equations to construct, manipulate, and analyze abstract models. In this account, then, mathematical modeling is a practice of constructing, manipulating, and analyzing abstract models for the purpose of scientific research.

My aim in this chapter is twofold. First, I want to show that for the purpose of understanding the practice of mathematical modeling, the metaphysics based perspective is misguided because the category of abstract models obscures rather than illuminates this practice. Second, I want to develop alternative answers to (i) and (ii) by showing that to better understand the modeling practice we should think of models as external representations. I argue that this cognitive science based perspective is more useful for analyzing the modeling practice, including mathematical modeling. The plan of the chapter is as follows: Sections 2.2 and 2.3 present the influential account of the practice of modeling developed by Ronald Giere and Michael Weisberg. Section 2.4 criticizes this account by showing that the underlying metaphysics based perspective creates serious problems. Section 2.5 states the basic commitments of the alternative account I favor, and Section 2.6 gives a different analysis of mathematical modeling. Section 2.7 considers some objections to my account, and Section 2.8 summarizes the main argument.

2.2 ACCOUNTS OF MODELING FROM THE METAPHYSICS BASED PERSPECTIVE

Giere and Weisberg have articulated an influential account of the practice of scientific modeling. Their account is developed from the metaphysics based perspective on models and is committed to the following theses:

M1 Some scientific models are concrete entities (call them *concrete models*, while others are abstract entities (call them *abstract models*).

M2 Scientists construct, manipulate, and analyze abstract models by constructing, manipulating, and analyzing external representations of the models, such as equations, graphs, and diagrams.

M1 represents the metaphysics based perspective outlined in Chapter 1 and applies the concepts of abstract and concrete entities to models. According to the common characterization of abstract entities, they are either non-spatiotemporally located or causally inefficacious or both.[2] Paradigmatic examples of abstract objects are mathematical objects, such as numbers, functions, and sets (Rosen 2012). Unlike abstract objects, concrete objects are spatiotemporally located and causally efficacious.[3] *M1* divides scientific models into abstract entities and concrete entities. The Lotka-Volterra model is an example of an abstract model, and the double helix model that Watson and Crick built from metal plates and rods is an example of a concrete model.

M2 states the relationship between abstract models and external representations, such as equations, graphs, and diagrams. These representations are not themselves models but are representa-

[2]This characterization follows what David Lewis calls "The Negative Way" (Lewis 1986, 83). For discussion of other ways to characterize abstract objects, see Lewis (1986, 81–86).

[3]This ontological distinction is different from another abstract/concrete distinction discussed in the literature (e.g., Cartwright 1989, Ch. 5; Cartwright 1999, 259–261). The other distinction has to do with the relation between an abstract (or general) description or concept and a concrete (or specific) description or concept.

tions of abstract models. They are related to abstract models in such a way that when scientists construct, manipulate, and analyze external representations, what they do is to construct, manipulate, and analyze abstract models. In other words, *M2* says what it is that scientists do in practice when models being used are abstract models.

In the rest of this section, I show that Giere and Weisberg are committed to *M1* and *M2*. I also explain their reasons for accepting these theses.

2.2.1 Giere's Account

Giere argues for *M1* through his observation of textbooks in classical mechanics:

> Mechanics texts continually refer to such things as "the linear oscillator," "the free motion of a symmetrical rigid body," "the motion of a body subject only to a central gravitational force," and the like. Yet the texts themselves make clear that the paradigm examples of such systems fail to satisfy fully the equations by which they are described. No frictionless pendulum exists, nor does any body subject to no external forces whatsoever. How are we to make sense of this apparent conflict? (Giere 1988, 78)

The descriptions of the linear oscillator and so on that Giere refers to are what Martin Thomson-Jones calls the descriptions of "missing systems" (Thomson-Jones 2010, 284). As Giere observed, the descriptions of the linear oscillator and so on in textbooks appear to be descriptions of actual systems, but as scientists acknowledge, no real-world system satisfies the descriptions. There is thus a puzzle about missing systems: How should we interpret the descriptions of missing systems?

Giere proposes *M1* as a solution to the puzzle about missing systems:

> I propose that we regard the simple harmonic oscillator and the like as *abstract entities* having all and only the properties ascribed to them in the standard texts. (Giere 1988, 78)

Giere's solution is to provide an ontology of missing systems. Missing systems exist not in the physical world but in the realm of abstract entities, and missing systems are abstract entities that have exactly the properties specified by descriptions of them. Thus, for Giere, the terms like 'the harmonic oscillator' refer to abstract entities that have exactly the properties specified in mechanics textbooks. Moreover, he suggests that abstract entities like the harmonic oscillator are models (Giere 1988, 79), which

> function as "representations" in one of the more general senses now current in cognitive psychology. Theoretical models are the means by which scientists represent the world—both to themselves and for others. They are used to represent the diverse systems found in the real world: springs and pendulums, projectiles and planets, violin strings and drum heads. (Giere 1988, 80)

What he calls "theoretical models" are also called "abstract models" (Giere 1999, 51). Thus, to resolve the puzzle about missing systems, Giere posits abstract entities exactly satisfying the descriptions of such systems, and he further claims that scientists use these abstract entities to represent the real-world systems.

There are three things in Giere's picture so far: statements or equations in the textbook (i.e., descriptions of missing systems), abstract models, and the real-world systems. How are these things related? According to Giere, statements or equations characterize or define models:

> The relationship between some (suitably interpreted) equations and their corresponding model may be described as one of characterization, or even definition. . . . The equations truly describe the model because the model is defined as something that exactly satisfies the equations. (Giere 1988, 79)

To characterize a model in Giere's sense is to construct a model, because he thinks that abstract entities are constructed by scientists with the help of representational devices, such as sentences and equations (Giere 1988, 78; Giere 2004, 747). His view, then, is that when scientists write down equations, such as Equations (2.1.1) and (2.1.2), and give them interpretations, scientists are constructing an abstract model. In addition, Giere understands the relationship between models and the real-world systems in terms of similarity, and scientists need to specify in what respect and to what degree that a given model is similar to a given real-world system. Giere calls claims that specify the similarity relation in this way "hypotheses" (Giere 1988, 80–81). Giere summarizes these relations in a diagram (Figure 2.1), which has been adopted by other philosophers (see, e.g., Weisberg 2003, 8; Godfrey-Smith 2006a, 733).

M2 is a claim about relation *R* in Figure 2.1, and as we saw, Giere understands *R* to be characterization or definition as well as construction. Scientists use linguistic entities, such as sentences

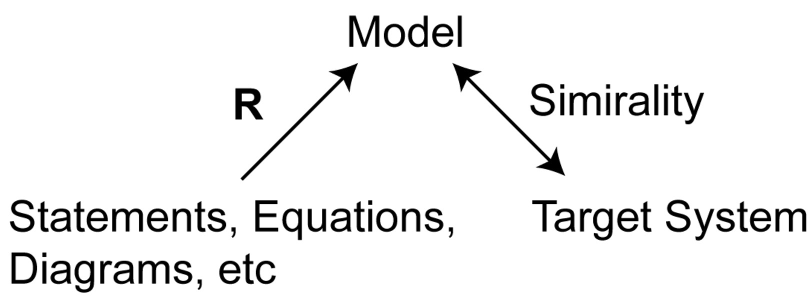


Figure 2.1: Giere's account of the relationship between external representational devises, models, and the real-world target systems. modeling practice. Giere characterizes relation R as statements *characterizing* or *defining* a model. Redrawn from Giere (1988, 83; 1999, 55).

and equations, to define and construct a model, which is not itself a linguistic entity (Giere 1988, 47).[4] Giere proposes *M2* as a better alternative to the logical positivist view of the relationship between a scientific theory and the world. He thinks that the problem with the logical positivist view is its preoccupation with language:

> Most theories of science, whether old or new, assume that any representational relationship between theory and reality would have to be understood as a "correspondence" between scientific statements and the world. The fate of any understanding of theories as somehow representing reality has thus been linked to the fortunes of a correspondence theory of truth. It is here that the battle is usually joined. The interpretation I have offered . . . undercuts these arguments by denying the common assumption. There is, on this account, no direct relationship between sets of statements and the real world. The relationship is indirect through the intermediary of a theoretical model, as pictured in [Figure 2.1]. (Giere 1988, 82; see also Giere 1999, 50, 55)

In other words, Giere wants to make the analysis of the relationship between models and the world not dependent on the analysis of the relationship between language and the world. For Giere, this means that linguistic and other representational devices used to construct and define models are less important than models themselves: "When viewing the content of a science, we find the models

[4]As we saw in Chapter 1, Giere has developed a general account of how models are constructed from principles or laws of a theory.

occupying center stage. The particular linguistic resources used to characterize those models are of *at most secondary interest*" (Giere 1988, 79; my emphasis). I highlight this point, as I argue below that to understand the *practice* of modeling it is important to pay close attention to external representations used by scientists.

2.2.2 Weisberg's Account

Weisberg uses the Lotka-Volterra model and its analysis as his primary case of mathematical modeling (Weisberg 2007, 210–212, 222–223; 2010; 2013, esp., 3–4, 10–13, 25–29, 36–37, 40–41, 74–81). In summarizing his account of mathematical modeling, I want to highlight his metaphysics based perspective, that is, his commitment to ***M1*** (Weisberg 2007, 216–217; see also Weisberg 2013, 7). Weisberg argues that the modeling practice has three distinct stages: construction of a model, analysis of a model, and comparison of a model with a real-world target system.[5] I focus on the first two stages where his commitment to ***M1*** and ***M2*** appears very clearly.

Following Giere, Weisberg distinguishes between a model and its description. In the case of mathematical modeling, for Weisberg, an interpreted equation or graph is not a model but only a description of a model, which is an abstract entity. Weisberg calls relation *R* in Figure 2.1 specification, which can be partial and is less strict than definition (Weisberg 2007, 217; 2013, 34–35). He says:

> At its base, the [specification] relationship is representational; model descriptions represent models. And while the relationship is not one-to-one between models and model descriptions, there is still a very tight link. Thus I will speak of model descriptions as *specifying* models, and of models as *realizing* model descriptions. (Weisberg 2013, 35)

He also says that different model descriptions can describe a single model and that a single model description, if it is vague or imprecise in some respects, can describe different models (Weisberg 2013, 34–35).

Weisberg gives two reasons for distinguishing between a model and a model description. The first reason is that this distinction is an insight of the semantic view of theories: "One of the most important insights behind the semantic view and other attempts to reconstruct theories as sets of

[5] Weisberg observes that the third stage is optional.

models is that a theory should not depend on a particular linguistic formulation" (Weisberg 2007, 217). As we saw, Giere has also argued that this is an advantage of his account over the logical positivist view of scientific theories. The second reason, according to Weisberg, is that in his view, scientists themselves use equations and other external representations to describe models:

> More importantly for understanding the practice of modeling, a modeler often conceives of a model in a vague way, writes down some equations to describe the model she thought she had in mind, studies the model actually specified by the equations, and determines whether or not they pick out the right model. Situations can arise where the modeler's imagination picks out some set of models and her model description picks out a different set of models, necessitating a refinement either to her imagination or to her model description. (Weisberg 2007, 217)

I want to note in passing that Weisberg's reasons are weak. First, an account of the practice of modeling need not be compatible with the semantic view. As we saw in Chapter 1, some proponents of the semantic view have a goal of formally reconstructing a scientific theory, and this goal is not necessarily the same as the goal of understanding scientific practice (Bailer-Jones 2009, 127; see also Downes 1992). Second, the situation Weisberg describes in the above passage does not call for abstract entities at all, for it seems to be the situation where a scientist uses external representations to think further about the content of her internal representation. In fact, Weisberg's own words—"conceives," "had in mind," "the modeler's imagination"—suggest that the situation can be understood as concerning the interaction between internal and external representations, rather than that between abstract entities and external representations.

Let us now turn to Weisberg's account of the first two stages of modeling. The first stage is the construction of a model. In the case of mathematical modeling, this means construction of an abstract entity. Weisberg says that a mathematical model, which is an abstract entity, is constructed by writing down external representations, such as equations and graphs, which describe the model: "In mathematical modeling, construction is achieved by writing down the model description for the model, typically in the form of equations or graphs" (Weisberg 2013, 75). The Lotka-Volterra model is described by Equations (2.1.1) and (2.1.2). The construction of the Lotka-Volterra model is complete when the equations are written down and interpretations are given.[6]

The second stage of the modeling practice is the analysis of a model. As Weisberg says, analysis is done in many different ways, and here I focus on his description of the analysis of the Lotka-Volterra model. According to Weisberg, scientists can manipulate and analyze mathematical models *only* by manipulating model descriptions:

> Equations or other kinds of statements specify mathematical objects and these objects satisfy their descriptions. However, unlike in the case of concrete models, mathematical models can be studied and manipulated only via their descriptions. While the Lotka-Volterra model itself is not a set of equations, it can be studied only through proxies such as these equations. (Weisberg 2013, 36–37)

Thus, scientists can analyze the Lotka-Volterra model only by analyzing its descriptions. In his presentation of the analysis of this model, Weisberg first says that the steady states of the model are found "by setting both equations [Equations (2.1.1) and (2.1.2)] to zero and solving for [X] and [Y]" (Weisberg 2013, 81). He then describes, by paraphrasing May (1973, 42), how the stability of the solutions of the Lotka-Volterra model is studied "by constructing the community matrix for the model" (Weisberg 2013, 81). I quote his description in full below; it will become apparent in Section 2.6 that his description is an inadequate account of how scientists analyze Equations (2.1.1) and (2.1.2).[7] Here's what he says (technical details are not important at the moment):

> To do this [i.e., to study the stability of the solutions], we start from the (unstable) equilibrium point and write down the community matrix, which has the following form:
>
> $$A = \begin{pmatrix} 0 & -\alpha m/\beta \\ \beta r/\alpha & 0 \end{pmatrix} \tag{2.2.1}$$
>
> If we solve for the eigenvalues of this matrix, we get the complex conjugate:
>
> $$\lambda = \pm i(\alpha\beta)^{1/2} \tag{2.2.2}$$

[6] Weisberg identifies four components—"an assignment, the modeler's intended scope, and two kinds of fidelity criteria" (Weisberg 2013, 76)—of interpretation of a model, which is described by an external representation. I skip this detail as it is irrelevant for my purpose.

[7] That is, May's (1973, 42) passage that Weisberg paraphrases concerns only a particular step in the analysis of Equations (2.1.1) and (2.1.2).

> Because this eigenvalue has real parts equal to zero, the oscillations will be neutrally stable. This means that they will continue indefinitely and, if they are disturbed in any way, they will not have a tendency to return to their original amplitude, nor will they become unstable. (Weisberg 2013, 81)

He concludes by saying, "We can learn about the behavior of the model over its entire domain *by doing mathematics*" (Weisberg 2013, 81; emphasis mine). It is true that scientists analyze mathematical models by doing mathematics, but this does not help us understand the practice of modeling much. We still need to know *how* scientists do mathematics, which will be discussed in Section 2.6.

2.3 THE METAPHYSICS BASED PERSPECTIVE AND DISTRIBUTED COGNITION

The accounts of the practice of modeling articulated by Giere and Weisberg are committed to *M1* and *M2*. In this section, I first describe research on distributed cognition and then present the recent attempts by Giere and Marion Vorms to interpret distributed cognition from the metaphysics based perspective. Their attempts are important for my purpose because in my view they have exposed a serious problem with *M1* and *M2*.

2.3.1 Distributed Cognition

The theoretical framework in cognitive science known as distributed cognition emphasizes how real-life human cognitive processes occur as people interact with other people and manipulate objects in their environment (Norman 1993, Zhang and Norman 1994, Hutchins 1995, Hollan et al. 2000, Kirsh 2006). In other words, real-life cognitive processes are not entirely internal to a person's mind but distributed across a group of people and the external environment. In this framework, a cognitive process—a process performing a cognitive task—is said to be distributed if it involves interaction between representations that are internal to the mind of a task performer and those that are external, such as physical objects in the environment. Here I describe the aspects of this research that I take to be important for understanding the modeling practice.

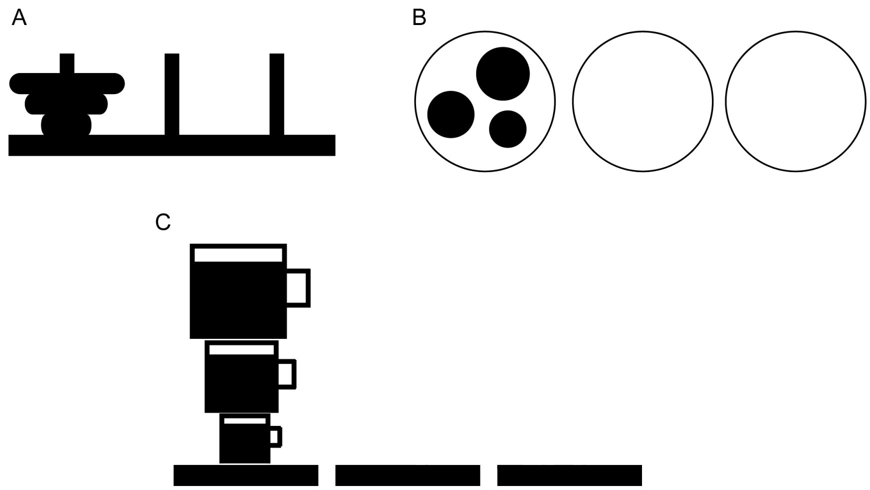


Figure 2.2: Three versions of the Tower of Hanoi: (A) The Tower of Hanoi (standard), (B) the Oranges Puzzle, and (C) the Coffee Cups Puzzle. Redrawn from (Zhang and Norman 1994, 92, 106; Norman 1993, 84, 87, 88)

Let us focus on a game often used in cognitive science, the Tower of Hanoi (TOH). Jiajie Zhang and Donald Norman (1994) developed three versions of TOH. In the standard TOH (Figure 2.2 A), the three disks are stacked on one pole at the start, and the task performer is to move one disk at a time to create a designated configuration. For example, for the starting configuration shown in Figure 2.2 A, the goal configuration has the large disk on the right pole, the medium disk on the left pole, and the small disk on the center pole. The task performer is to follow two rules: (1) Only one disk can be moved from one pole to another at a time, and (2) a disk can only be moved to a pole where it will be the largest. There is another rule that is not stated in the standard TOH: (3) Only the largest disk on a pole can be moved to another pole (Zhang and Norman 1994, 93). In the standard TOH, if the task performer memorizes and follows Rules 1 and 2, then it is not necessary to learn Rule 3, because the physical structure of TOH, together with Rules 1 and 2, guarantees that Rule 3 be followed. Thus, Zhang and Norman say that in the standard TOH, Rules 1 and 2 are represented internally, while Rule 3 is represented externally:

> Internal rules are memorized rules that are explicitly stated as written propositions in the instructions for experiments. External rules are not stated in any form in the instructions.

> They are the constraints that are embedded in or implied by physical configurations and can be perceived and followed without being explicitly formulated. (Zhang and Norman 1994, 93)

Zhang and Norman developed two other versions of TOH to vary which rule is represented internally or externally. In the version of TOH called "The Oranges Puzzle" (Norman 1993, 86; Zhang and Norman 1994, 108), instead of disks on poles, there are three oranges with different sizes that are placed on plates (Figure 2.2 B). At the beginning of the task, the oranges are placed on one plate, and the task performer is to move one orange at a time to create a designated configuration. The rules are:

> Rule 1: Only one orange can be moved from one plate to another at a time.
>
> Rule 2: An orange can only be moved to a plate where it will be the largest:
>
> Rule 3: Only the largest orange on a plate can be moved to another plate.

In the Oranges Puzzle, all these rules are to be represented internally: they have to be memorized.

In the third version of TOH called "The Coffee Cups Puzzle" (Norman 1993, 87; Zhang and Norman 1994, 108), there are three different cups of coffee and plates (Figure 2.2 C). Each plate has space for only one cup, and at the beginning, the three cups are placed on top of each other, the smallest one at the bottom and the largest one on top. The task performer is to move one cup at a time without spilling coffee to create a designated configuration. The rules are:

> Rule 1: Only one cup can be moved from one plate to another at a time.
>
> Rule 2: A cup can only be moved to a plate where it will be the largest.
>
> Rule 3: Only the largest cup on a plate can be moved to another plate.

In this case, only Rule 1 is internal. Rule 2 is represented externally by the fact that coffee will spill if a smaller cup is stacked on top of a larger one. Rule 3 is also represented externally by the fact that a plate holds only one cup. Thus, physically, the only way to place two cups on a plate is to stack a larger cup on top of a smaller one. So the task performer does not need to internalize Rules 2 and 3.

In their experiments, Zhang and Norman had subjects do different versions of TOH and evaluated their performances. They found that the subjects took most time and most steps, and made most errors when given the Oranges Puzzle, and the subjects took less time and fewer steps, and made fewer errors when given the standard TOH or the Coffee Cups Puzzle (Zhang and Norman

1994, 109–110). Zhang and Norman thus concluded that the presence of external rules improved the subjects' performance.

Solving TOH is a cognitive task, and all three versions of TOH can be shown to have the same formal structure (Zhang and Norman 1994, 98–99). Thus, we can say that a task performer performs the same cognitive task when she is given any version of TOH. But Zhang and Norman's study suggests that in order to understand what she actually does to perform the task, we must pay attention to the particular combination of internal and external representations she uses, because the actual cognitive process involves her interaction with external representations. For a task performer, according to Zhang and Norman, the three versions of TOH are "simply three different problems" (Zhang and Norman 1994, 90).

2.3.2 Distributed Cognition and Abstract Models

In his recent work, Giere argues that cognitive processes in science, such as experiments and modeling, are distributed cognitive processes (Giere 2002a,b; Giere and Moffatt 2003; Giere 2006, 100). After showing how we can easily understand cognitive processes involving concrete models as distributed (Giere 2006, 101–105), Giere argues that interaction with abstract models can also count as a distributed cognitive process. But he notes that there are special problems:

> Abstract models provide what is probably the most difficult case for understanding reasoning with models as an example of distributed cognition. It is not clear in what sense an abstract model can be external. Nor is it clear how a person can interact with an abstract model. Yet many, if not most, models used in the sciences are abstract models. Think particularly of models in quantum physics or cosmology. So some account of reasoning with abstract models is needed. (Giere 2006, 105)

Distributed cognition involves an agent and a representation external to her. Thus, as Giere notes, there are at least two ontological problems with abstract models. First, there is *the problem of externality*. Diagrams, pictures, and other physical objects are unproblematically external to an agent, but since an abstract object is not a physical object, it is not clear how an abstract object can be external to an agent. Second, there is *the problem of interactivity*. It is unproblematic that an agent interacts with physical objects, but since an abstract object is neither spatiotemporally located nor capable of standing in a causal relation, it is not clear how she can interact with an abstract object. These problems are particularly serious for Giere's account of modeling because he believes that many scientific models are abstract models.

Giere responds to the externality problem by suggesting that we understand an abstract model like a planned party, a possible but not actual entity. He insists that this response avoids philosophical problems associated with abstract objects. He writes:

> [E]ven if we agree that abstract models are in some sense external, there remains a question of just what this means. This question threatens to lead us into the arid land of the philosophy of mathematics, where one worries about what numbers might be. I think we would do well to avoid this detour and take a safer route. As in our understanding of time, we traffic in abstract entities every day without worrying about what they are or how we interact with them. Consider plans and planning, well-known topics in cognitive science. Here abstract models are simply taken for granted. Suppose three friends are planning a party. The planned party is, at least in part, an abstract model of a party. It is assigned a date and time in the future and potential guests may even be designated in a written list. The party starts out as an abstract entity, a mere possibility, because it may in fact never materialize. (Giere 2006, 105–106)

Here Giere seems to think that understanding a planned party as an abstract object external to us is not problematic because we do so without worrying about potential ontological problems, such as what such an object is or how we interact with it. I will return to this response below, and for the moment I note that this answer begs the question since the externality problem denies that what Giere claims we do everyday—e.g., talking about a potential party as if it is an external object—is unproblematic.

To deal with the interactivity problem, Giere argues that scientists interact with abstract models by using language (Giere 2006, 106). Referring to the party example, he says:

> The three friends in my example build up their model of the party by talking about it. Moreover, they can reason about it as well, realizing, for example, that the number of potential guests has become too large for the intended space. It does not follow, however, that the possible party is itself in any way propositional, a mere linguistic

> entity. My three friends are not talking about what they are saying; they are talking about a possible party. (Giere 2006, 106)

We saw above that for Giere, the relation *R* between language and other representational devices and models is one of characterization of definition (Figure 2.1). He is now making a further claim that language and other representational devices are the means by which we create and interact with abstract models. Giere concludes his discussion as follows:

> Even this rudimentary understanding of abstract models as abstract entities is enough to support our understanding of the development and use of abstract models as an instance of distributed cognition. Traditional scientists sitting alone with pencil and paper are already distributed cognitive systems. They interact physically with their diagrams and equations and, thereby, abstractly with an assumed more complex abstract model. (Giere 2006, 106)

Thus, for Giere, scientists interact with the Lotka-Volterra model by manipulating Equations (2.1.1) and (2.1.2).[8]

In her critical discussion of Giere's account, Vorms (2011) develops a potentially viable solution to the externality problem that does not reject *M1* or *M2*. Vorms accepts *M1* and *M2* (Vorms 2011, 288, 290) but wants to resist the idea, explicitly defended by Giere (1988, 79), that particular linguistic devices used to characterize abstract models are of secondary importance in an account of the practice of modeling. Vorms argues that even if scientists have external representations of the same system, the actual process and efficiency of their reasoning depend on how these representations convey information about the system: they depend on what she calls the "format" of representations (Vorms 2011, 289).[9] She also insists that an abstract model "has to be accessed by a representation of some sort, since one cannot have a direct perceptual access to it" (Vorms 2011, 290). In another sentence, she says: "Since [an abstract model] does not exist in the spatio-temporal world, it has to be accessed by means of our representational capacities such as language" (Vorms 2011, 290). Although the meaning of "access" is not entirely clear, these claims can be seen as a version of *M2*. Vorms appeals to her version of *M2* to argue that whenever scientists use

[8]As we saw, Weisberg also holds a very similar view.

[9]Very roughly, an equation and a graph of an equation convey information differently and thus count as having different formats. For a more precise characterization of formats, see (Vorms 2011, 289–290).

or reason with abstract models, they use external representations of those models (Vorms 2011, 293).[10]

Perhaps Vorms could evade the externality problem by saying that we do not need an explanation of how abstract models can be external objects since scientists never reason with abstract models themselves. Scientists must use representations of abstract models, and these representations are external to them. But this response works only if there is an explanation for *M2*, because it is Vorms's version of *M2* that supports the claim that scientists must use representations of abstract models.

I agree with Vorms that external representations are important in the practice of modeling and that their importance is suggested by research on distributed cognition. But as I argue in the next section, I think her acceptance of *M1* and *M2* undermines her account as well as Giere's and Weisberg's.

2.4 PROBLEMS WITH THE METAPHYSICS BASED PERSPECTIVE

In my view, the problems of externality and interactivity that Giere has identified undermine the metaphysics based perspective.

As we saw, Giere tries to reconcile his view that many or most scientific models are abstract models—i.e., *M1*–with the idea that modeling is a distributed cognitive process where models are external to scientists. Giere offers what he considers a safe response: ontologically, an abstract model is like a possible object, such as a planned party, which in everyday life people talk about as if it is an external object. I think Giere is convinced of the safety of this response because he simply takes for granted the idea that possible objects (e.g., plans) are abstract entities.[11] But this amounts to the claim that possible objects exist outside of our minds and are abstract entities. This is hardly

[10] Ultimately, Vorms argues that philosophers should focus on a representational relationship between statements, equations, and diagrams (which characterize an abstract model) and the target system rather than the relationship between an abstract model itself and the target system (Vorms 2011, 294). In other words, using Figure 2.1, we can say that in Vorms's account, there is an arrow emanating from statements and other things displayed at bottom left to the target system displayed at bottom right. And she questions the value of wondering about the double-headed arrow connecting a model and a target system in the figure.

[11] Giere appears to think that it is taken for granted *in cognitive science* that plans are abstract entities (Giere 2006, 105–106). But at least some cognitive scientists think of plans as internal representations (Suchman 1987).

a safe route, for it is a substantive ontological position about possible objects, which should force Giere to deal with ontological problems about possible objects (see Yagisawa 2013). In short, either Giere substituted one ontological problem (the problem of externality of abstract entities) with another (the problem of possible objects), or he reintroduced the problem of externality in his response by suggesting that possible objects, such as a planned party, are abstract entities.

We also saw that both Giere and Weisberg argue that external representations are means for constructing, manipulating, and analyzing abstract models. For example, manipulation of Equations (2.1.1) and (2.1.2) results in manipulation of an abstract model called the Lotka-Volterra model. This is *M2*. Now, since abstract models cannot stand in a causal relation, it needs to be explained just how our constructing, manipulating, and analyzing external representations like Equations (2.1.1) and (2.1.2) result in the construction, manipulation, and analysis of abstract entities. But in both Giere's and Weisberg's accounts, no such explanation is provided: their accounts are committed to *M2* but do not explain how it can be true. I suggest that the failure to provide any explanation for *M2* tells us that there is a basic problem with the metaphysics based perspective. That is, the metaphysics based perspective does not provide a viable explanation for *M2* because such an explanation is not forthcoming or extremely hard to develop. This makes sense, because providing a viable theory of how concrete objects like external representations interact with mathematical objects conceived as abstract entities is one of the most difficult problems in metaphysics and philosophy of mathematics.[12]

Two lessons emerge from the above discussion. First, if we hold *M1*, it is hard to develop an account of the practice of modeling in which many aspects of the practice are seen as distributed cognitive processes. Second, although it *is* natural to hold *M2*, given *M1*, it is very hard to give an explanation for *M2*. But such an explanation is needed if *M2* is to give us an insight into the practice of modeling. Both lessons suggest that for the purpose of understanding the practice of modeling, *M1* and *M2* are problematic commitments to base our account on. The metaphysics based perspective obscures rather than illuminates the practice of modeling in science.

[12]It is worth mentioning another problem with *M1* and *M2* that is independent of the problems I just raised. For Giere and Weisberg, Equations(2.1.1) and (2.1.2) characterize an abstract entity, which scientists use as a model of a real-world system. Equations(2.1.1) and (2.1.2) say that the values of X and Y change over *time*. But an abstract entity cannot have a temporal property as it does not exist in space and time. So a model description cannot be true of an abstract entity that it is supposed to characterize or define. For a discussion of this problem, see Thomson-Jones (2010).

2.5 THE COGNITIVE SCIENCE BASED PERSPECTIVE: MODELS AS EXTERNAL REPRESENTATIONS

In this and subsequent sections, I develop the cognitive science based perspective as an alternative to the metaphysics based perspective.

My account of the practice of modeling draws on research on distributed cognition. As we saw, when the task performer solved the standard TOH and the Coffee Cups Puzzle, her cognitive process was distributed and was more efficient and reliable than the process for the Oranges Puzzle. For my purpose, the important lesson of distributed cognition research is that our perceptual and bodily interactions with physical objects in a given environment are part of how we perform a cognitive task and that difference in our cognitive and epistemic performance can be explained in terms of how we are able to use our perceptual and bodily processes. As Zhang and Norman argue, external representations transform a cognitive task in the sense that for a task performer, external representations are constitutive of the task (Zhang and Norman 1994, 118–119).

I will develop my account by gradually extending the framework of distributed cognition to modeling. My account differs from Giere's and Weisberg's accounts because the central analytic categories are those of internal and external representations rather than those of concrete and abstract objects. My account is based on the following theses:

E1 Scientific models are external representations.

E2 Scientific modeling is a practice of constructing, manipulating, and analyzing external representations in service of cognitive and epistemic aims of research.

E1 represents the cognitive science perspective outlined in Chapter 1. *E1* may be interpreted to mean that models are concrete entities, since external representations are concrete entities. But this interpretation is misleading as it can encourage those who accept *M1* to react that *E1* leaves out abstract models. Rather *E1* should be understood as a rejection of the use of the categories of abstract and concrete entities.[13]

[13] In a series of papers, Tarja Knuuttila has argued that to understand how a model gives us knowledge, we should regard a model as a concrete, artifactual entity. If a model is a concrete artifact, she argues, it is not difficult to see how scientists' manipulation of it can be cognitively and epistemically valuable despite the fact that a model is not an accurate representation of its target (Knuuttila and Voutilainen 2003; Knuuttila 2005; 2011; Boon and Knuuttila 2009; Knuuttila and Boon 2011). Although her question (how do models give us knowledge?) is different from mine, the characteristics of models (understood as concrete entities) that she highlights (Knuuttila 2011, 267–270) are also characteristics of external representations. Thus, in this sense, my account is complimentary to hers. But her argument

A more detailed formulation of *E1* is needed if we want to tell whether a given external representation counts as a model. Clearly not every external representation is a scientific model. For example, the inscriptions on this page are external representations of English sentences. We interpret them as representing sentences in English. But these inscriptions do not count as scientific models because, as I noted in Chapter 1, models are representations of parts of the world that scientists use for cognitive and epistemic purposes relevant to their inquiry. I will combine this point about models and the characterization of external representations given in Chapter 1 into a more detailed formulation of *E1*. To do so, I adopt the following terminology. I use the noun 'representation' to refer to a structure that bears informational content: a representation is a vehicle that carries information about something. I use the verb 'to represent' to mean "to carry information about." To say that the inscriptions on this page represent English sentences is to say that the inscriptions carry information about English sentences, and the information in this case is the meaning or content of the sentences. I also use 'representation of' to mean the same thing: 'X is a representation of Y' means X represents Y.

> Scientific Models as External Representations (*E1*): External representations are structures in a person's external environment that bear informational content. A given external representation is a scientific model if a scientist interprets some part or property of the representation as representing (in the sense of carrying information about) some part or property of the object of an investigation and if the scientist uses the representation as a tool to approach the scientist's specific research questions or aims.

Not all external representations satisfy this criterion, and thus we can see scientific models as forming a class of external representations.[14]

against the view that some models are abstract entities (e.g., Knuuttila and Voutilainen 2003, Knuuttila and Boon 2011) can be easily countered by the defenders of this view. For they can accept her positive claims about models as claims about concrete models while maintaining that abstract models require special treatment, or they can accept her claims about models as claims about model descriptions.

[14]My criterion is compatible with but narrower than Paul Teller's. His view is that "in principle, anything can be a model, and that what makes a thing a model is the fact that it is regarded or used as a representation of something by the model users" (Teller 2001, 397). In my view, it is not just "something" but specifically a part or property of of a model that is interpreted as representing a part or property of the object of an investigation. In addition, a model is used as a tool to approach research questions or aims. I also assume that a part or property of a scientific model that is interpreted as representing a part or property of the object of an investigation can stand in a variety of relations with the latter that interest a scientist, depending on the sort of an object the model is, the scientist's interpretation of its parts and properties, and the sort of analysis the scientist plans to apply to the model. For example, if a scientist had a mathematical model with structural properties that she interprets as representing properties of the object of her investigation, and she planned to use algebraic techniques to analyze her model, then she would be

There is an obvious objection to *E1*, which has to do with the identity and individuation of external representations. For example, the system of Equations (2.1.1) and (2.1.2) counts as a model according to *E1*, but so do the system of equations that we obtain by substituting different letters into Equations (2.1.1) and (2.1.2) (N_1 for X, N_2 for Y, etc). So too does a graph of solutions to these systems of equations. It is absurd, so the objection goes, to say that these systems of equations and graphs of their solutions are all different models. Giere and Weisberg, taking the metaphysics based perspective on models, can simply say that these systems of equations and graphs are different descriptions of the same abstract model, but *E1* does not allow me to resort to this solution. So what can we say about the identity of models understood as external representations? I address this question in Section 2.7.

E2 means that scientists' perceptual and bodily interactions with models, understood as external representations, are central to the practice of modeling.[15] This incorporates the insight of distributed cognition research. In the next section I develop an account of mathematical modeling based on *E1* and *E2*, using the Lotka-Volterra model as an example.

2.6 AN ACCOUNT OF MATHEMATICAL MODELING FROM THE COGNITIVE SCIENCE BASED PERSPECTIVE

From the cognitive science based perspective I favor, Equations (2.1.1) and (2.1.2), together with their interpretation in biological terms, are a model. Mathematical modeling is then seen as a practice of constructing, manipulating, and analyzing these interpreted equations. In this section, I show how this perspective illuminates the standard method—phase plane analysis—of ordinary differential equations (ODEs) like Equations (2.1.1) and (2.1.2).

interested in isomorphism or other morphic relations between the structural properties of the model and the structural properties, themselves described mathematically, of the object of her investigation described. If her model were a concrete, physical object some parts of which she interprets as representing parts of the object of her investigation and if she planned to use analogical reasoning to analyze the model, then the relevant relation would be a certain type of similarity relation between parts of a model and parts of the object of her investigation.

[15] Of course, *E2* does not imply that scientists' internal representations are not involved in modeling.

2.6.1 Phase Plane Analysis

To summarize the phase plane analysis of ODEs, I rely primarily on Garrett Odell's (1980) insightful presentation of the method. His presentation distills the basic results of the qualitative theory of ODEs (e.g. Andronov et al. 1973) but does so by highlighting the importance of graphical techniques. Speaking about the aim of qualitative theory, Odell says:

> To *use* differential equations for mathematical modeling, it is not necessary to know how to write down formulae for their solutions (or approximations thereunto). In most cases, no such formulae exist. In many cases for which an exact or approximate (asymptotic) analytical solution can be discovered, the solution formula is so complicated that it discloses nothing about the nature of the solution until a (geometrical) graph of it is drawn. We aim to generate the graphs, at least their general shapes, directly, without bothering with formulae for them. (Odell 1980, 649)

A particular solution to an ODE tells us how the value of the dependent variable (e.g., population size) changes as the independent variable (e.g., time) changes. For example, as the value of the independent variable increases, the value of the dependent variable may increase monotonically until it reaches a stable steady state, or it may oscillate between certain values without settling down to a particular state. Odell is referring to these characteristics when he says "the nature of the solution." A solution to an ODE is a curve, so it may be written down as a function. But as Odell says, for most ODEs, such a function does not exist, or even if it did, it would be so complicated that we would not be able to infer the nature of the solution from it. Either way, Odell says, we have to make a graph of a solution. In other words, as I illustrate further below, to analyze ODEs, scientists rely on external representations like graphs that will enable them to use their visual perception, especially their ability to detect spatial patterns, to identify characteristics of a solution.

To take a closer look at phase plane analysis, consider a system of two ODEs:

$$\frac{dX}{dt} = F(X,Y) \tag{2.6.1}$$

$$\frac{dY}{dt} = G(X,Y). \tag{2.6.2}$$

We want to know the characteristics of the solutions to Equations (2.6.1) and (2.6.2): our cognitive aim is to understand the nature of the solutions. In phase plane analysis, we construct a *phase portrait*—a graph of particular solutions in the X–Y plane—without finding formulae of the solutions. Following Odell, I divide the method into three steps:

> Step 1: Find all the steady state points and draw a graph of tangent vectors.
>
> Step 2: Use linearization to identify the nature of each steady state point.
>
> Step 3: Draw a phase portrait, combining the results of Steps 1 and 2.

Odell nicely summarizes these steps in one sentence: In phase plane analysis, he says,

> geometry, graphical sketching, and local magnifying glass views via linearization, collaborate in a simple way to determine (rigorously) most information about the behavior of a class of [ODEs]. (Odell 1980, 650)

Equations (2.1.1) and (2.1.2) are instances of Equations (2.6.1) and (2.6.2), and phase plane analysis is routinely applied to this external representation. I argue that each step of phase plane analysis involves the construction, manipulation, and analysis of models (external representations), in which scientists use various perceptual and bodily processes as well as internal processes to investigate the nature of the model. This means that external representations are essential for achieving the aim of understanding the nature of the solutions to Equations (2.6.1) and (2.6.2).

2.6.1.1 Phase Plane Analysis: Step 1 We begin the first step of phase plane analysis by finding the steady state points for ODEs. That is, for Equations (2.6.1) and (2.6.2), we want to find the points (X_0, Y_0) at which both $F(X, Y)$ and $G(X, Y)$ are zero. For example, Equations (2.1.1) and (2.1.2) have two such points: $(0, 0)$ and $(d/ba, r/a)$. To find these points, we manipulate Equations (2.1.1) and (2.1.2). If we set $dX/dt = 0$, we can solve for Y:

$$\begin{aligned} 0 &= rX - (aX)Y \\ (aX)Y &= rX \\ Y &= \frac{r}{a} \end{aligned}$$

We apply the similar procedure to the other equation to obtain $X = d/ba$. Instead of using the pencil-and-paper technique just described, we can use a computer program to find these points (often the only way to find the steady state points for complicated equations) (see, e.g., Roughgarden 1998, 267).

Note that these methods—paper or computer—call for different sets of internal (mental) and bodily (perceptual and motor) processes. In manipulating Equations (2.1.1) and (2.1.2) on paper, we recall the rules of algebra (internal process) and write down symbols accordingly (bodily process). That is, we follow our internalized rules of algebra when we move symbols to the left or the right of the equation or eliminate them from both sides. In addition, by writing down a new line of the equation each time we make changes to the previous one, we create a spatial representation of the sequence of our manipulation and application of the rules of algebra. In the example above, I used a spatial orientation (top to bottom) to represent the temporal sequence. This record is useful for planning our next move, and can be inspected if we want to check the accuracy of our result: the record provides both cognitive and epistemic benefit. On the other hand, if we manipulate Equations (2.1.1) and (2.1.2) on a computer, we need a different set of internal and bodily processes. Instead of recalling and applying the rules of algebra, we have to recall the rules of the programming language being used. Instead of creating the external record of our algebraic manipulation, we type (and save) our code on a computer.

Now consider any point in the X–Y plane. Each point has a vector, whose components are $[F(X,Y), G(X,Y)]$, and it can be drawn as an arrow. Call this arrow a "flow arrow" (Odell 1980, 664). In this step of the analysis, our goal is to draw flow arrows in the X–Y plane. If we use a computer, we can draw many arrows as shown in Figure 2.3, but it would also be sufficient if we chose several points and drew flow arrows by hand. Once we draw flow arrows, we can sketch trajectories that are tangential, at every point, to the flow arrows. Such curves are shown in Figure 2.3, and these curves depict parts of solutions to Equations (2.1.1) and (2.1.2).

As in the example of finding steady state points by hand, to find tangent vectors by hand, we have to recall the rules of algebra and arithmetic as well as the concept of vector space. We then have to draw a graph like Figure 2.3. Alternatively, to let a computer program to generate a graph like Figure 2.3, instead of recalling the rules of algebra and arithmetic and the concept of vector space, we need to recall the rules of the programming language we are using and type our instructions in that language.

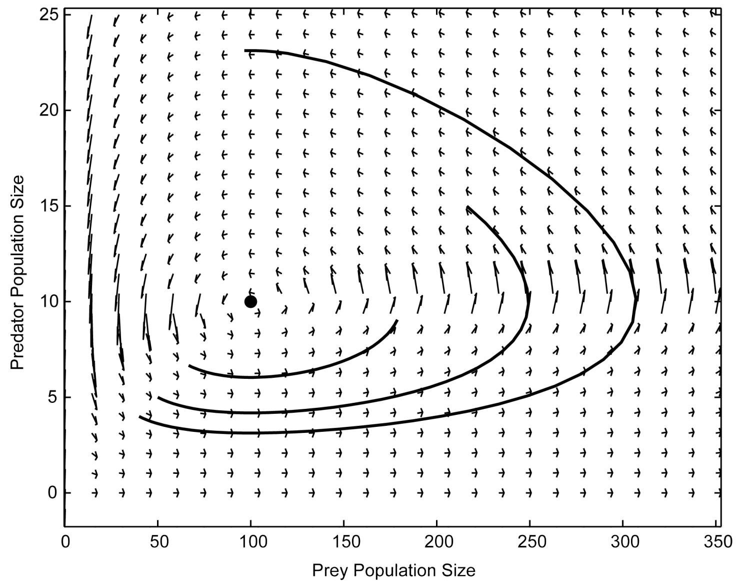


Figure 2.3: Flow arrows for Equations (2.1.1) and (2.1.2), with $r = 1, a = 0.1, b = 0.05$, and $d = 0.5$. The dot indicates the nontrivial steady state point. Three trajectories are partially sketched.

Thus, our interaction with Equations (2.1.1) and (2.1.2) result in our performing a particular set of internal and bodily processes, and these processes are done in service of the cognitive aim of understanding the nature of the solutions to Equations (2.1.1) and (2.1.2).

2.6.1.2 Phase Plane Analysis: Step 2 The second step is to identify the nature of the steady state points. In particular, we study the characteristics of Equations (2.6.1) and (2.6.2) at the points near (X_0, Y_0). As quoted above, Odell describes this step as providing "local magnifying glass views via linearization" (Odell 1980, 650), and this description, though metaphorical, should be taken seriously. For the goal of this step is to get a *visual* understanding of the behavior of the solutions to ODEs at points in a very small neighborhood around a steady state point. We cannot do this, however, by literally magnifying Figure 2.3 around a steady state point, because we are interested in the points whose distance from the steady state point is infinitesimally small. Thus, we use a technique called linearization and a mathematical theory known as stability theory as our "magnifying glass."

To do this, let x and y be small displacements from (X_0, Y_0). We then substitute $X = X_0 + x$ and $Y = Y_0 + y$ into Equations (2.6.1) and (2.6.2). Taking the Taylor series expansion of F near (X_0, Y_0) and ignoring the higher order terms, we obtain the linearization of Equation (2.6.1):

$$\begin{aligned}\frac{d}{dt}(X_0 + x) &= F(X_0 + x, Y_0 + y) \\ &= F(X_0, Y_0) + \frac{\partial F}{\partial X}(X_0, Y_0)x + \frac{\partial F}{\partial Y}(X_0, Y_0)y.\end{aligned}$$

The similar procedure will yield the linearization of Equation (2.6.2). Since (X_0, Y_0) is a steady state point, $F(X_0, Y_0)$ and $G(X_0, Y_0)$ are both zero. Thus, the linearization of F and G leads to:

$$\frac{dx}{dt} = a_{11}x + a_{12}y \tag{2.6.3}$$

$$\frac{dy}{dt} = a_{21}x + a_{22}y \tag{2.6.4}$$

where

$$\begin{aligned} a_{11} &= \frac{\partial F}{\partial X}(X_0, Y_0), \qquad & a_{12} &= \frac{\partial F}{\partial Y}(X_0, Y_0), \\ a_{21} &= \frac{\partial G}{\partial X}(X_0, Y_0), \qquad & a_{22} &= \frac{\partial G}{\partial Y}(X_0, Y_0).\end{aligned}$$

There is a well-developed mathematical theory, called stability theory, that allows us to analyze Equations (2.6.3) and (2.6.4). To apply this theory, we let $\beta = a_{11} + a_{12}$ and $\gamma = a_{11}a_{22} - a_{12}a_{21}$. Then, according to the values of β and γ, the theory allows us to identify the nature of the steady state point as one of the six types displayed in Figure 2.4.

If we linearize Equations (2.1.1) and (2.1.2) near $(d/ba, r/a)$ and apply stability theory, we will find that $\beta = 0$ (Odell 1980, 687; Edelstein-Keshet 2005, 219–220). In Figure 2.4, we see that this result implies that the steady state point is a *center*. If the steady state point is a center, solutions do not approach or run away from it: they circle around the steady state point. This property of the steady state point is also called neutral stability (Odell 1980, 672).[16]

Like the first step, this step involves manipulation of Equations (2.1.1) and (2.1.2). In this step, we also construct a new external representation like Equations (2.6.3) and (2.6.4). Applying stability theory to the new model, we gain visual or geometric understanding of the nature of the

[16]Weisberg notes this result, but not the graphical method, in his discussion of the Lotka-Volterra model (Weisberg 2013, 81).

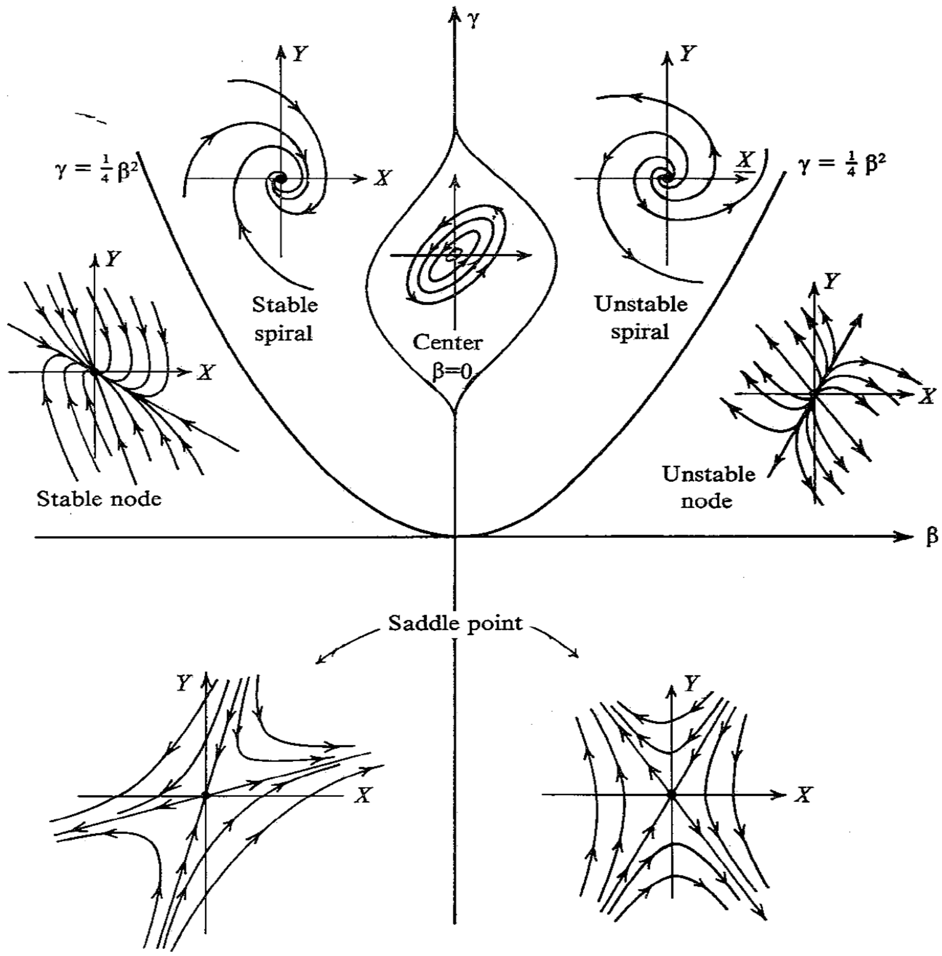


Figure 2.4: Types of steady state point. From Odell (1980, 673). Copyright © Cambridge University Press 1980. Reprinted with the permission of Cambridge University Press.

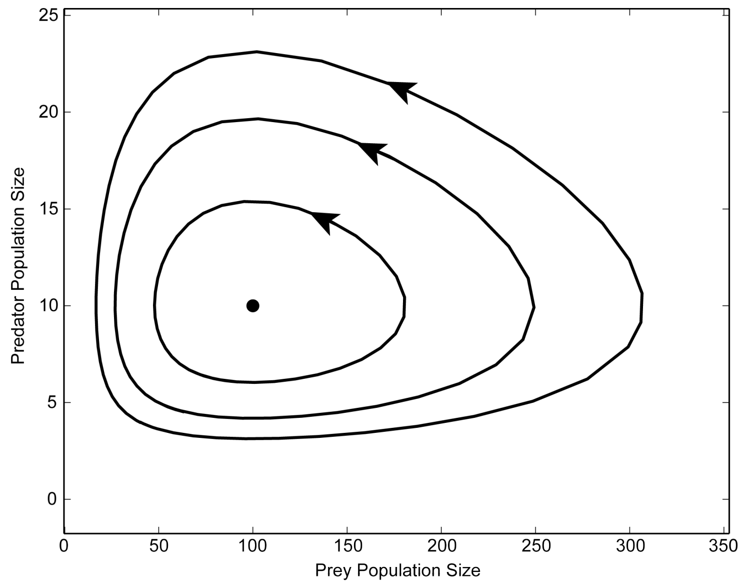


Figure 2.5: The phase portrait for Equations (2.1.1) and (2.1.2), with $r = 1, a = 0.1, b = 0.05$, and $d = 0.5$. The steady state point is a center.

steady state point of Equations (2.1.1) and (2.1.2). That our understanding is visual is illustrated by the fact that the types of steady state points shown in Figure 2.4 are characterized by visual concepts, such as a spiral, a saddle point, and a center.

2.6.1.3 Phase Plane Analysis: Step 3 The last step is to construct a phase portrait. To do this, we finish the diagram (Figure 2.3) we began in the first step by taking into account the visual understanding of the steady state point gained in the second step. We have learned that the nontrivial steady state point of Equations (2.1.1) and (2.1.2) is a center. Thus, our solution curves should form closed loops. Figure 2.5 shows the finished portrait.

As in the first step, we can draw a phase portrait by hand or let a computer to make one. I used a computer to make Figure 2.5. Note that we do not know and did not even try to find the formulae for any of the solution curves in Figure 2.5. Thus, Figure 2.5 is indispensable for gaining any understanding of the solutions to Equations (2.1.1) and (2.1.2).

In sum, each step of phase plane analysis involves the construction, manipulation, and analysis of models understood as external representations, which enable us to use perceptual and bodily processes to achieve the goal of obtaining an understanding of the solutions to Equations (2.1.1) and (2.1.2). Thus, in phase plane analysis, external representations play an essential role for understanding the nature of the solutions to a system of ODEs.

2.6.2 Advantages of the Account

I want to highlight two advantages of my account of mathematical modeling over Giere's and Weisberg's. First, my account overcomes the problems with the metaphysics based perspective discussed in Section 2.4. As noted above, Giere's account faces the problem of the externality of abstract models, and as a consequence it is difficult to characterize mathematical modeling as a distributed cognitive process. My account does not face this problem, because models are external representations. Thus, it easily incorporates the insight of distributed cognition research into an account of the practice of mathematical modeling. We also noted above that Giere's and Weisberg's accounts face the problem of interactivity, which makes it difficult to explain *M2*. That is, their accounts have so far failed to explain how constructing, manipulating, and analyzing external representations, such as equations and graphs, result in the construction, manipulation, and analysis of abstract models. Without an explanation of this relationship between external representations and abstract models, we cannot hope to have an account of how scientists use abstract models. It is hard to explain how what we do to an equation (say) is an interaction with an abstract entity, which does not stand in any causal relation. My account does not face this problem, since this particular explanatory need does not arise in the first place.

Second, as illustrated by phase plane analysis, scientists' perceptual and bodily interactions with external representations are crucial for achieving their cognitive and epistemic aims. My account brings this embodied aspect of scientific practice to the center stage of a philosophical account of the practice of modeling.[17] By doing so, my account also explains why manipulation of a model is important for gaining better understanding of the model. Manipulation of an external representation allows scientists to use their perceptual abilities to interact with the model. This

[17] In this sense, my account is in line with Vorms', although Vorms takes the metaphysics based perspective on models.

interaction can give scientists a perceptual (e.g., visual) understanding of the nature of the model. By contrast, Giere's account makes external representations peripheral to a philosophical account of scientific practice (Giere 1988, 79). Vorms tries to resist this tendency, while holding, as Weisberg does, that external representations are the only means by which scientists can have access to abstract models. But without an explanation of the nature of scientists' access to abstract models, that is, without an explanation for *M2*, Vorms's or Weisberg's accounts ultimately fail to illuminate how manipulation of an external representation helps scientists gain better understanding of the model.

2.7 THE PROBLEM OF IDENTITY

In Section 2.5, I mentioned an important objection to *E1*, which is the claim that models are external representations: it leads to an absurd answer to the identity question, whereas *M1*, which allows models to be abstract entities, does not. I now respond to this objection.

To illustrate the objection, consider Equations (2.1.1) and (2.1.2). By writing $u(\tau) = baX/d$, $v(\tau) = aY/r$, $\tau = rt$, and $\alpha = d/r$, we obtain the following equations (Murray 2002, 80):

$$\frac{du}{d\tau} = u(1 - v) \tag{2.7.1}$$

$$\frac{dv}{d\tau} = \alpha v(u - 1). \tag{2.7.2}$$

Since Equations (2.7.1) and (2.7.2) are the result of substituting new symbols into Equations (2.1.1) and (2.1.2), the two sets of equations are logically equivalent. The objection is that given my account, we have to say that the two sets of equations are two different models because different symbols are used. This cannot be an important difference, so the objection goes, between two models.

This objection, however, ignores other ways in which two sets of equations can be different that we would want to recognize as important. For example, although Equations (2.7.1) and (2.7.2) are logically equivalent to Equations (2.1.1) and (2.1.2), the variables in the former set of equations are defined in such a way that they do not have units (i.e., non-dimensional), while those in the latter do.[18] Thus, if scientists want to compare a model with data on the sizes of predator and prey populations over time, a more useful model is Equations (2.1.1) and (2.1.2). This is one way in which these models differ.

The two models make another difference in the practice of modeling. If scientists do phase plane analysis of Equations (2.7.1) and (2.7.2), their first step is to find the steady state points—$(0, 0)$ and $(1, 1)$ in this case—and draw some flow arrows, and the second step—linearization—is to identify the characteristic of the steady state points (see Murray 2002, 81–82). The cognitive processes, including both internal and bodily processes, involved in these steps will be similar to those in the case of Equations (2.1.1) and (2.1.2). But the cognitive processes involved in the third step—drawing a phase portrait—may be different from those discussed earlier. Recall that in the case of Equations (2.1.1) and (2.1.2), we did not find or even try to find the formulae of the solution curves in Figure 2.5. Rather we finished drawing the solution curves we started in the first step. But Equations (2.7.1) and (2.7.2) enable scientists to deploy a different set of processes in this step. In the u–v plane, scientists can transform Equations (2.7.1) and (2.7.2) into

$$\frac{dv}{du} = \alpha \frac{v(u-1)}{u(1-v)}. \tag{2.7.3}$$

This equation *can* be integrated exactly. In other words, scientists can obtain a formula for the solution curves in the u–v plane:

$$\alpha u + v - \ln u^{\alpha} v = H, \tag{2.7.4}$$

where H is a constant. In the third step of phase plane analysis of Equations (2.7.1) and (2.7.2), scientists can write down a formula for the solution curves. But as Odell says, it is hard to understand the nature of the solutions simply by looking at Equation (2.7.4) alone, so scientists would have to graph Equation (2.7.4) for some values of H to obtain a phase portrait in the u–v plane.[19] The solution curves form closed loops as in Figure 2.5 (see Murray 2002, 81).

Thus, even if we accept *E1*, we can say that Equations (2.7.1) and (2.7.2) are not the same model as Equations (2.1.1) and (2.1.2) because these external representations make difference in

[18] In Equations (2.1.1) and (2.1.2), t has the unit of time, r, a, and d have the unit of per time, and other variables do not have units. To produce Equations (2.7.1) and (2.7.2), we define new variables so that the units cancel out.

[19] For a detailed discussion of the roles of sentential representations like Equation (2.7.4) and diagrammatic representations like a phase portrait in human cognition, see Larkin and Simon (1987).

the practice of modeling.[20] We can say this even though the two sets of equations are logically equivalent. And if two models do not make any practical difference, we can say that they are practically equivalent models.

Underlying this response is the strategy of individuating models by citing differences they make or do not make in the practice of modeling. This strategy helps us individuate other models that we saw in the discussion of phase plane analysis. For example, we can say that the system of Equations (2.1.1) and (2.1.2) and the phase portrait (Figure 2.5) are different models, because the phase portrait enables us to visually understand the nature of the solutions to Equations (2.1.1) and (2.1.2), an understanding that we cannot obtain by looking at the equations.

This strategy, of course, is implicit in Zhang and Norman's view, noted in Section 2.3, that the three versions of TOH are different problems for a task performer although they can be represented as having the same formal structure. The three versions of TOH require a task performer to use different sets of internal and external representations, and thus her internal processes as well as perceptual and bodily activities will differ depending on which version of TOH she does. These differences legitimize Zhang and Norman's individuation of the TOH games. Similarly, my strategy appeals to differences that external representations make in the modeling practice as basis for individuation.

2.8 CONCLUSION

This chapter has focused on two questions regarding mathematical modeling: (i) What is a mathematical model? (ii) What is the practice of modeling? Taking the metaphysics based perspective on models, Giere and Weisberg developed accounts that answer these questions. According to their accounts, mathematical models are abstract, non-physical entities, which are not to be identified with equations and graphs used in mathematical modeling. Moreover, in their accounts, scientists use equations to construct, manipulate, and analyze abstract models. Mathematical modeling is a

[20]For additional cases that support my point, see Vorms (2010, 539–543). Vorms discusses cases in which logically equivalent external representations in physics make differences in scientists' inferential processes.

practice of constructing, manipulating, and analyzing abstract models for the purpose of scientific research.

I have argued that the basic commitments—*M1* and *M2*—of Giere's and Weisberg's accounts obscure rather than illuminate the practice of modeling. I have then developed an alternative account from the cognitive science based perspective on models. In my view, scientific models are external representations, and modeling is a matter of constructing, manipulating, and analyzing external representations in service of cognitive and epistemic aims of research. This account successfully takes into account the insights from research on distributed cognition and gives pride of place to the embodied aspect of the practice of modeling. I have shown how this account illuminates and is reinforced by phase plane analysis, the standard method of analysis of ordinary differential equations models. Finally I have argued that my account can deal with the identity question regarding models.

3.0 USES OF MODELS, CRITERIA OF SUCCESS, AND AIMS OF RESEARCH: A CASE STUDY OF SEWALL WRIGHT'S RESEARCH ON *LINANTHUS PARRYAE*

3.1 INTRODUCTION

In the previous chapter I argued for the cognitive science based perspective on models: a scientific model is an external representation of aspects of the real-world system being investigated. I also suggested that to analyze scientists' uses of models in the ongoing process of inquiry, we focus on the perceptual and bodily interactions between scientists and models. To do so, in this and next chapters, I examine the historical records of scientific research to analyze how scientists used them in their research.[1]

This chapter concerns the relationship among uses of a model, particular aims of research in which scientists use the model, and criteria of success relevant to a given use of the model. The main theses of the chapter are:

1. The relationship between uses of a model and particular aims of research is dynamic: uses of the same model can shift over the course of scientists' research in response to the shift in aim.
2. Criteria of success for one use of a model can be different from those for another use of the same model.[2]

[1]Hereafter I simply write 'model' to mean models understood as external representations.

[2]To forestall a potential misinterpretation of my thesis, I emphasize that what I am going to argue is that a model can have many different uses, that there are certain criteria of success that are relevant to each use, and that the criteria of success for one use may be quite different from those for another use. I will not be claiming that different criteria of success necessarily or often stand in trade-off relations of the kind discussed by philosophers of science following Levins's work (see, e.g., Levins 1966; Orzack and Sober 1993; Levins 1993; Odenbaugh 2003). The uses of a model and relevant criteria of success are important to know, even if the criteria do not stand in any trade-off relations, because such knowledge helps the evaluation of the model in a given context of scientific research and the search for ways to improve the model.

After clarifying the key concepts—a use of a model, a criterion of success, and an aim of research—I argue for these theses using a detailed case study of Sewall Wright's research on a specific plant population (*Linanthus parryae*). The case study reveals that the context of Wright's research radically changed over the course of research between 1941 and 1978, while he continued using the same models developed near the beginning of his research. My analysis of Wright's research shows that he put models to a variety of uses and that some of Wright's uses of models changed in response to the changing aims of his research and that criteria of success for one use of his model were different from—even irrelevant to—another use of the same model.

3.2 USES OF MODELS, CRITERIA OF SUCCESS, AND AIMS OF RESEARCH

In this section, I clarify what I mean by an *aim of research*, a *use of a model*, and *criteria of success*, and explain how these are related. My aim here is to characterize these concepts in such a way that it makes sense to interpret uses of models as part of scientists' activities toward the aims of research.

3.2.1 Aims of Research

When I say aims of research, I am concerned with aims that we would understand as cognitive, epistemic, or practical rather than those that we would understand as social (e.g., to prove oneself to be a legitimate member of a particular community of scientists), political (e.g., to gain more power in the community), or financial (e.g., to secure more grant support). As long as we can identify a given aim of research as cognitive, epistemic, or practical, it is not important for my purpose to decide exactly which sort of aims they are. For example, we can say that to obtain knowledge of some phenomenon is an epistemic aim, but if we also think that to have knowledge is to be in a particular cognitive (mental) state, then we can also say that this is a cognitive aim. To give another example, we can say that to understand a causal process that produces some phenomenon is a cognitive aim because to have an understanding is to be in a certain cognitive state, but if we also think that to understand a causal process is to be able to perform intervention in that process to produce a desired effect, we can also say that it is a practical aim. For my purpose, these ambiguities do not matter as long as a case of research we want to analyze has an aim that can be understood as cognitive, epistemic, or practical.

We can characterize an aim of research with varying details. At the most general end, there are various views about the aims of science. For example, according to Philip Kitcher, one of the most traditional and popular views of science regards science as aimed at "discovering the truth, the whole truth, and nothing but the truth about the world" or, more specifically, "discovering truth about those aspects of nature that impinge most directly upon us, those that we can observe (and, perhaps, hope to control)" (Kitcher 1993, 3). Larry Laudan, to take another example, sees problem solving as the aim of science (in order to explain the development of science) (Laudan 1977, 11–12). Ian Hacking highlights representing (theorizing) and intervening (experimenting) as the aims of science (Hacking 1983, 31). Bas van Fraassen's constructive empiricism holds that science aims at empirically adequate theories (van Fraassen 1980, 12). At the more specific end, there are particular aims of particular research projects. For example, James Watson and Francis Crick's research between 1951 and 1953 aimed at determining the molecular structure of DNA. This was their epistemic aim. To figure out what a given case of research aimed at, we often need historical investigations.

In this chapter, I am interested in the relationship between uses of models and specific, rather than general, aims of research, because this relationship can be dynamic. General aims of science are typically understood as defining features of science; thus, they are not supposed to change over time. But specific aims of research can change over time: scientists can redefine or redirect their research programs. Such change is routine in actual science, and a question arises as to what happens to the ways in which scientists use their models when the aims of research change.

3.2.2 Uses of Models in Science

I have mentioned the aim of Watson and Crick's research, and I will use their research to explain what I mean by uses of a model. I quote passages from Watson's *The Double Helix* where Watson describes how he used models. Here's how Watson recounts his work[3]:

[3]This event took place on February 27 and 28, 1953 in Watson's office, which he shared with Crick and Jerry Donohue (Olby 1994, 410–412; Olby 2009, 165–167; Judson 1996, 148–149).

> Fortunately, when [Crick and I] walked upstairs [in the Cavendish Laboratory], I found that I had an excuse to put off the crucial model-building step for at least several more hours. The metal purine and pyrimidine models, needed for systematically checking all the conceivable hydrogen-bonding possibilities, had not been finished on time. At least two more days were needed before they would be in our hands. This was much too long even for me to remain in limbo, so I spent the rest of the afternoon cutting accurate representations of the bases out of stiff cardboard. ...
>
> When I got to our still empty office the following morning [on February 28], I quickly cleared away the papers from my desk top [sic] so that I would have a large, flat surface on which to form pairs of bases held together by hydrogen bonds. Though I initially went back to my like-with-like prejudices [i.e., each base bonds with another base of the like kind], I saw all too well that they led nowhere. When Jerry [Donohue] came in I looked up, saw that it was not Francis, and began shifting the bases in and out of various other pairing possibilities. Suddenly I became aware that an adenine-thymine pair held together by two hydrogen bonds was identical in shape to a guanine-cytosine pair held together by at least two hydrogen bonds. All the hydrogen bonds seemed to form naturally; no fudging was required to make the two types of base pairs identical in shape. Quickly I called Jerry over to ask him whether this time he had any objection to my new base pairs.
>
> When he said no, my morale skyrocketed, for I suspected that we now had the answer to the riddle of why the number of purine residues [i.e., adenine and guanine] exactly equaled the number of pyrimidine residues [i.e., cytosine and thymine]. (Watson 1968, 114; see also Olby 1994, 411–412; Judson 1996, 148)

The next morning Watson saw Crick in the lab, "flipping the cardboard base pairs about an imaginary line. As far as a compass and ruler could tell him, both sets of base pairs neatly fitted into the backbone configuration" (Watson 1968, 117). Soon after, in the first week of March 1953, the metal models were delivered, and Watson and Crick began holding the metal pieces together by clamps and rods and eventually built the double helix model on the office table. Next they put a plumb line and a measuring rod against the model and read off the coordinates of atoms making up nucleotides (Watson 1968, 117–118). During the following weeks, they showed the model to other people, including Lawrence Bragg, Maurice Wilkins, Rosalind Franklin, and Alexander Todd, letting them inspect various aspects of the model (Watson 1968, 118–127).

What was the *use* of Watson's cardboard models of bases? We might say that Watson used the models to find the plausible patterns of base pairs in DNA. This answer, however, is too general for my purpose. The aim of Watson and Crick's research was to determine the structure of DNA, and part of this aim was to find the plausible base paring patterns. So the answer we just gave simply says that the use of the cardboard models was to achieve part of the aim of Watson and Crick's research. A more adequate answer needs to say *how* Watson used the models or *what activity* the models enabled him to do that helped him find the correct base paring patterns. This activity that scientists do with a model is what I mean by a *use* of a model. Thus, I would say that the use of the cardboard models was to systematically check the possible patterns of hydrogen bonding among four nucleotide bases (adenine, guanine, cytosine, and thymine). Moreover, Watson and Crick used the models to see if the base pairs would fit into the backbone, and to do that, they took measurements by applying a compass and rule to the cardboard models. Note that checking all the hydrogen-bonding possibilities for their plausibility is not the same as finding the plausible base paring patterns, but it is an activity that could (and did) help Watson find them.

Let us also consider the use of the double helix model, the mockup built from metal plates and rods. Again it would be too vague to say that the use of the double helix model was to determine the three-dimensional structure of DNA—i.e., to achieve the aim of their research. It does not tell us what activity Watson and Crick did with the models that helped them achieve their aim. I would say instead that the model had at least three uses. First, Watson and Crick used it to visualize a structure that would satisfy their assumptions about the geometry and arrangement of nucleotides and the sugar-phosphate backbone. Second, they used the model to take precise measurements of the atomic coordinates of the nucleotides by putting a plumb line and a measuring rod against the double helix model and reading off the atomic coordinates. Third, Watson and Crick used the model to enable other experts, such as Wilkins and Franklin, to visually inspect their proposed structure. None of these uses by itself amounts to the achievement of the aim of Watson and Crick's research, but together these uses advanced their aim.[4]

[4]Other steps needed to accomplish their aim included detailed comparison with experimental evidence (see, e.g.,Franklin and Gosling 1953a,b; Wilkins et al. 1953; Watson and Crick 1953b).

In short, when a scientist uses a model in her research, she uses it to perform some action that would promote the aim of her research. I refer to this action as a use of a model, and the brief discussion of Watson and Crick's research shows that a use of a model does not always amount to the achievement of the aim of research. In such a case, it is important to distinguish uses of a model from the aim of research in which the model is used so that we can be clear about what scientists do with the model in order to advance the aim of their research.

3.2.3 Criteria of Success

By criteria of success, I mean criteria according to which scientists can evaluate how successful their *use* of a model is. Some criteria apply to properties of a model itself, when the properties affect the outcome of a use of the model. Other criteria apply to the properties of the outcome.

Let me illustrate these points with the Watson-Crick example. Recall that Watson used the cardboard models of the bases to systematically check all the conceivable hydrogen-bonding possibilities. In Watson's account quoted above, one criterion of success relevant for this use of the cardboard models was how stereochemically accurate they were, because Watson says that he purposefully designed them to be stereochemically accurate. This criterion applies to the spatial properties of the models themselves, and this was a relevant criterion because spatial properties of each model affect the spatial configurations they form when Watson moved them about to check bonding possibilities. Recall also that one of Watson and Crick's uses of the double helix model was to take precise measurements of the atomic coordinates of the nucleotides with a plumb line and a measuring rod. For this use, stereochemical accuracy just mentioned was relevant, but we can think of other relevant criteria. First, the metal plates and rods making up the double helix model had to be rigid so that the model does not collapse while Watson and Crick take measurements. Second, the parts of the double helix model had to be adequately spaced so that Watson and Crick can access all parts by hand and take measurements. Note that this last criterion had to be met without violating the criterion of stereochemical accuracy. Both of these additional criteria apply to the properties of the model, because if the model met these criteria as well as the stereochemical accuracy criterion, Watson and Crick would be able to successfully use the model to take precise measurements. Finally, given the aim of their research, it was reasonable to require as a criterion of success that the measurements taken from the double helix model be consistent with the experimental evidence. This criterion applies to the outcome of the use of the model.

3.3 DYNAMICS OF RESEARCH: A CASE STUDY

The central theses of this chapter are that criteria of success for one use of a model can be different from those for another use of the same model and that the relationship between uses of a model and particular aims of research is dynamic: uses of the model can shift over the course of scientists' research in response to the shift in aim.

Scientific research can take many different trajectories. For example, consider a scientist working on her side project. She develops a model during her work on this project, but the project turns out to be more important than she initially imagined. Over time the project grew to one of her main projects, and along the way her specific research questions changed. In this case, we might find that the same model was put to different uses at different phases of her project; we might also find that the model was put to the same use throughout the project.[5] At any given point in the project, the model might or might not have met the criteria of success for one or more of uses.

To analyze the dynamics of research, it is necessary to follow the development of research over time. To do this, I provide a detailed historical analysis of Wright's research. It turns out that his research took a trajectory like the one just described.

Here is an outline of the case study: In Section 3.4, I provide background to Wright's research and identify three temporal phases of his project. I will then devote one section to each phase (Sections 3.5–3.7). Each section presents a historical narrative followed by an analysis of the aims of Wright's research, his uses of models, and criteria of success.

[5] Other combinations are of course possible.

3.4 BACKGROUND TO THE *LINANTHUS* RESEARCH

3.4.1 *Linanthus* Facts

Linanthus parryae is a diminutive desert annual in the Mojave Desert in California. It has blue and white flower color morphs, the former being dominant to the latter (Epling et al. 1960, 238). It is pollinated exclusively by a species of soft-winged flower beetles, whose flight distance is one to ten feet, and seeds are dispersed passively (Epling et al. 1960, 240, 243; Schemske and Bierzychudek 2001, 1270).

The life cycle of *Linanthus* shows two patterns. In wet years, when there is enough rainfall in winter, seed germination occurs, and plants flower in early to late April, shedding seeds in late May to early June. In dry years, no seed germination occurs although seeds can remain dormant in the soil for seven years or longer (Epling et al. 1960, 240, 250; Schemske and Bierzychudek 2001, 1270). In a favorable wet year, thousands of plants bloom and cover the desert as if snow has fallen: hence the common name "desert snow" (Epling and Dobzhansky 1942, 318).

3.4.2 1941 Survey

In April 1941, the population of 10 to 100 billion blooming *Linanthus* plants covered an 840-square-mile region of the Mojave Desert (Epling and Dobzhansky 1942, 329–330; Wright 1943a, 141). The distribution of flower color exhibited interesting patterns. Overall white flowers were most abundant, and in some areas there were only white flowers. But in three separate areas–referred to as the "variable areas" (Epling and Dobzhansky 1942, 323)—blue and white flowers coexisted. There were no obvious geographical barriers that might have been responsible for these patterns.

Carl Epling and his students did an extensive survey of this population. They created a station every half-mile along the roads forming a rough grid in the desert and ran, at each station, a transect at approximately right angles to the road. Along each transect, they made four equally spaced sampling points, at each of which, when there were plants, they counted plants until 100 and recorded the numbers of blue and white. Epling and his students made 1261 sampling points, counting a total of 113,955 whites and 12,145 blues (Epling and Dobzhansky 1942, 322–325).

Station Number	Numbers of Blues			
14	39	58	57	52
19	0	0	0	0
20	–	86	72	4

Table 3.1: Three rows taken from Epling and Dobzhansky's table. Redrawn from Epling and Dobzhansky 1942, 324.

Between May and October 1941 Epling and Dobzhansky analyzed the data and prepared a manuscript on *Linanthus*. In it they presented "a condensed summary of the data" (Epling and Dobzhansky 1942, 323), using a table and a map. Together the table and the map summarize the location of each sampling point and the proportion of blue and white recorded at each point. The rows of the table looked like this (see Table 3.1). According to Epling and Dobzhansky, the table is to be read as follows:

> The numbers of the blue flowered plants found in the four samples at each station are indicated consecutively from left to right [in a row of the table]. Thus, the numbers 39, 58, 57, and 52 at station No. 14 mean that 39 blue flowered plants were found in the sample which was taken 70 paces on the left of the road, 58 in that 20 paces on the left, 57 in that 20 paces on the right, and 52 in that 70 paces on the right of the road. The corresponding numbers of the white flowered plants found, but not cited, were hence, 61, 42, 43, and 48. Zero means that no blue flowered plants were found in a given sample; the sign − means that no plants at all were encountered. (Epling and Dobzhansky 1942, 323)

Epling and Dobzhansky gave the location of station No. 14 in the accompanying map (Epling and Dobzhansky 1942, 320–321). The number 39 in the first row of Table 3.1 means that 39 blue flowered plants were found in the sample of a hundred plants counted at 70 paces on the left of the location marked 14 on the map.[6]

[6]It is a bit unclear which is the "left" of the road because there is no explicit mention of the direction of the road, but if we assume that the direction is given by the way the number 14 is written on the map, then the left of station No. 14 is the south of the road (see Figure 3.1).

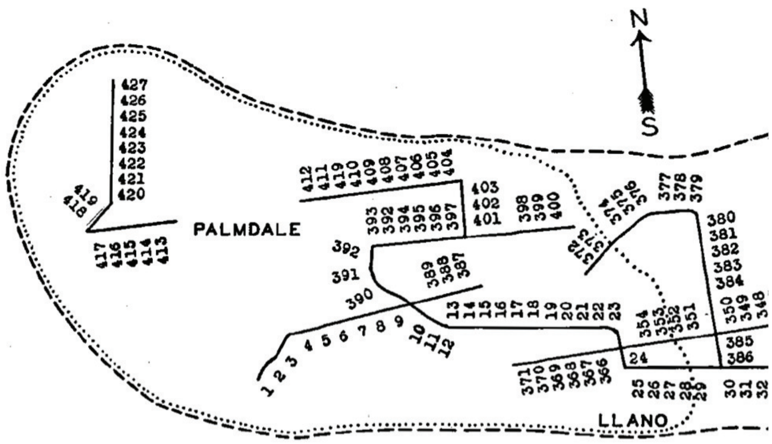


Figure 3.1: A portion of Epling and Dobzhansky's map of stations. Solid lines indicate the roads traveled during the survey. The broken line marks the geographical limit of the occurrence of *Linanthus*. Inside the dotted line is the variable area, where blue and white flowers coexisted. Reproduced from Epling and Dobzhansky (1942, 320) with permission from the Genetics Society of America.

Epling and Dobzhansky's map and table were "a condensed summary" of the raw data in two respects. First, Epling and Dobzhansky omitted the data on stations where no plants were observed, so stations No. 15 through 18, which appear in Figure 3.1, are skipped in Table 3.1. Second, they also omitted the numbers of whites although these numbers can easily be inferred from those of blues given in the table.

3.4.3 Three Phases

With hindsight and with the surviving records of Wright's work, we can identify three major phases of Wright's research on *Linanthus*.[7] The first phase started when Wright was given the *Linanthus* data in 1941 and ended with Wright's paper on *Linanthus*, completed in 1942 and published in 1943. The second phase occurred between 1960 and 1962 when Wright wrote two unpublished manuscripts on *Linanthus* in response to Epling and colleagues' new paper, written in 1959 and

[7]For the basic history of the *Linanthus* research, see Provine (1986, 370–381, 484–488).

published in 1960. The third phase occurred between 1972 and 1978 when Wright reanalyzed the original *Linanthus* data and published the new analysis as part of his 1978 book.

3.5 FIRST PHASE: ISOLATION BY DISTANCE IN *LINANTHUS*

3.5.1 Wright's 1941 Letter

In September and October 1941, Dobzhansky gave the *Linanthus* data, the map of some 400 stations, and the manuscript to his then collaborator Sewall Wright (Dobzhansky to Wright, October 17, 1941; Dobzhansky to Wright, October 30, 1941).[8]

In their manuscript Epling and Dobzhansky showed that the statistical distribution of samples containing various proportions of blues resembled the U-shaped distribution of gene frequencies, which, according to Wright (1931, 122–128), was expected if the effective population size and mutation and migration rates were so small that change in gene frequency in each generation was dominated by stochastic factors, such as genetic drift. Thus, Epling and Dobzhansky argued that effective population size is quite small in *Linanthus*, thereby implying that effects of random evolutionary factors are not negligible in this population (Dobzhansky to Wright, October 17, 1941; Dobzhansky to Wright, October 30, 1941; Epling and Dobzhansky 1942, 331–332).

In November 1941 Wright wrote a detailed response to Dobzhansky, which began as follows:

> I have gone over the manuscript on *Linanthus* that you sent me. It is certainly a very interesting case. It appears to be a good example of isolation by distance that I discussed in my symposium paper at the Columbus meeting (p. 244–246) (Wright to Dobzhansky, November 1941).

The "symposium paper" refers to Wright's paper on "Breeding Structure of Populations in Relation to Speciation," which Wright delivered at a symposium on speciation at the meeting of the American Association for the Advancement of Science on December 28, 1939. The paper was published

[8] Wright's correspondence cited in this chapter is available in Sewall Wright Papers, Series I, American Philosophical Society, and it will be cited by the correspondent and date. By the time Dobzhansky sent the *Linanthus* data to Wright, they had already published a paper in Dobzhansky's *Genetics of Natural Populations* series (Dobzhansky and Wright 1941) and were working on another paper to be published in 1942 (Wright et al. 1942). For a fine discussion of the history of the collaboration between Dobzhansky and Wright, see Provine (1986, chs. 10 and 11).

in the next year, and the page numbers given in the quote refers to the section on isolation by distance in the published version (see Wright 1940, 244–246).

In the rest of the letter Wright presented his own analysis of the *Linanthus* data given in Epling and Dobzhansky's manuscript. To analyze the data, Wright immediately faced an obstacle. Epling and Dobzhansky's table and map of data summarized the geographical distribution of *phenotype* frequencies in the *Linanthus* population. If the mode of inheritance of the flower color (e.g., whether blue is dominant and white is recessive) were known via breeding experiments, it would have been relatively easy to infer gene or genotype frequencies from the table. However, in 1941, Epling was not able to germinate seeds of *Linanthus* and determine the mode of inheritance of the flower color (Epling and Dobzhansky 1942, 332, 332; Dobzhansky to Wright, October 3, 1941; Dobzhansky to Wright, October 30, 1941).[9]

Proceeding without the knowledge of the mode of inheritance, Wright made different assumptions about which flower color is dominant and converted the distribution of phenotype frequencies into that of gene frequencies. Under the assumption that blue was recessive and that mating was random within each sample, according to Wright, the frequency of the recessive gene in a sample is a square root of the frequency of blue phenotype in the sample (Wright to Dobzhansky, November 1941).[10] Thus, Wright converted the phenotype frequency of 0.01 in a sample (i.e., 1% of a sample was blue) into the gene frequency of 0.1 in the sample, the phenotype frequency of 0.02 into the gene frequency of 0.141, and so on. In this way, Wright converted the phenotype frequency of each sample recorded in Table 3.1 into the gene frequency.

Once he had the data in terms of gene frequency, Wright calculated the mean frequency $\bar{q}$ of a gene for blue and standard deviation σ_q of the frequencies of q in samples that characterized this converted data under two different hypotheses about the mode of inheritance. He presented the results in the first two rows of Table 3.2 (Wright to Dobzhansky, November 1941).[11] From $\bar{q}$ and σ_q, Wright calculated the product Nm of population size N and migration rate m. But since no separate measurement of N or m was available, Wright had to constrain the value of at least one

[9] It would take Epling twenty years to determine the mode of flower color inheritance. Blue turned out to be dominant to white (Epling et al. 1960, 238).

[10] Let p be the phenotype frequency of blue in a sample and q the frequency of the gene for blue. Under the given assumptions, the genotype of blue-flowered plants is homozygous recessive, and its frequency in a sample should be q^2 and equal to p, because there is only one genotype that is associated with blue phenotype. Hence, q is a square root of p.

[11] Wright would try other assumptions in his published analysis (Wright 1943a).

Blue Recessive	White Recessive
$\bar{q} = .1291$	$\bar{q} = .0718$ (blues give)
$\sigma_q = .2823$	$\sigma_q = .2056$
$\frac{\sigma_q}{\sqrt{\bar{q}(1-\bar{q})}} = .8420$	$\frac{\sigma_q}{\sqrt{\bar{q}(1-\bar{q})}} = .7962$
$Nm = \frac{1}{4}[\frac{\bar{q}(1-\bar{q})}{\sigma_q} - 1] = .1026$	$Nm = .1443$
If $N = 100$, $m = .001026$	*If* $N = 100$, $m = .001443$
If $N = 10$, $m = .01026$	*If* $N = 10$, $m = .01443$

Table 3.2: Wright's 1941 analysis of the *Linanthus* data. $\bar{q}$ is the mean gene frequency in the total population, σ_q the standard deviation in the gene frequency q among smaller territories within the total population, N the effective population size, and m the migration rate. The expression $\sigma_q/\sqrt{\bar{q}(1-\bar{q})}$ is the square root of a statistic Wright would later call F_{IS}. Redrawn from Wright to Dobzhansky, November 1941 with permission from the American Philosophical Society.

of these parameters and calculate the value of the other. The bottom two rows of Table 3.2, Wright considered two possible values of N, 10 and 100, and calculated the corresponding values of m. These values of N refer to the effective size of a local population of breeding individuals rather than the total population size, which was estimated to be 10 to 100 billion.

When Wright assumed the effective population size of 10 or 100 individuals, he was working with the idea of isolation by distance. In the Symposium paper cited at the beginning of his letter, Wright discussed, under the heading "Isolation by Distance," a "model of breeding structure which may be of considerable importance" (Wright 1940, 244–245). He said:

> Suppose that a population is distributed uniformly over a large territory but that the parents of any given individual are drawn from a small surrounding region (average distance D, effective population N). How much local differentiation is possible merely from accidents of sampling? Obviously the grandparents were drawn from a larger territory [defined in terms of D and N]. The ancesters [sic] of generation n came from an average distance $\sqrt{n}D$ and from a population of average size nN. . . . If the parents are drawn from local populations of effective size greater than 1,000, the situation differs little from panmixia even over enormous areas. There is considerable fluctuating local

> differentiation of unit territories where their effective size is of the order of 100, but not much differentiation of large regions unless effective N is much less. (Wright 1940, 245)

There are two key assumptions here: (1) Individuals in a population are uniformly distributed over a large territory, and (2) individuals disperse over relatively short distances from their birthplaces to locations where they produce offspring. As long as the geographical range of the population is greater than a region from which parents of a given individual are drawn, there is a "unit" population whose effective size is smaller than the total population size. In this Symposium paper and an earlier 1938 note, Wright showed that if effective size of unit populations is small, say, 100 or less, frequencies of a gene in unit populations can exhibit considerable random fluctuations (e.g., a gene may be fixed in some unit populations, lost in some, and in intermediate frequencies in others) (Wright 1938, 1940).

Making assumptions (1) and (2) for the *Linanthus* population, Wright inferred that the total *Linanthus* population was not a single population of billions of randomly mating individuals but was composed of numerous, small local populations that were more or less isolated merely by distance. Hence, he considered the effective population size of 10 and 100. Assumptions (1) and (2) were reasonable for the population of insect-pollinated plants distributed over 840 sq mi. In addition, Wright assumed that these local populations have the same effective population size N. This assumption can be called the spatial homogeneity assumption, and at least in the early 1940s, Wright made this assumption primarily because of its convenience, and he was aware that in reality the assumption was almost certainly false (Dobzhansky and Wright 1941, 35).[12]

Wright then calculated the distribution of gene frequencies without selection and mutation by substituting the parameter values given in Table 3.2 into the following equation (Wright to Dobzhansky, November 1941):

[12]The assumption is actually more general and takes other parameters of interest, such as selection coefficient and migration rate, to be spatially homogenous as well. That is, any parameter values in Equation (3.5.1) and the like are constant throughout the geographical region occupied by the population in question.

Blue (%)	Observed Number	Calculated (blue recessive)	Diff.	Calculated (blue dominant)	Diff.
0	987	982.6	+4.4	947.4	+39.6
1–10	68	88.6	-20.6	118.2	-50.2
10–20	28.5	26.0	+2.5	32.9	-4.4
20–30	19.5	17.6	+1.9	20.9	-1.4
30–40	17.5	14.4	+3.1	16.2	+1.3
40–50	23.5	13.0	+10.5	13.9	+9.6
50–60	28	12.5	+15.5	12.8	+15.2
60–70	20	12.7	+7.3	12.5	+7.5
70–80	10.5	14.0	-3.5	13.1	-2.6
80–90	8	17.9	-9.9	15.9	-7.9
90–99	16.5	37.1	-20.6	33.1	-16.6
100	34	24.6	+9.4	24.1	+9.9
Total	1261	1261		1261	

Table 3.3: Wright's comparison of the theoretical distribution of gene frequencies with the *Linanthus* data. Redrawn with modified labels from Wright to Dobzhansky, November 1941 with permission from the American Philosophical Society.

$$\varphi(\bar{q}) = C\bar{q}^{4Nm\bar{q}-1}(1-\bar{q})^{4Nm(1-\bar{q})-1}. \tag{3.5.1}$$

The calculated distribution of gene frequencies, however, could not be directly compared with Epling and Dobzhansky's data on *Linanthus*. The unit of the distribution was the mean gene frequencies $\bar{q}$, but the unit of Epling and Dobzhansky's data was phenotype frequencies.[13]

To overcome this problem, Wright did two things. First, he constructed a frequency distribution of number of samples in Epling and Dobzhansky's table. One dimension of this distribution was percentages of blue divided into 12 classes, and the other dimension was the number of samples out of 1261 whose percentages of blue correspond to each class.[14] He displayed it as what statisticians call a contingency table (the left two columns of Table 3.3). Second, Wright converted the calculated distribution of gene frequencies into the distribution with the same unit, domain, and

[13]Since the distribution of gene frequencies is supposed to be a probability distribution, the area under the curve must be 1, the condition guaranteed by the coefficient C in Equation (3.5.1).

[14]If this array is presented by a two dimensional graph, the horizontal axis will be percentages of blue and the vertical axis the number of samples having a given percentage of blue.

range as the two-dimensional array. Wright referred to this conversion process as "retransforming the scale [of the calculated distribution of gene frequencies] to phenotypes and groupings" (Wright to Dobzhansky, November 1941). He displayed the converted distributions in the third and the fifth columns from the left in Table 3.3 as well as the difference between the observed and calculated distributions.

Referring to Table 3.3, Wright argued that the fit is not good "because of the hump at 40% to 70%, the deficiencies at 1 to 10%, 80–99% and the excess at 100%" (Wright to Dobzhansky, November 1941). However, he suggested that adding a slight selection for the heterozygotes to the above equation would make the theoretical distribution exhibit a hump similar to the observed (Wright to Dobzhansky, November 1941). He told Dobzhansky:

> Such selection would account for the persistence of both blue and white in nature. Minute differences in mutation rates in the two directions or in selection of the homozygotes could account for the value of $\bar{q}$. The variability of gene frequency would still be due to isolation by distance. (Wright to Dobzhansky, November 1941)

In other words, in *Linanthus*, σ_q had the value it did because local populations were isolated by distance and ended up having widely different gene frequencies.

After presenting this argument, Wright gave another argument for the idea that effective population size was small (about 10) in the *Linanthus* population. For this argument, he presented a new method of data analysis developed "shortly after" (Wright to Dobzhansky, November 1941) he wrote the Symposium paper. According to Wright, the new method was "much more useful" (Wright to Dobzhansky, November 1941) than the one given in the Symposium paper, because whereas the latter required the knowledge of effective population size (or the size of the unit population) to begin with, the new method did not.[15] To present this method, Wright drew a diagram (Figure 3.2) and wrote:

> Let N_t be the size of the total population and suppose this to be subdivided into K territories of effective size $N_i = N_t/K$ and these into territories of size N_u within which there is random mating. Then the amount of variability of gene frequency calculated for the territories of size N_i is a function of the unknown size of the random breeding

[15]Thus, the previous method was "not of much use" (Wright to Dobzhansky, November 1941) for the purpose of inferring the effective population size.

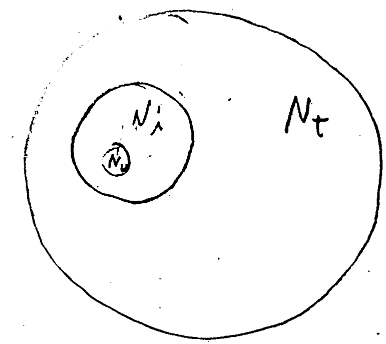


Figure 3.2: Wright's diagram of hierarchical population structure. N_t is the size of the total population, and N_u is the size of a unit population. N_i is the size of an intermediate population within the total and containing certain number of unit populations. Reproduced from Wright to Dobzhansky, November 1941 with permission from the American Philosophical Society.

> unit (actually the number of individuals from which the mate of any single individual is drawn). Curiously enough Figure 3 [in the Symposium paper; see Figure 3.3] will practically do for the case of area continuity [like the *Linanthus* case]. ...(Wright to Dobzhansky, November 1941)

Figure 3 of the Symposium paper (Figure 3.3 in this chapter) correlated effective population sizes with values of the variability in gene frequency, measured by $\sigma_q/\sqrt{\bar{q}(1-\bar{q})}$. Following the above passage, Wright argued that given plausible assumptions about the average distances among the samples of each station and about the average number of possible samples within the area of a station, the variability of gene frequencies in *Linanthus* will be the value given in the third row of Table 3.2. Furthermore, given the assumptions, "generations" on the horizontal axis in Figure 3 of the Symposium paper would be on the order of 10^8 (Wright to Dobzhansky, November 1941). Thus, in the figure Wright looked at the solid curves (i.e., area) (since the *Linanthus* population occupies a continuous area rather than a linear range) and extrapolated them into the 10^8 region lying to the right of the region actually displayed in the figure. He could then locate the value of the variability in the x-axis and look to the right to see if any curve has or comes near that value. Take, for example, $\sigma_q/\sqrt{\bar{q}(1-\bar{q})} = .8420$. Wright was able to see that the solid line for $N = 10$

would reach this value if the line were extrapolated into the 10^8 region. He thus concluded: "*N* of the random breeding unit [i.e., N_u] is about 10" (Wright to Dobzhansky, November 1941).

3.5.2 Wright's 1943 Papers

On November 5, 1942, Wright sent to Dobzhansky two manuscripts. The first paper, entitled "Isolation by Distance," according to Wright, "has been around in one form or another at least since 1939 when I based my Symposium paper in part on it" (Wright to Dobzhansky, November 5, 1942). The second paper was on *Linanthus*. Write said:

> The other manuscript is an account of an analysis which I made of the data which you and Epling published on *Linanthus Parryae*. In view of your statement in a recent letter [probably a reference to: Dobzhansky to Wright, October 16, 1942] that Epling is continuing work on this, it becomes a question what I should do with it. There are undoubtedly parts which should be modified in the light of his results although not necessarily very much since the primary purpose is methodological, merely using *Linanthus* as illustrative material. (Wright to Dobzhansky, November 5, 1942)

Wright intended the *Linanthus* paper to be an illustration of the methods developed in the Isolation paper. Here I highlight the aspects of these two papers that will become important in my analysis of Wright's uses of models in his research.

In "Isolation by Distance," Wright showed that isolation by distance has significant evolutionary consequences. For example, heterozygosity will decrease in a population under isolation by distance because local individuals become more and more genetically related and thus tend to be homozygous at a given locus. The degree of inbreeding increases in a population under isolation by distance. To demonstrate these consequences, Wright adopted the inbreeding coefficient F, which he introduced in 1921 (Wright 1921, 118), as a quantitative measure of the effect of population structure on gene and genotype frequencies, especially local genetic differentiation.

The quantity F is defined as follows (Wright 1943b, 122):

$$F = \frac{\sigma_x}{\sqrt{q_y(1-q_y)}} \tag{3.5.2}$$

where σ_x is the standard deviation of the frequencies of a gene of interest (say, A) among subgroups x's in a population. The total population or any subpopulation that contains x's is referred to as

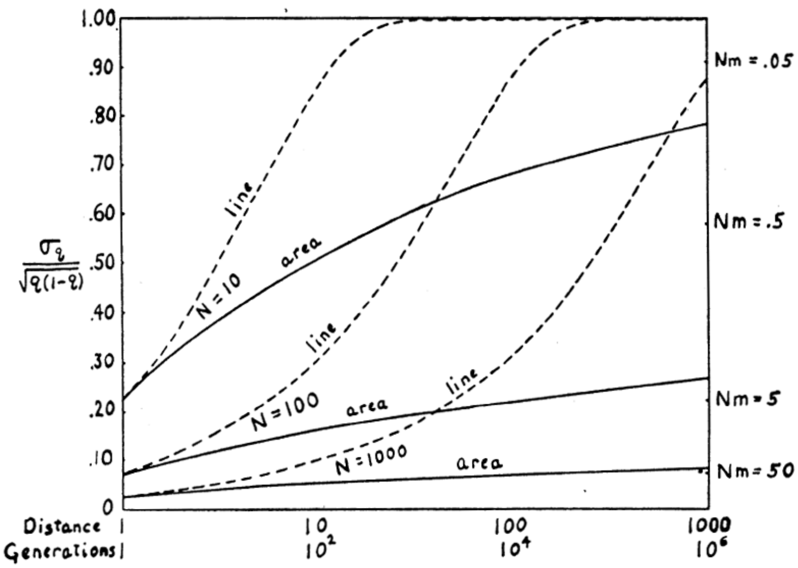


FIG. 3. The standard deviation of the mean gene frequencies of unit random breeding territories (N = 10; N = 100; N = 1000), in relation to mean distance. The case in which the population is distributed uniformly over an area is represented in solid lines, that in which it is distributed along one dimension by broken lines.

Figure 3.3: The relationship between effective population size N and the amount of variability in gene frequency $\sigma_q/\sqrt{\bar{q}(1-\bar{q})}$. The curves marked "area" are relevant to the *Linanthus* population. Wright's own caption is included. Reproduced from Wright (1940, 246) with permission from the American Society of Naturalists and the University of Chicago Press.

population y. q_y is the frequency of A in y. $1 - q_y$ is the frequency of the other allele (say, B) at the same locus. $\sqrt{q_y(1-q_y)}$ gives the limiting value of variability of genes in y: If a gene of interest is fixed or lost in every x, then the distribution of gene frequencies in y has the greatest spread (x's are completely differentiated). And the standard deviation of such distribution is equal to $\sqrt{q_y(1-q_y)}$, for any given value of q_y. If a gene is fixed in half of x's and lost in the other half, then $q_y = 0.5$ and the distribution of gene frequencies in y has the greatest possible value among all possible values of q_y. In any case, $F = 1$ means that x's are maximally differentiated for the given value of q_y.

An investigator can devise a hierarchy of F values. For example, in Figure 3.2, unit populations u's would be the x's, and the subpopulation i or the total population t would be the y. Or subpopulations i's can be taken to be the x's and the total population t be the y. To avoid confusion, as Wright would later do (Wright 1951, 1965), we can put subscripts to F, like F_{xy}, and substitute appropriate symbols for x and y. In the example just given, we have F_{ui}, F_{ut}, and F_{it}, which, in Wright's later notations, correspond to F_{IS}, F_{DT} (= F_{IT}), and F_{ST}, respectively.[16] These notations are meant to be suggestive; T stands for a total population, S for subpopulations, D for demes regarded as unit populations, and I for individuals (Wright 1965, 401). According to Wright, $F_{ST} = (F_{IT} - F_{IS})/(1 - F_{IS})$ (Wright 1965, 402).

Wright showed two results concerning F_{IS} and F_{ST} that became relevant to his analysis of the *Linanthus* data. First, for a given value of N_u, F_{IS} increases as the number of unit populations within a subpopulation increases or as the area occupied by a subpopulation increases. And for a given number of unit populations within a subpopulation, F_{IS} increases as N_u decreases (Figure 3.4). Second, for a given value of N_u, F_{ST} decreases as N_i increases, and for a given value of N_i, F_{ST} increases as N_u decreases (Figure 3.5). Greater F values mean greater amount of local genetic differentiation: unit populations have widely different frequencies of a given gene, including fixation in one unit population and loss in another. In particular, Wright said that if the gene frequency in the total population is 0.5, which gives the greatest value of the denominator in Equation (3.5.2), the F_{IS} value of greater than 0.577 means that the distribution of gene frequencies

[16]Technically, F_{DT} is equivalent to F_{IT} if there is, as usually assumed, random mating within the unit population (Wright 1965, 401). In the later notations, I stands for individual, D for deme, S for subpopulation, and T for total population.

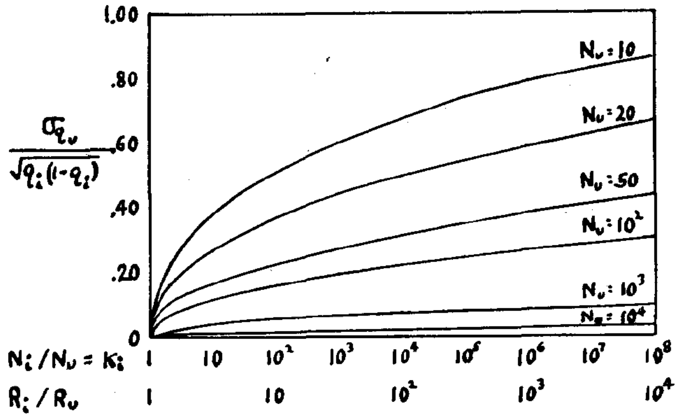


Figure 3.4: Theoretical curves of F_{IS}. F_{IS} (the vertical axis) is defined according to Equation (3.5.2). K_i is the number of unit populations of size N_u within a subpopulation of size N_i. Reproduced from Wright (1943b, 122) with permission from the Genetics Society of America.

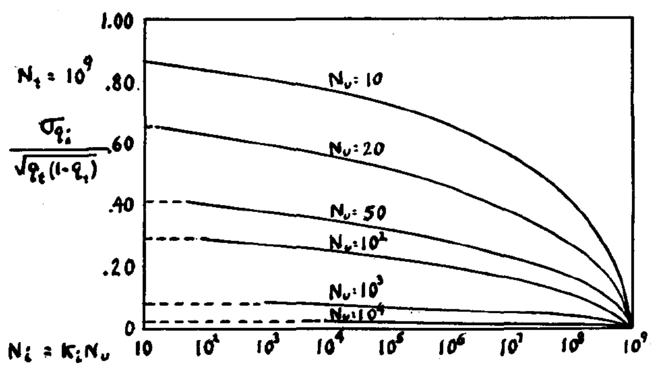


Figure 3.5: Theoretical curves of F_{ST}. F_{ST} (the vertical axis) is defined according to Equation (3.5.2). The size N_i of the subpopulation under consideration (the horizontal axis) is equal to K unit populations of size N_u. The total population N_t is constant. Reproduced from Wright (1943b, 122) with permission from the Genetics Society of America.

Smaller Subdivisions	Primary Subdivisions (East to West) I	II	III	IV	V	VI	Total
Secondary	5	5	5	5	5	5	30
Tertiary	20	20	20	20	20	20	120
Stations	57	59	60	60	61	59	356
Samples	198	211	214	214	218	203	1258

Table 3.4: Wright's hierarchy of subdivisions in the *Linanthus* data. Redrawn from Wright (1943a, 140) with permission from the Genetics Society of America.

among unit populations is U-shaped (Wright 1943b, 123–124). That is, in many unit populations, the gene is fixed, and in equally many unit populations, it is lost.

Wright had been working on the paper on isolation by distance at least since 1939 (Wright to Dobzhansky, November 5, 1942; for earlier publications, see Wright 1938, 1940). By the summer of 1942 Wright was analyzing the *Linanthus* data, using the F values he described above. He published the results in the companion paper to "Isolation by Distance." In this paper, Wright analyzed the 1941 *Linanthus* data in terms of isolation by distance and the method of analysis based on F_{IS} and F_{ST}. In addition to the hypotheses that blue is dominant and that it is recessive, he also considered the possibilities of self-fertilization and no dominance.

As in 1941, the spatial homogeneity assumption—the assumption that N_u is constant throughout the region occupied by the total population of *Linanthus*—underlay all aspects of Wright's analysis of the *Linanthus* data. Wright's hierarchical division of the *Linanthus* population which he used throughout his analysis was based on the spatial homogeneity assumption. As shown in Table 3.4, Wright divided the total population into six primary subdivisions, each of which is divided into five secondary subdivisions. Each secondary subdivision is divided into 20 tertiary subdivisions, each of which contains about 60 sampling stations. Each station contains four samples (100 plants each) (Wright 1943a, 140). Each station is assumed to contain 200 random breeding units.

Wright also constructed a map of the primary and secondary subdivisions (Figure 3.6). He apparently wanted to make these divisions as uniform as possible, and in the paper, he only says, "For more detailed mathematical analysis, it is convenient to define a hierarchy of subdivisions" (Wright 1943a, 140). Of course, uniform subdivisions would simplify computations and compar-

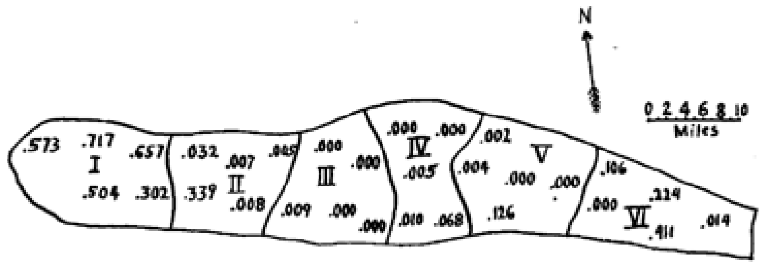


Figure 3.6: Wright's map of the distribution of flower colors in *Linanthus*. Frequencies of blue in primary (I–VI) and secondary subdivisions are shown. Reproduced from Wright (1943a, 149) with permission from the Genetics Society of America.

ison of the results, but if convenience were the only reason, he could be criticized for arbitrarily dividing the population and hence the data. But we can see that his assumption of the spatial homogeneity of parameter values also supported uniform subdivisions. Under this assumption, each random breeding unit is subject to the same conditions as any other, and the same must hold for any population in the hierarchy of subdivisions. Thus, given the spatial homogeneity assumption, it was reasonable for Wright to divide the *Linanthus* population uniformly.

Obviously there are many ways to divide the population uniformly, so it seems impossible to eliminate an element of arbitrariness from the way in which the hierarchy of subdivisions was created. The historical question that we can address is how Wright actually—arbitrarily or not—created the map in Figure 3.6 from the map he received from Dobzhansky, which was like Figure 3.1. Consider Wright's drafts of the map (Figure 3.7). I interpret Figure 3.7a to be earlier than Figure 3.7b for two reasons: First, Figure 3.7a shows the roads along which Epling created sampling stations, and an area of the roads is circled to indicate a secondary subdivision. Compared to this map, Figure 3.7b is more simplified and similar to Figure 3.6. Second, only in Figure 3.7b does Wright directly refer to the manuscript (e.g., "A in paper," "B in paper," and so on), suggesting that Figure 3.7b is later than Figure 3.7a. If this ordering is correct, Figure 3.7 suggests that Wright first divided each primary subdivision into secondary subdivisions by following the grid of the roads along which stations were created and by making divisions at each level as uniform as possible (see Table 3.4).

a

b

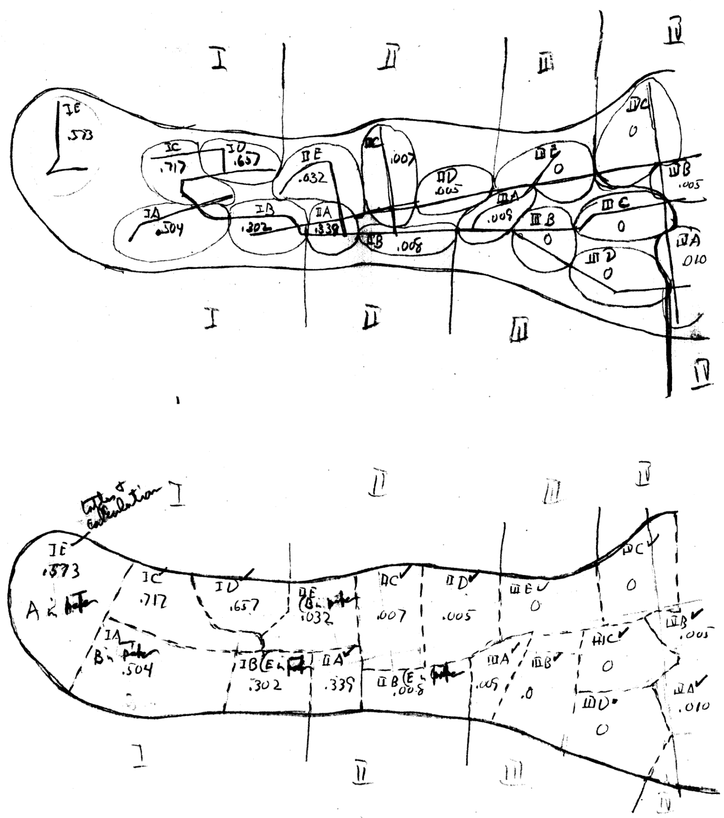


Figure 3.7: Wright's drafts of the map of the distribution of flower colors in *Linanthus*. Each represents the western half of the *Linanthus* population. The eastern side of each map is now shown. Reproduced from Sewall Wright Papers, Series IIa, Folder 5 with permission from the American Philosophical Society.

In the 1943 paper, Wright used both F_{IS} and F_{ST} to determine the size of a unit population in *Linanthus*. For each hypothesis about the mode of inheritance, Wright computed the average F_{IS}. Table 3.5 shows the basic quantities derived from the data on the assumption that blue is dominant. Wright calculated these quantities on September 18 and 19, 1942 (Figure 3.8, 3.9, 3.10). Wright considered only those stations in which the mean frequency of the blue gene among samples within a station was between 0.10 to 0.90. In the left column in Figure 3.8 (see the close-up in Figure 3.9), he wrote down station numbers in the order of increasing values of gene frequency. In the second column from the right, he recorded the value of F_{IS} for each station where station was taken to be the subpopulation (see the column indicated by the arrow in Figure 3.9).[17]

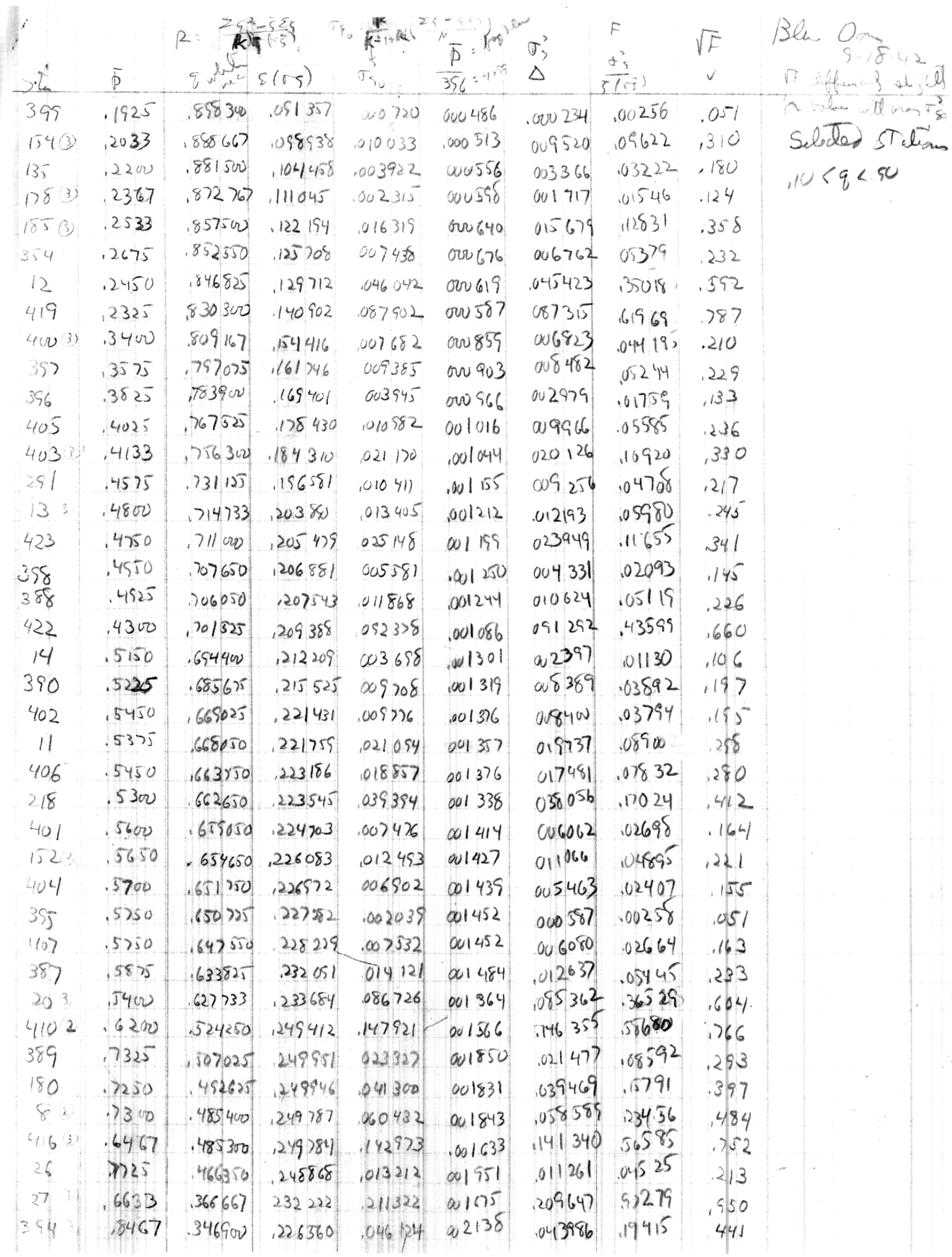
On September 19, 1942, Wright calculated the mean F_{IS} of 0.2095 in the space on the right of Figure 3.10 and rounded the value to 0.210 in Table 3.5. This is a squared value, and the value that is directly comparable to Figure 3.4 is $\sqrt{0.210} = 0.46$. As shown in the bottom two rows of Table 3.5, Wright considered two hypotheses about the number of unit populations within a station: 2×10^4 units per station (hypothesis A) and 2×10^2 units per station (hypothesis B). Hypothesis A was more plausible given the abundance of flowers in 1941, but in case 1941 was an exceptional year, Wright considered the other hypothesis. Using these quantities, Wright interpolated the size of N_u between the curves for $N_u = 20$ and 50 in Figure 3.4. He concluded that N_u would be about 45 under hypothesis A and 25 under B (Wright 1943a, 145–146).

Wright's analysis of the *Linanthus* data in terms of F_{ST} was based on the data as organized in Table 3.4 and Figure 3.6, both of which reflected the assumption of the spatial homogeneity of parameter values (e.g., Wright 1943a, 149–155). Wright calculated the value of F_{ST} from the *Linanthus* data by taking each level of the hierarchy in Table 4 as the subpopulation that F_{ST} refers to. He then compared the values thus obtained with the predicted values of F_{ST} for N_u = 10 and 20 (Figure 3.11; Wright 1943a, 152). Wright concluded that the effective size of a unit population is between 14 and 27. Such small size in turn implied that changes in gene frequency among unit populations and possibly among stations (lower level of the hierarchy) are random. Therefore, he argued that random changes in gene frequency in local populations could explain the local patterns of the distribution of flower color (e.g., the fact that two adjacent stations had widely different frequencies of blue). However, Wright also found that the variability among higher levels

[17] Note that F_{IS} according to Equation (3.5.2) would be the square root of the values given in Figure 3.8.

Cross Fertilization Single Gene Difference	
	Blue Dominant
Range (q)	.10 to .90
No.	47
$\bar{q}$	.380
$\sigma_{\bar{q}}$	.202
$\sqrt{\bar{F}}$	.380
$\sigma_{\sqrt{\bar{F}}}$	.256
r	$+.56 \pm .10$
b	$+.70 \pm .16$
$\bar{F}$	.210
N (A)	25
N (B)	14

Table 3.5: Selected columns from Wright's table of values. Symbols used: $\bar{q}$, mean gene frequency; $\sigma_{\bar{q}}$, standard deviation of the distribution of gene frequencies; $\sqrt{\bar{F}}$, mean F value; $\sigma_{\sqrt{\bar{F}}}$, standard deviation of the distribution of F values; r, correlation between $\sqrt{\bar{F}}$ and $\bar{q}$; b, regression of $\sqrt{\bar{F}}$ on $\bar{q}$; $\bar{F}$, mean F value (squared); N (A), effective size of a unit population under hypothesis A (2×10^4 units per station); N (B), effective size of a unit population under hypothesis B (2×10^2 units per station). Redrawn from Wright (1943a, 145) with permission from the Genetics Society of America.



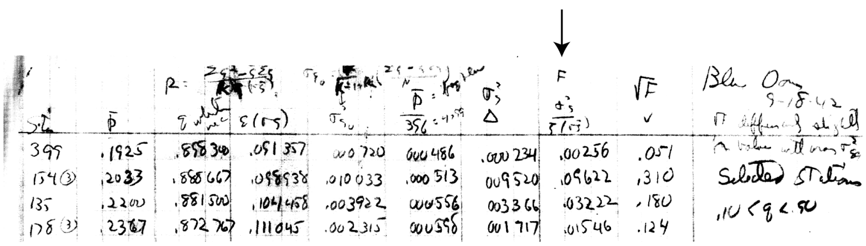
St.	$\bar{p}$	q wtd.	S(rs)	$\sigma_{s.o.}^2$	$\bar{p}$	σ_s^2 Δ	F σ_s^2/S(rs)	$\sqrt{F}$ v
399	.1925	.858340	.051357	.000720	.000486	.000234	.00256	.051
154(3)	.2033	.888667	.098938	.010033	.000513	.009520	.09622	.310
135	.2200	.881500	.104458	.003922	.000556	.003366	.03222	.180
178(3)	.2367	.872767	.111045	.002315	.000598	.001717	.01546	.124
155(3)	.2533	.857500	.122194	.016319	.000640	.015679	.12831	.358
354	.2675	.852550	.125708	.007438	.000676	.006762	.05379	.232
12	.2450	.846825	.129712	.046042	.000619	.045423	.35018	.592
419	.2325	.830300	.140902	.087902	.000587	.087315	.61969	.787
400(3)	.3400	.809167	.154416	.007682	.000859	.006823	.04419	.210
357	.3575	.797075	.161746	.009385	.000903	.008482	.05244	.229
396	.3825	.783900	.169401	.003945	.000966	.002979	.01759	.133
405	.4025	.767525	.178430	.010982	.001016	.009966	.05585	.236
403(3)	.4133	.756300	.184310	.021170	.001044	.020126	.10920	.330
251	.4575	.731125	.196581	.010411	.001155	.009256	.04708	.217
13(3)	.4800	.714733	.203880	.013405	.001212	.012193	.05980	.245
423	.4750	.711000	.205479	.025148	.001199	.023949	.11655	.341
358	.4550	.707650	.206881	.005581	.001250	.004331	.02093	.145
388	.4925	.706050	.207543	.011868	.001244	.010624	.05119	.226
422	.4300	.701825	.209388	.092338	.001086	.091252	.43599	.660
14	.5150	.694400	.212209	.003698	.001301	.002397	.01130	.106
390	.5225	.685675	.215525	.009708	.001319	.008389	.03892	.197
402	.5450	.669025	.221431	.009776	.001376	.008400	.03794	.195
11	.5375	.668050	.221755	.021094	.001357	.019737	.08900	.298
406	.5450	.663750	.223186	.018857	.001376	.017481	.07832	.280
218	.5300	.662650	.223545	.039394	.001338	.038056	.17024	.412
401	.5600	.659050	.224703	.007476	.001414	.006062	.02698	.164
152(3)	.5650	.654650	.226083	.012493	.001427	.011066	.04895	.221
404	.5700	.651750	.226972	.006902	.001439	.005463	.02407	.155
355	.5750	.650725	.227282	.002039	.001452	.000587	.00258	.051
407	.5750	.647550	.228229	.007532	.001452	.006080	.02664	.163
387	.5875	.633825	.232051	.014121	.001484	.012637	.05445	.233
20(3)	.5400	.627733	.233684	.086726	.001364	.085362	.36529	.604
410(2)	.6200	.524250	.249412	.147921	.001566	.146355	.58680	.766
389	.7325	.507025	.249951	.023327	.001850	.021477	.08592	.293
180	.7250	.452625	.249846	.041300	.001831	.039469	.15791	.397
8(2)	.7300	.485400	.249787	.060432	.001843	.058589	.23436	.484
416(3)	.6467	.485300	.249784	.142973	.001633	.141340	.56585	.752
26	.7725	.466350	.248868	.013212	.001951	.011261	.04525	.213
27(3)	.6633	.366667	.232222	.211322	.001675	.209647	.90279	.950
354(3)	.8467	.346900	.226360	.046124	.002138	.043986	.19415	.441

Blue Orange 9-18-42

Selected Stations

.10 < q < .90

Figure 3.8: A page from Wright's analysis of the *Linanthus* data. The page is dated September 18, 1942 (upper right). Reproduced from Sewall Wright Papers, Series IIa, Folder 11 with permission from the American Philosophical Society.



399	.1925	.888340	.091357	.000720	.000486	.000234	.00256	.051	
154 (3)	.2033	.888667	.098938	.010033	.000513	.009520	.09622	.310	Selected Stations
135	.2200	.881500	.104458	.003922	.000556	.003366	.03222	.180	.10 < q < .90
178 (3)	.2367	.872767	.111045	.002315	.000598	.001717	.01546	.124	

Figure 3.9: Close-up of Figure 3.8. Reproduced from Sewall Wright Papers, Series IIa, Folder 11 with permission from the American Philosophical Society.

of the hierarchy, measured by F_{ST}, was greater than could be explained by random gene frequency changes alone. In other words, it was difficult to explain the global pattern of the distribution (e.g., the fact that there is a large area of all whites) by appealing solely to random drift. Wright thus suggested that the cumulative effect of mutation between blue and white, occasional long-distance migration, and slight selection for white could explain the global pattern (Wright 1943a, 155).

3.5.3 Aims of Research

Having reconstructed Wright's research on *Linanthus* in the early 1940s, we can now take a deeper look at his work. This subsection focuses on the aims of his research during this period and the next subsection on his uses of models.

In the light of the above reconstruction, we can formulate two hypotheses about the aims of the first phase of Wright's research.

> *Illustration*. Wright aimed to illustrate isolation by distance with a case from natural populations, that is, to substantiate the idea that effective population size can be much smaller than apparent population size even if there are no obvious geographical barriers.
>
> *Explanation*. Wright aimed to explain how the observed geographical patterns of flower color dimorphism are maintained in the *Linanthus* population.

These hypotheses are not mutually exclusive, and both are plausible.

Illustration gains support from Wright's 1941 letter and his paper on *Linanthus*. Wright took the *Linanthus* case to be "a good example of isolation by distance" (Wright to Dobzhan-

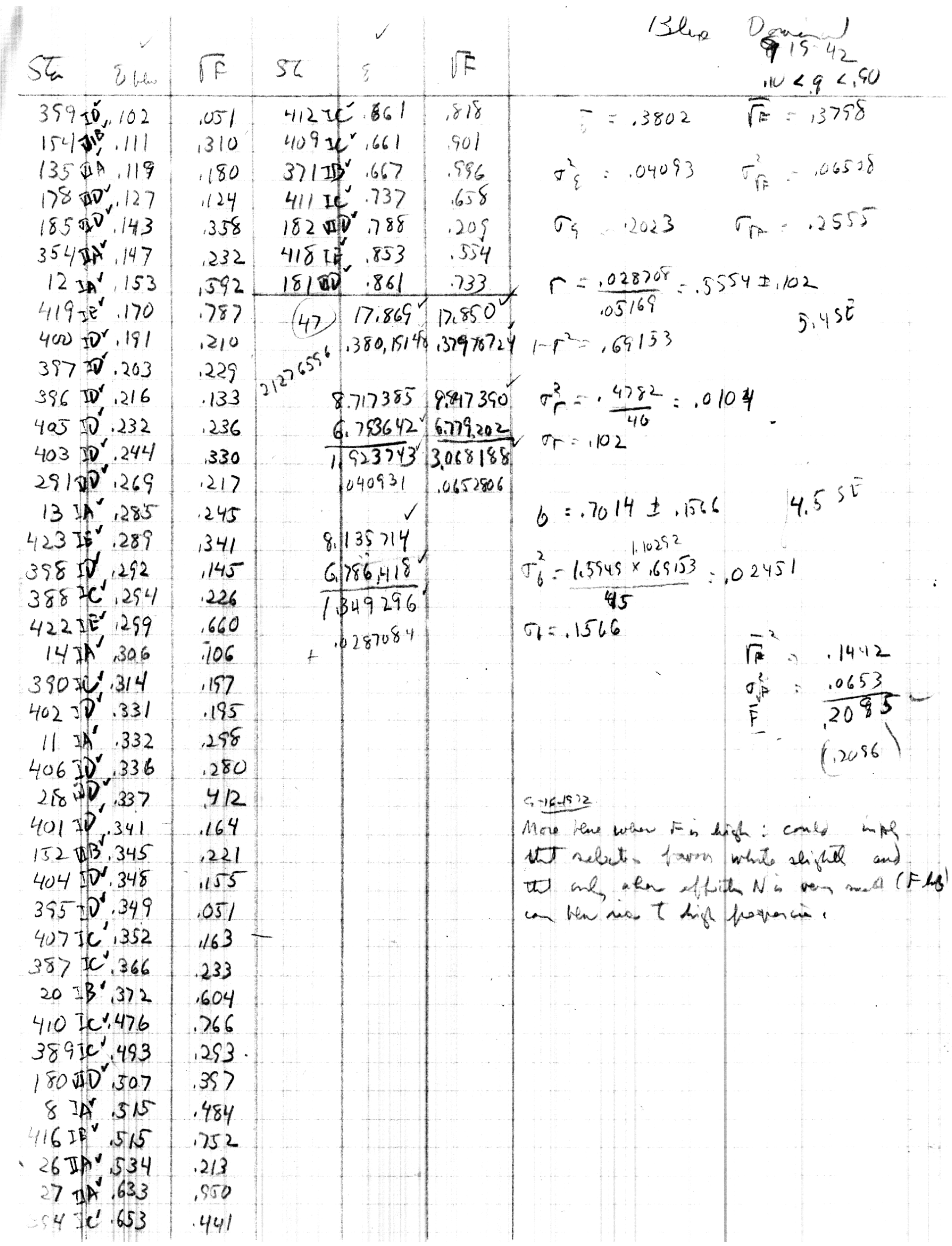


Figure 3.10: A page from Wright's analysis of the *Linanthus* data. The page is dated September 19, 1942 (upper right). Reproduced from Sewall Wright Papers, Series IIa, Folder 11 with permission from the American Philosophical Society.

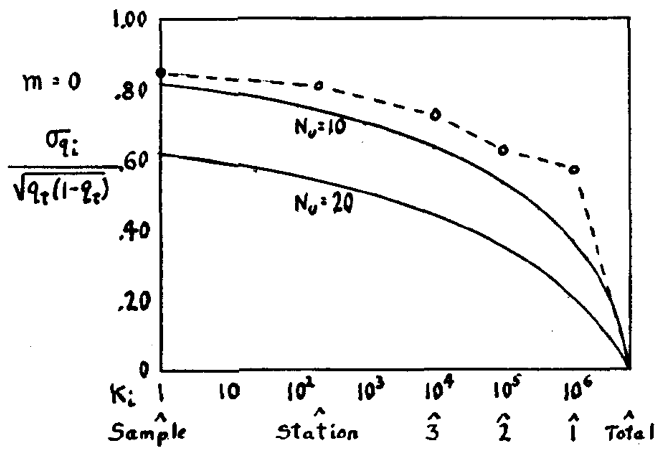


Figure 3.11: Wright's comparison of values of F_{ST} derived from the *Linanthus* data (dashed line) and calculated on the assumption of N_u = 10 and 20 (solid lines). Reproduced from Wright (1943a, 152) with permission from the Genetics Society of America.

sky, November 1941) and that he wrote his own analysis of the *Linanthus* data as a potentially interesting application of the technical results developed in his paper on isolation by distance (see, especially, Wright 1943a, 139, 144, 155). As we saw above, Wright attempted to determine the effective population size of the *Linanthus* population and to estimate the amount of variability due to random differentiation among unit populations (Wright to Dobzhansky, November 1941; Wright 1943a). His activities make sense as part of the strategy to show that the *Linanthus* case is an example of isolation by distance in nature.

Explanation gains support from the fact that Wright *did* suggest possible evolutionary factors which could explain the observed distribution of flower colors. Explanation here is to show how the observed flower color distribution in the *Linanthus* population could be maintained, given a combined influence of such evolutionary factors as selection, mutation, migration, and genetic drift on frequencies of genes in the population.[18] By 1941, Wright had already attempted to provide explanations in this sense for other natural populations. He addressed the distribution of self-sterility alleles in populations of *Oenothera organensis* (the Organ Mountains evening prim-

[18]I do not mean that this is the only form of explanation in classical population genetics; there may be others (see Plutynski 2004).

rose) in New Mexico, using Sterling Emerson's (1938, 1939) data (Wright 1939; see also Provine 1986, 488–491). And with Dobzhansky, he worked on the distribution of lethals in populations of *Drosophila pseudoobscura* in California (Dobzhansky and Wright 1941). We can regard the first phase of Wright's *Linanthus* research as belonging to this series of work. In his letter to Dobzhansky, Wright provided an explanation in the above sense, and he further elaborated it in his paper on *Linanthus*. In each case, he suggested a set of evolutionary factors that could maintain the patterns of flower color distribution observed in the 1941 survey (Wright to Dobzhansky, November 1941; Wright 1943a, 155–156).

Given these considerations, we can say that Wright had a basic aim of illustrating isolation by distance and a more ambitious aim of explaining the observed distribution of flower colors. Moreover, we can also assume that in the first phase of his research, Wright primarily aimed to investigate the patterns of the distribution of gene frequencies in the *Linanthus* population and their evolutionary causes. I will use this assumption to identify models and analyze their uses.

3.5.4 Uses of Models and Criteria of Success

From the cognitive science based perspective on models, we can identify the following external representations as some of the models in Wright's research.[19] For convenience, I refer to them by suggestive names.

Steady State Equation (Equation 3.5.1)

Isolation-by-Distance Diagram (Figure 3.2)

F-Equation (Equation 3.5.2)

F-Curves (Figure 3.3, 3.4, 3.5)

Here I analyze the uses of these models. As we saw above, the first phase of Wright's research focused primarily on the patterns of the distribution of gene frequencies in the *Linanthus* population and their evolutionary causes.

3.5.4.1 Steady State Equation In 1941 Wright used the steady state equation (Equation 3.5.1) in three ways: to compute quantities that made up the distribution of gene frequencies, to predict

[19]The list is not meant to be exhaustive.

the distribution of gene frequencies in the *Linanthus* population, and to infer a combination of evolutionary factors that could explain the observed distribution of flower color (see Table 3.3).

Different criteria of success were likely to be relevant to these uses of the steady state equation. First, some reasonable criteria of success for the computational use of the equation were that the equation should be mathematically well-defined and should enable Wright to do computation efficiently and precisely within the available resources. Taking into account that Wright had to do all the computations by hand, we can say that the steady state equation met these criteria well. Second, for the predictive use, one reasonable criterion was that it should be possible to give a biological interpretation of the equation. For example, the equation should give a value between 0 to 1 so that it can be interpreted as *gene frequency*, and it of course did. Another reasonable criterion was that the predictive use of the equation should produce a pattern that could be compared to the statistical pattern brought out in Table 3.3. For Wright's strategy in his letter to Dobzhansky was to use the comparison between the two patterns to infer the relevant evolutionary factors and their magnitudes in maintaining the flower color dimorphism in *Linanthus*. As we saw above, the predictive use of the steady state equation failed to meet this criterion in the sense that the distribution of gene frequencies was not directly comparable to the frequency distribution of samples given in Table 3.3. But Wright was able to convert the distribution of gene frequencies into the same unit and scale of the other distribution, making the uses of the steady state model indirectly meet the criterion of comparability. Third, for the use of the equation to infer evolutionary factors, one reasonable criterion of success was that it should be possible to describe the effects of different evolutionary factors by the equation. One of Wright's major accomplishments in his 1931 paper (Wright 1931) was to derive a steady state equation like Equation (3.5.1) that properly captured the effects of various evolutionary factors, such as mutation, selection, migration, and genetic drift.

3.5.4.2 Isolation-by-Distance Diagram Wright used the diagram of isolation by distance (Figure 3.2) to describe a structure of population for the analysis of which he developed F-statistics. No surviving notes from 1941 and 1942 contain a diagram like this. Wright's early, surviving notes on population structure, one of which is dated September 29, 1938, contain similar diagrams (Sewall Wright Papers, Series IIa, Folder 9), but it is difficult to infer from the notes what he was doing with the diagrams. Thus, we can only say that Wright put the isolation-by-distance diagram

(Figure 3.2) to a descriptive use in his letter to Dobzhansky (Wright to Dobzhansky, November 1941).

Wright's descriptive use of the diagram resulted in the following claims about the *Linanthus* population:

> *Hierarchical*. The total *Linanthus* population of size N_t is divided into subpopulations, each of which has the size N_i and is further divided into unit populations of size N_u.
>
> *Circular*. The geographical area occupied by the *Linanthus* population, subpopulations, and unit populations is circular.

Wright clearly states *Hierarchical* by referring to Figure 3.2 (see the passage, quoted above, from his letter to Dobzhansky). *Circular* also resulted from Wright's use of the diagram as a description of population structure, because Wright interpreted N_t, N_i, and N_u as an area occupied by a fixed number of individuals. This latter point can be seen clearly in his interpretations of F values explained above.

There were at least two reasonable criteria of success relevant to the descriptive use of the isolation by distance diagram. One is suggestiveness: Wright was trying to communicate an idea of isolation by distance to Dobzhansky, that is, the idea of hierarchical population structure, so the diagram should suggest this idea. The *nested* figures in the diagram met this criterion well. Another relevant criterion is consistency with Wright's assumptions about the area of unit populations. Wright assumed that the area of a unit population was circular, so the diagram should be consistent with this assumption. The circular figures in the diagram met this criterion.

3.5.4.3 F-Equation and F-Curves Wright put the F-Equation and F-Curves to a variety of uses. First, as we saw in his notes (Figure 3.8 and 3.10), Wright used the F-Equation (Equation 3.5.2) to calculate values of F from Epling and Dobzhansky's data. Let us take a closer look at Figure 3.9. Wright wrote the right hand side of the squared F equation in the second column from the right. To the left of this column he wrote the quantities to be substituted into the equation. He wrote the value of the numerator of the F-Equation in the third column from the right and the value of the denominator in the fourth column from the left. Following the F-Equation, Wright divided the former by the latter and recorded the result in the second column from the right. Second, Wright took quantities computed according to the F-Equation as points on F-Curves shown in

Figures 3.4 and 3.5. He used these F-values to describe genetic characteristics of the *Linanthus* population, in particular, the extent of local genetic differentiation relative to the limiting value. Third, Wright used the F-Curves to infer the effective population size (i.e., the size of the unit population).

Like the steady state equation, different criteria of success were likely to be relevant to these uses of the F-Equation. First, reasonable criteria of success for the computational use of the F-Equation included that the equation should be mathematically well-defined and enable Wright to do computation efficiently and precisely. Second, for the descriptive use, one reasonable criterion was that the F-Equation should make biological sense: the right hand side of the equation should admit a biologically reasonable interpretation. The equation did have such an interpretation, for the right hand side was the ratio of the actual spread of the distribution of gene frequencies among subgroups to its limiting value. In other words, the value of F means how far subgroups are differentiated: the closer the value is to 1, the more differentiated subgroups are. Third, for the use of the F-Curves to infer the effective population size, one reasonable criterion was that it discriminates different effective sizes. The F-Curves met this criterion as Wright showed that distinct F-Curves arise for different effective sizes (e.g., Figure 3.4).

I have just given an analysis of Wright's uses of models and criteria of success that would have been relevant to his uses of models. In the following sections I focus on his uses of the steady state equation, the F-Equation, and the F-Curves in the second and third phases of his research. But it is important to emphasize that by the end of 1942 Wright had no plan of continuing to work on *Linanthus*. How he would work on *Linanthus* was open-ended, but as we shall see, he decided to use the same models while his specific aims changed.

3.6 SECOND PHASE: EPLING, WRIGHT, AND THE HARDENING OF THE MODERN SYNTHESIS

In 1944, Epling began a long-term study of *Linanthus*. In each year he collected data at transects set up in an area much smaller (20 sq mi) than that surveyed in 1941 (840 sq mi) and periodically sent the new data to Wright (see, e.g., Dobzhansky to Wright, May 15, 1944; Sewall Wright

Papers, Series IIa, Folder 10). Wright spent countless hours analyzing the new series of data. Meanwhile, in 1959, the year of the Darwin Centennial, Epling and his collaborators Harlan Lewis and Francis Ball submitted to *Evolution* a manuscript on their long-term study of *Linanthus*. Contrary to the conclusion reached by Epling, Dobzhansky, and Wright in the early 1940s, Epling and colleagues argued that selection, rather than genetic drift, is the primary cause of the observed patterns of distribution of flower colors (Epling et al. 1960, 254). Wright refereed the manuscript and recommended it for publication (Provine 1986, 486).

In response to Epling and colleagues' paper, Wright wrote two manuscripts, one in 1960 and another in 1962. In both manuscripts, Wright criticized a particular form of inference to natural selection Epling and colleagues used. The form of inference in question goes from the evidence that the influence of random drift is not or cannot be strong in a population to the conclusion that natural selection must be operating in that population. This form of inference reflected a characteristic of the hardened evolutionary synthesis, and Wright's response clarifies what, for him, would count as an adequate explanation of the flower color dimorphism in *Linanthus*.

3.6.1 Hardening of the Modern Synthesis

According to Stephen Jay Gould, in its early days (1930s), the so-called modern synthesis in evolutionary biology was pluralistic in that a range of mechanisms, including Darwinian natural selection and random genetic drift, was accepted as legitimate explanations of evolutionary phenomena. But Gould argued that the synthesis gradually hardened into a monistic view that took natural selection as the primary, if not the only, mechanism of evolution. He wrote:

> The original version of the synthesis . . . did not attempt to crown any particular cause of change, but to insist that all permissible causes be based on known Mendelian mechanisms. In particular, it did not insist that adaptive, cumulative natural selection must underlie nearly all, or even most, change—though many synthesists personally favored this view. Any theory of change would be admitted, so long as its causal base lay in known Mendelian genetics. In the 1930s, for example, genetic drift was often granted a predominant role in phenotypic change, not only at the level of demes, but also in the origin of many species. . . . I have called this original version "pluralistic" because

> it admitted a range of theories about evolutionary change, Darwinian and otherwise, and insisted only that explanations at all levels be based upon known genetic causes operating within populations and laboratory stocks.
>
> This pluralistic version was slowly and subtly altered, primarily during the 1940s (and perhaps with the 1947 Princeton conference as a focal point), as the intent of explanation by known genetics shifted to the content of one particular theory—neo-Darwinism and its insistence that cumulative natural selection leading to adaptation be granted pride of place as the mechanism of evolutionary change. The synthesis hardened by elevating one theory to prominence among the several that supported the primary methodological claim of the original version—and eventually … by insisting to the point of dogma and ridicule that selection and adaptation were just about everything. (Gould 1983, 74–75)

Gould illustrates this "hardening of the modern synthesis" (Gould 1980, 1982, 1983) by showing the hardening in Dobzhansky, Simpson, and Wright. It has also been documented by other historians (Provine 1986, Ch. 12; Smocovitis 1996, 146–147). The heyday of the hardened synthesis was obvious at the 1959 Darwin Centennial Celebration held at the University of Chicago (Gould 1983, Smocovitis 1999).[20]

There is a particular form of inference to selection that was licensed by the hardened synthesis and which Wright consistently resisted. According to the hardened synthesis, selection is the most important and most prevalent mechanism of evolutionary change, so the following forms of inference are all valid:

Selection only.

There is evidence for selection for a trait in a given population.

So:

[20]Provine has argued that Wright's shifting balance theory hardened. This may seem counterintuitive because throughout his career, Wright emphasized a balance of evolutionary factors, such as selection, mutation, migration, and drift. Selection was always one of many factors of evolution. But, according to Provine, in the 1930s Wright held that taxonomic differences above the species level were largely nonadaptive and therefore random in character, but by the 1950s he held that only local differences within a species were random. Within Wright's shifting balance theory, the level at which random genetic drift played a role came to be more restricted than before (Provine 1986, 361–362). The relation between Wright's theory and the hardening of the modern synthesis is thus not straightforward: On the one hand, Wright's theory hardened in the way just described, but on the other hand, his theory did not harden because he never regarded selection as the only important mechanism of evolution. As I show below, Wright's response to Epling and colleagues supports this latter point.

Selection explains the origin and maintenance of the trait of interest.

No Drift then Selection.

There is no evidence for drift being a significant causal influence on the frequency of a trait in a given population.

So:

Selection explains the origin and maintenance of the trait of interest.

To reject both *Selection only* and *No Drift then Selection*, one may point out that selection is not the only cause of evolutionary change and that a combination of different causes may explain the origin and maintenance of a trait. In addition, to reject *No Drift then Selection*, one may argue that the lack of evidence for one evolutionary factor is not evidence for some other factor. To emphasize a combination of different causes is to reject the central commitment of the hardened synthesis that natural selection is the only, most important, or most prevalent evolutionary mechanism. The rejection of the hardened synthesis just outlined, however, is not exactly a reversion to what Gould calls the original, pluralistic version of the synthesis. For the rejection is based not only on the idea that there are selective (or adaptive) and non-selective (or non-adaptive) evolutionary mechanisms—the central idea of the original version of the synthesis—but also on the idea that there are different evolutionary mechanisms operating in concert to produce whatever phenomena we find in a given population. This latter idea, for example, is invoked in the rejection of *Selection only*. To distinguish these ideas, we can call the former pluralism and the latter interactionism.

Wright's view was interactionism, and he strongly resisted *No Drift then Selection* when it was employed by Epling and colleagues in their 1959 paper.

3.6.2 Epling and the Hardening of the Modern Synthesis

After the original 1941 survey, Epling launched what would become more than twenty years of observational and experimental studies of *Linanthus*. In the spring of 1944 Epling and his associates set up a permanently marked transect in the area where both blue and white had coexisted at least since the 1941 survey. The area was located in the western part of the Mojave Desert, near Pearblossom, California. The transect was a half mile long and ten feet wide, running from west to east. It was further divided into 260 quadrats (10 ft 10 ft each) (Epling et al. 1960, 241–242). In 1944, the density of *Linanthus* was not uniform in each quadrat, and Epling and his associates recorded the locations of the areas of greater density for each quadrat. They also sampled the frequencies of blue and white by first determining, for each quadrat, the densest square foot area and then recording the numbers of blue and white within that area. They recorded the location of the densest area (that is, their sampling location) in each quadrat. Epling and his associates collected the data in this way every year since 1944 except when there were no or too few plants to count (1950, 1951, 1955, 1956, and 1958).[21] Epling and colleagues calculated the mean frequency of blue for each quadrat over those ten years where counts were made. They also compared the mean frequencies for the entire transect in the period 1944–1947 and that in the period 1953–1957 (excluding 1955 and 1956 as no counts were made in these years). According to Epling and colleagues, the comparison showed that "the cline of frequencies illustrated has remained constant" (Epling et al. 1960, 245).

In their paper, Epling and colleagues made two arguments for the claim that selection, as opposed to random genetic drift, is the primary cause of the observed distribution of flower color in *Linanthus*. In the first argument, Epling and colleagues said that the stability of the cline of phenotype frequencies suggests that "if genetic drift has played a role, it has been of only local consequence and not persistent in its effects" (Epling et al. 1960, 254). "*Conversely*," they claimed, the stability of the cline suggests "an intense local selection because the blues are concentrated in certain areas and because persisting clines of blue and white frequencies have been found" (Epling et al. 1960, 254; emphasis mine). Note the pattern of inference here: According to Epling and colleagues, the stability of the cline suggests that drift is not a significant factor, and it "conversely" suggests that selection must be the main factor. This is a clear instance of the eliminative inference to selection of the form referred to above as *No Drift then Selection*.

[21] In addition to this transect study, Epling and his associates did the following: In 1944, they established three stations in the area where blue and white coexisted. Every year since 1944, they recorded the frequencies of blue and white in these plots. In 1948, they established two new plots for the elimination experiment where plants of *Linanthus* were removed before they left seeds in each year so that the next year's plants had to grow from whatever seeds dormant underground. In November 1954, they transplanted the seeds obtained from blue flowered plants living in an all-blue area to an all-white area in order to test the viability of seeds in different areas. The results of all these studies were reported in Epling and colleagues' paper (1960).

In the second argument, Epling and colleagues noted that seeds of *Linanthus* could remain dormant at least seven years in the soil.[22] After noting that the seed storage increases the effective population size (Epling et al. 1960, 254), Epling and colleagues said:

> The conclusion seems warranted, *therefore*, that the frequencies of blue and white flowered plants are in the long run the product of selection operating at an intensity we have been unable to measure; and that the large size of the effective population, and the localized dispersion of pollen and seeds, has precluded significant changes in pattern during 15 seasons. (Epling et al. 1960, 254; emphasis mine)

Again note the form of inference: According to Epling and colleagues, the presumably large size of the seed storage makes the effective population size large. In a large population, drift cannot be a significant factor. "Therefore," selection must be the main factor although its intensity is too small to be detected. Like the first argument, the inference is eliminative: Epling and colleagues infer selection from the fact that drift cannot be the main factor.

3.6.3 Wright's Response

In response to Epling and colleagues, in January 1960, Wright began a manuscript on *Linanthus*, an apparently unfinished, handwritten version of which has survived (Sewall Wright Papers, Series IIa, Folder 29).[23] In the introduction of the manuscript, Wright explicitly noted that Epling and colleagues argued for selection only by elimination of random drift:

> They [Epling, Lewis, and Ball] conclude that the frequencies of blue and white flowered plants are in the long run the product of selection. This conclusion was however arrived at *only by elimination* since studies of topography, soil samples, and of associated vegetation in areas in which one or the other color predominates have given no indication of any basis for differential selection. (Sewall Wright Papers, Series IIa, Folder 29; emphasis mine)

[22] This was found in the elimination experiments.

[23] Yet, by June 1962, he returned to *Linanthus*, producing a complete typescript with tables and figures (Sewall Wright Papers, Series II, Box 2). In an interview with Provine, Wright said that in the 1960s he was planning to publish his analysis in a joint paper with Epling but that the project stalled (Sewall Wright Papers, Series IIIa, Tape 11, 1979; Provine 1986, 486–488). Moreover, at that time, Wright devoted himself to writing his treatise *Evolution and the Genetics of Populations*. He thus did not publish the *Linanthus* manuscripts.

Epling and colleagues did not find any positive evidence for selection. But interpreting the apparent stability of the frequencies of blue and white as evidence against drift, they concluded that somehow selection must be in play. This reasoning was unacceptable for Wright. Thus he wrote:

> The problem of finding positive evidence for any interpretation of the pattern thus remains: The purpose of this paper is to consider whether statistical analysis of the new data gives any positive evidence of any sort. (Sewall Wright Papers, Series IIa, Folder 29)

One section of the manuscript was entitled "The Problem," and in it, Wright emphasized that the important problem concerning *Linanthus* is far more complex than the problem of determining whether selection or drift is responsible for the patterns of distribution of flower color. He wrote:

> The problem presented by distribution of blue and white flowered plants is not as simple as a decision between two sharply distinct alternative[s]: control by selection or by random drift. Selection may be involved in diverse ways, there are different sorts of random drift to be considered, and selection and random [drift] may be combined in any degrees and may conceivably interact to produce a more heterogeneous pattern than either by itself. (Sewall Wright Papers, Series IIa, Folder 29)

The first sentence criticizes Epling and colleagues' eliminative inference, and the rest of the passage illustrates the way in which Wright's evolutionary theory did not harden and what he regarded as an adequate explanation of the flower color dimorphism in *Linanthus*. For Wright, selection was not the only cause of evolution, and an adequate explanation would have to appeal to a particular combination of evolutionary factors, notably selection and drift. The above passage does not refer to mutation and migration, but these must also be considered (as Wright did in the early 1940s).

At some points in the 1962 manuscript, Wright used F-Curves and the 1941 data to argue for an interactionist hypothesis that the observed pattern of distribution of flower color was the result of interaction between random drift in subpopulations and migration (Sewall Wright Papers, Series II, Box 2; see also Series IIa, Folder 29 for F-Curves Wright drew in 1962). He produced the curves similar to Figure 3.5 and explained the idea of unit population (now called "neighborhood") and how the size of unit population changes the curves. He then showed the curves of F_{ST} for each of the six primary subdivisions (see Table 3.4 and Figure 3.6). He showed that given the average F_{ST} values for variable areas (primary subdivisions I, II, and VI), the best fit F_{ST} curves were those calculated according to the hypothesis that the size of a unit population is 10 and that there is certain amount of migration closely (Sewall Wright Papers, Series II, Box 2).[24]

Wright paid special attention to clinal selection and contrasted it with an explanation based on the interaction between random drift and dispersal. A cline, introduced by Julian Huxley (1938), refers to a spatial gradient of the distribution of phenotypes or genotypes. Such gradient is supposed to arise from corresponding differences in environmental conditions. Thus, the presence of a cline can be taken to be an indication of spatially varying selection. Wright considered two forms of a cline: plane cline (uniform gradient in one direction) and conical cline (gradient falling off uniformly in all directions from a point). Wright tested how well the values of F_{ST} based on interaction between drift and migration, the plane cline hypothesis, and the conical cline hypothesis fit the 1941 *Linanthus* data. The interaction between drift and migration turned out to fit the data better than clinal selection hypotheses. Moreover, Wright argued that it would be surprising if there were fine-grained clinal selection in the areas where blue and white were mixed:

> To account for the observed distribution of blue on a largely selective basis would require a distribution of selective values, favorable and unfavorable to blue, in a fine-grained pattern that happens to simulate very closely that expected from random drift and dispersion. This would be a surprising pattern of selection to find in an apparently uniform environment since it requires such a delicate balance between opposed selective advantages that there is reversal an enormous number of times within any mixed area. The evidence from the variance-area curve [i.e., F_{ST} curves] thus points strongly toward the joint effects of random drift and dispersion as the principal explanation of the pattern. (Sewall Wright Papers, Series II, Box 2)

Clinal selection is directional (for or against blue) so that in order for it to maintain dimorphism within a small area, its direction needs to change from time to time. Since in the 1941 data there were many such mixed areas in an apparently uniform environment, Wright argued that a kind of selection that could maintain dimorphism in these areas would be a surprising form of selection.[25]

[24] According to Wright's surviving notes, he calculated relevant F values in late June 1962 (Sewall Wright Papers, Series IIa, Folder 29).

[25] This was not the first time Wright considered the hypothesis of local environmental selection with regard to *Linanthus*. The hypothesis was suggested by William Hovanitz in 1942 after he saw Epling and Dobzhansky's paper (Hovanitz to Wright, May 28, 1942; Wright to Hovanitz, June 11, 1942; see also Provine 1986, 375–376). Wright acknowledged Hovanitz' suggestion in his 1943 paper on *Linanthus* (Wright 1943a, 155).

In both of his unpublished manuscripts Wright explicitly stated that the problem of *Linanthus* was not the simple choice between selection and drift as a main cause of the flower color distribution but the determination of the balance or interaction among various factors, one of which is selection, that can produce the observed pattern of distribution. He showed the weakness of the eliminative inference used by Epling and colleagues (1960) and that in explaining the patterns of the distribution of flower color, he appealed to the interaction between random drift and migration. Wright's manuscripts were a criticism of the eliminative inference and the hardened synthesis, which took natural selection as the main mechanism of evolution.

3.6.4 Aims of Research and Uses of Models

In the second phase of his research on *Linanthus*, Wright primarily aimed to find any positive evidence for evolutionary factors that can explain the observed patterns of the distribution of flower colors. This had been a second, more ambitious aim of the first phase of his research, but it became the central aim in the second phase.

To find positive evidence for evolutionary factors involved in *Linanthus*, Wright tried to show that the patterns of the distribution of flower color were more likely to be the consequence of drift and migration than selection alone. This was done by using models: He computed F values using the F-Equation and constructed F_{ST}-curves under different hypotheses concerning evolutionary factors and compared them against the F_{ST} values derived from the data. He found that the interaction of drift and migration gave the best fit result. He thus argued that there is evidence for drift and migration as important evolutionary factors in the *Linanthus* population. In other words, in the second phase, Wright used F-Curves to make a case for interactionism in *Linanthus*, whereas in the first phase, he used F-Curves to make a case for isolation by distance in nature.

Like the first phase, Wright used values of F_{ST} to describe the extent of local genetic differentiation. But in the second phase he put F-Curves to a new use. He used F-Curves to infer evolutionary factors that could explain the observed patterns of the distribution of flower colors.[26] In the first phase Wright used F-Curves to infer the effective population size of the *Linanthus* population, but given a more explanatory aim of the second phase, it made sense for Wright to exploit

[26] In the first phase Wright did this by using the steady state equation.

the fact that F-Curves could be used to compare different explanatory hypotheses against F values derived from data.

It appears that the change in the aim of research in the second phase did not demand change in the criteria of success for the computational use of the F-Equation and the descriptive use of F-Curves.[27] That is, F-Equation had to be mathematically sound and computationally efficient in order for Wright's computational use of it to be successful, and it met this criterion well. Values of F had to make biological sense in order for the descriptive use of F-Curves to be successful at all. As we saw above, values of F met this criterion. But there was probably an additional criterion of success for the computational use of the F-Equation and the inferential use of F-Curves in the second phase. For F values had to be calculated not only for the hypothesis of drift and migration—which Wright knew how to do in the early 1940s—but also for clinal selection hypotheses. That is, it needed to be possible for Wright to manipulate the F-Equation in a mathematically sound way to compute F values for different hypotheses. The success of this computation enabled inferential uses of F-Curves in this phase.

3.7 FINAL PHASE: EXPLAINING GEOGRAPHICAL PATTERNS

Recall that in 1941 Wright found that the theoretical distribution of gene frequencies did not fit well with the *Linanthus* data. Recall also that in his analysis of the *Linanthus* data in the 1940s, Wright assumed that the parameter values are spatially homogenous. In the 1970s, he came to relax this assumption and allowed the effective population size to vary from one locality to another. With this relaxation of the spatial homogeneity assumption, he was able to improve the fit between the theoretical and the observed distributions of gene frequencies (Wright 1978, 209–211). This new development began in 1972 while Wright was preparing to include the *Linanthus* case in his treatise.

[27] This does not mean that in general the same uses are associated with the same criteria; rather, this is the case where change in the aim of research did not have impact on the uses of models and the criteria of success.

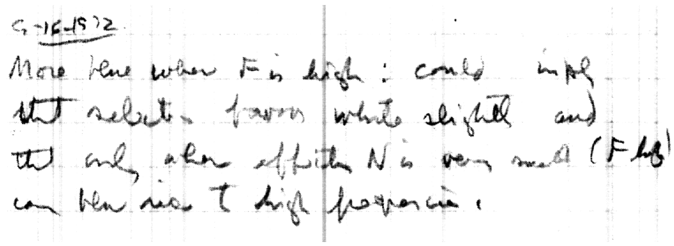

9-16-1972
More blue when F is high: could imply that selection favors white slightly and that only when effective N is very small (F high) can blue rise to high frequencies.

Figure 3.12: Close-up of Figure 3.10. Reproduced from Sewall Wright Papers, Series IIa, Folder 11 with permission from the American Philosophical Society.

3.7.1 Wright's 1972 Analysis

In September and October 1972, Wright reanalyzed the 1941 *Linanthus* data, referring back to his notes from 1942 (Sewall Wright Papers, Series IIa, Folder 40). On September 16, he wrote a new note on the table of F values he produced in 1942 (Figure 3.10 and 3.12). Various statistics derived from the table, written on the right hand side of the page, were published in Wright's 1943 paper on *Linanthus* (see Table 3.5; Wright 1943a, 145). Figure 3.10 lists sampling stations (the first column from the left) in the ascending order of frequencies of blue (the second column) together with the values of F (the third column). Wright's note on the bottom of the page, dated September 16, 1972, reads (Figure 3.12):

> More blue when F is high: could imply that selection favors white slightly and that only when effective N is very small (F high) can blue rise to high frequencies. (Sewall Wright Papers, Series IIa, Folder 11)

Wright discovered a pattern in the 1941 data: the frequency of blue is correlated with the value of F. He then inferred that blue may be slightly selected against and that its frequency can increase only when the effective population size is so small that selection is ineffective.

Wright's inference seems to be based on the following consideration. As can be seen in Equation (3.5.2), F is greater when the standard deviation σ_x of gene frequency q in a subgroup x is greater. For Wright, greater σ_x implies that change in q is predominantly random, because σ_x represents random, as opposed to directional, change in q. Now if random change is due primarily to accidents of sampling and directional change to selection, and if selection pressure is constant throughout the geographical range of the population in question, then greater F for a given subgroup implies that the effective population size N of that group is smaller. Smaller F implies larger N, which makes selection more effective. So the correlation between the frequency of blue and the value of F suggested to Wright that N is not constant throughout the *Linanthus* population: N may vary from place to place.[28]

Thus Wright's notes suggest that he came to realize the possibility of the heterogeneity of the parameter values in *Linanthus* when he was going over his original notes taken in 1942. After September 16, Wright worked steadily. By September 29, he produced the table and graphs that would appear in his last published analysis of the *Linanthus* data (Sewall Wright Papers, Series IIa, Folder 40; Wright 1978, 207–212).

3.7.2 Wright's Final Analysis

Wright's final analysis, which addressed both the 1941 survey and Epling's long-term study, was published in the fourth volume of his treatise in 1978 (Wright 1978, 194–223).[29] Like his first analysis of the 1941 data, Wright compared the theoretical and the observed distributions of gene frequencies. Unlike the first analysis, however, he did not require the parameter values to be homogenous throughout the geographical range of *Linanthus*. Instead, he allowed the effective population size to vary from one locality to another. With this relaxation of the spatial homogeneity assumption, he was able to improve the fit between the theoretical and the observed distributions of gene frequencies (Wright 1978, 209–211).

Wright began his new analysis by reproducing his 1941 result, using Equation (3.5.1) (Wright 1978, 209). He compared the theoretical and the observed distributions of gene frequencies under the assumption of the spatial homogeneity of parameter values, failing to obtain a good fit between the two distributions because of the same hump he emphasized in 1941 (Figure 3.13; Wright 1978, 207–209). However, in this analysis, by dividing the total population into western and eastern regions, he was able to locate the hump in the western region: the theoretical and the observed distributions fit relatively well for the eastern region, but they do not fit well for the western region.

[28]The spatial homogeneity of parameter values requires that all relevant parameter values be homogenous throughout the population, but the spatial heterogeneity only requires that at least one of the relevant parameter values be heterogeneous.

[29]Wright's analysis of Epling's long-term study is essentially the same as that found in Wright's 1962 manuscript.

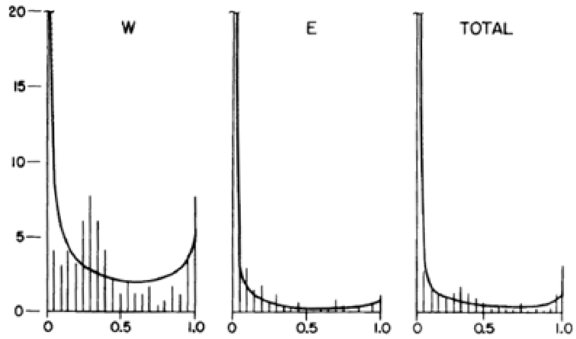


Figure 3.13: The theoretical and observed distributions of gene frequencies in *Linanthus* under the assumption of the spatial homogeneity of the parameter values. Frequencies of genes for blue (assumed to be dominant) in western, eastern, and the total populations are shown. Reproduced from Wright (1978, 207) with permission from the University of Chicago Press.

He then argued that such difference in the goodness of fit "could come about if the population is heterogeneous with respect to the parameters" (Wright 1978, 209). Referring to Epling and Dobzhansky's map of the relative frequencies of blue and white in sample stations (Epling and Dobzhansky 1942, 326; Wright 1978, 198), Wright argued that the data from the western region can be divided into two components, western central, W_C, and western peripheral, W_P. The frequency of blue was intermediate in sample stations in W_C (that is, the variance in q and hence F were low), whereas it was either high or low in those in W_P (that is, the variance in q and F were high). Relaxing now the spatial homogeneity assumption, Wright calculated the theoretical distributions of gene frequencies for W_C, W_P, and W_T (western total), using different values of size N_u of a random breeding unit and the migration rate m for each (Figure 3.14). The theoretical distribution now exhibited a hump like the observed distribution. From the values of F, Wright estimated the effective population size for W_P to be 7 or 8 and that for W_C to be about 100 (Wright 1978, 209–211).

In the light of this analysis, Wright noted that the shifting balance theory offers a plausible explanation of the observed patterns of flower color distribution in *Linanthus*. He said that it is

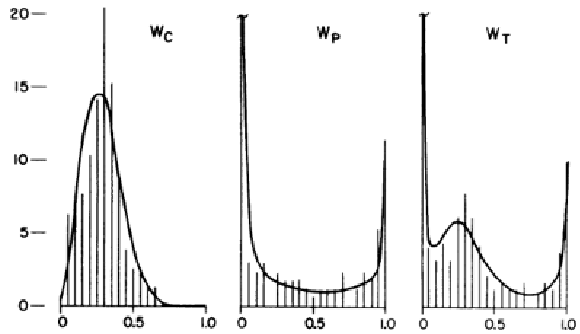


Figure 3.14: The theoretical and observed distributions of gene frequencies in *Linanthus* under the assumption of the spatial heterogeneity of the parameter values. Western central, peripheral, and total populations are shown. Reproduced from Wright (1978, 212) with permission from the University of Chicago Press.

> plausible that among the continually varying genetic compositions of local populations, arrived at by random drifting of the frequencies at all other heterallelic loci, favorable interaction systems may be arrived at which spread over large areas by interdeme selection and incidentally have some effect on the selective advantage of white over blue. (Wright 1978, 223)

In other words, according to Wright, the interaction of random drift in local populations, migration between local populations, and selection for favorable genotypes (i.e., individual selection) can explain the observed patterns of flower color distribution at all levels from local patterns in the variable areas to global patterns in the entire geographical range of *Linanthus* in the Mojave Desert.

3.7.3 Aims of Research and Uses of Models

In the final phase of his research on *Linanthus*, Wright aimed to revise his earlier analysis in order to identify a more complete set of evolutionary factors that can explain the observed patterns of the distribution of flower colors. He thus referred to the section on *Linanthus* as “a revision and extension of my early analysis” (Wright 1978, 194).

Wright used F values to make an inference to the spatial heterogeneity of N_u and selection for white. That is, he realized that F values could be put to a new inferential use with respect to the *Linanthus* data: positive correlation between frequency of blue and F could imply selection for white and variation in the size of unit populations (Figure 3.12).

Like his earlier uses of the steady state equation, Wright put the equation to computational and predictive uses. But he relaxed the homogeneity assumption so that different parameter values can be used for different regions of the *Linanthus* territory. The curves shown in Figure 3.14 corresponded to three different sets of parameter values for the steady state equation.

One criterion of success relevant to the new inferential use of F values was the reliability of making inferences about the effective population size from the values of F. Such reliability would have been needed to support Wright's inference to the heterogeneity of N from the correlated change in F with respect to the frequency of blue. As Wright had shown in the 1940s, F values vary with values of N_u (see Figure 3.4 and 3.5). Thus, it seems that Wright had good reason to believe that F values met the above criterion of success for the inferential use to which he put them.

For the predictive use of the steady state equation, in this final phase, Wright apparently had a more stringent criterion of success, because he rejected the 1941 result of the predictive use as inadequate. Consistent with his aim of revising the original analysis, he demanded more accuracy in the fit between the theoretical and observed distributions of gene frequencies. It was not that the steady state equation itself was inadequate in 1941; rather, it was the way he used the model in 1941 that was inadequate. In 1941, Wright used the model by assuming the homogeneity of parameter values and failed to produce a theoretical distribution of gene frequencies that fit well with the data (see Figure 3.13). In the 1970s, he relaxed the homogeneity assumption, and for different subdivision of the *Linanthus* population, he used different sets of parameter values for the steady state equation. This use of the equation resulted in theoretical distributions of gene frequencies that fit the data quite well (see Figure 3.14).

3.8 CONCLUSION

The relationship between Wright's uses of three models (steady state equation, F-equation, and F-curves) that I have focused on and the aims of his research is summarized in Tables 3.6 and 3.7. As we can see, uses of the same model can shift over the course of scientists' research in response to the shift in aim. We have also seen that for each use of a model, certain criteria of success were relevant and that criteria of success for one use of a model can be different from those for another use of the same model.

Model: Steady State Equation

Phase	Use	Aim
1	To compute gene frequencies. To predict the distribution of gene frequencies. To infer evolutionary factors.	To illustrate isolation by distance and explain the observed patterns.
2	Not used.	To find evidence for evolutionary factors beside selection.
3	To compute gene frequencies. To predict the distribution of gene frequencies. To infer evolutionary factors.	To revisit phase 1 and explain the observed patterns more fully.

Table 3.6: A summary of Wright's uses of the steady state equation model and the aims of his research.

Model: F-Equation and F-Curves

Phase	Use	Aim
1	To compute F values. To describe the extent of local genetic differentiation. To infer effective N (F-Curves).	To illustrate isolation by distance and explain the observed patterns.
2	To compute F values. To describe the extent of local genetic differentiation. To infer effective N (F-Curves). To infer evolutionary factors (F-Curves).	To find evidence for evolutionary factors beside selection.
3	To compute F values. To describe the extent of local genetic differentiation. To infer the spatial heterogeneity of N_u and selection. (F-Curves not used.)	To revisit phase 1 and explain the observed patterns more fully.

Table 3.7: A summary of Wright's uses of F-Equation and F-Curves and the aims of his research.

4.0 REALISM, INSTRUMENTALISM, AND USES OF MODELS IN SCIENCE

4.1 INTRODUCTION: STEIN'S CONJECTURE

In his 1989 paper, Howard Stein argues that James Clerk Maxwell was an *instrumentalist* when he used his mechanical models of the electromagnetic field—the ether models—"as mere aids" (Stein 1989, 61) to develop a system of equations governing the dynamics of that field. At the same time, according to Stein, Maxwell was also a *realist* when he saw the ether models as providing clues about the true nature of the electromagnetic field (Stein 1989, 62–64). Referring to Maxwell's instrumentalist and realist attitudes toward the ether models, Stein says: "This dialectical tension . . . between a realist and an instrumentalist attitude, existing together without contradiction, seems to me characteristic of the deepest scientists" (Stein 1989, 64).[1]

Stein contrasts Maxwell with two other scientists, Henri Poincaré and Lord Kelvin, and claims that the latter two were not as successful as Maxwell in their work on electrodynamics because of the strong adherence to instrumentalism—in the case of Poincaré—and realism—in the case of Kelvin. According to Stein, Poincaré regarded the ether as a useful fiction for organizing our experience and did not take the ether model as a source of important hypotheses and questions about the electromagnetic field—for example, whether momentum exchange occurs with the ether (Stein 1989, 56). Kelvin, on the other hand, demanded that the ether model provide an accurate description of the nature of the electromagnetic field and could not take the model merely as a tool for developing a better system of equations. He remained skeptical of Maxwell's equations since he failed to find a satisfactory mechanical model of the ether (Stein 1989, 64).

[1]On Maxwell's realist and instrumentalist attitudes, see also Cat (2001, 437).

Unfortunately, Stein does not say exactly what he takes to be realism and instrumentalism, nor does he clarify what he means by the dialectical tension between these two positions.[2] But we can glean from his paper an important conjecture about scientific practice:

> STEIN'S CONJECTURE: In successful scientific research, a scientist uses a model according to the methodological principles of realism and instrumentalism despite the tension they create among the uses of the model.[3]

I will characterize this conjecture in detail shortly, but first let us ask why this conjecture would be important if it were true. It implies that there are methodological principles of realism and instrumentalism that are directly relevant to successful scientific practice. A philosophical task is to examine these principles and explain how they create a a certain tension among the uses of a model while contributing to the success of scientific research. But *methodological* principles of realism and instrumentalism are unrecognized in the philosophical debate over scientific realism, as the following list of representative formulations of various realist and instrumentalist positions show.

Scientific realism standardly takes the form of a metaphysical, semantic, epistemic, or axiological position:

[2] And he does not explain what he means by the "deepest" (Stein 1989, 64) scientists. Rather than try to make sense of the thorny notion of depth (of a scientist, research, explanation, or something else), I will simply replace Stein's talk of the deepest scientists with the talk of successful scientific research and a scientist working in it. Throughout this paper I only consider paradigmatically successful research, that is, research that should be regarded as successful by any notion of successful research in science.

[3] By 'successful scientific research' I mean the sort of research that should count successful in anyone's axiology, whether realist or not. For example, if there is a case of scientific research that contemporary scientists themselves regard as one of the most successful in history, we have prima facie reason to think that this case should count as successful in anyone's axiology. Thus, Stein's conjecture should be compatible with both axiological realism and instrumentalism that I present below.

In addition to his views on Maxwell, Poincaré, and Kelvin mentioned above, my formulation of the conjecture is inspired by Stein's remark that the true contrast between realism and instrumentalism is methodological rather than metaphysical (Stein 1989, 56). Although I am using Stein's work as an inspiration and do not claim that my formulation is the correct interpretation of his view, here I offer a brief exegesis in favor of my formulation. I think Stein sees realism and instrumentalism as being opposed because he says that they stand in a "dialectical *tension*" (Stein 1989, 64; my emphasis) and thinks that Poincaré and Kelvin exhibited the opposite attitudes. But it is misleading to say that realism and instrumentalism coexist "*without contradiction*" (Stein 1989, 64; my emphasis), because this locution may suggest that realism and instrumentalism are not opposed at all. I suggest we interpret Stein as saying that realism and instrumentalism can coexist—hence there is no strict contradiction between them—despite the differences that create a certain tension between them. Understood in this way, Stein's conjecture is different from the view that realism and instrumentalism amount to the same position (e.g., Nagel 1961).

1. *Metaphysical realism*: The real world, which scientists investigate, exists independently of how scientists theorize about it (Boyd 1983, 45; Psillos 1999, xix; Niiniluoto 1999, 21; Chakravartty 2013).[4]

2. *Semantic realism*: What a theory or model says about the world is to be interpreted as truth-conditioned claims about the world (Boyd 1983, 45; Psillos 1999, xix; Chakravartty 2013).

3. *Epistemic realism*: Predictive success of a scientific theory or model is evidence that when interpreted as describing the mind-independent world, the theory or model is approximately true of the world. Consequently, we are justified in believing what a predictively successful theory or model says about the world (Boyd 1983, 45; Psillos 1999, xix).[5]

4. *Axiological realism*: The aim of science is to construct a theory or model that is true of both observable and unobservable parts of the world (van Fraassen 1980, 8; Niiniluoto 1999, 160; Godfrey-Smith 2003, 176; Lyons 2005).[6]

Instrumentalism is compatible with metaphysical realism (Fine 1986, 156) but is opposed to one or more of semantic, epistemic, and axiological realist positions:

1. *Semantic instrumentalism*: What a theory or model says about the unobservable part of the world is not truth-conditioned and is merely a linguistic devise to efficiently make inferences among claims about the observable part of the world (Stanford 2006, 191–192).[7]

2. *Epistemic instrumentalism*: Predictive success of a scientific theory or model is evidence that the theory or model is a reliable instrument for achieving our practical ends rather than evidence that it is approximately true of the world. Consequently, we should only

[4]Alternatives to metaphysical realism are the views that reject the existence of the external world (e.g., idealism) or the mind-independence of the world (e.g., neo-Kantianism and constructivism) (see Boyd 2010, Chakravartty 2013).

[5]For slightly different formulations, see Niiniluoto (1999, 79) and Chakravartty (2013). Since the epistemic realist about a given scientific theory or model would believe that the entities posited by the theory or model exist, we can say that she is also a metaphysical realist about those entities.

[6]In the context of the realism debate, this position is axiological because it is specifically about the aims of science. This label should not imply that consideration of aims, norms, or values does not matter to other positions on the list. Epistemic or cognitive norms matter to epistemic realism, but this fact does not turn epistemic realism into axiological realism since epistemic or cognitive norms may be different from the aims of science.

[7]This is a rough statement of instrumentalism concerning the language of science, and the linguistic variety of instrumentalism has met serious criticisms (see Psillos 1999, 17–39; Stanford 2006, 188–193.

make practical use of a predictively successful theory or model without believing what it says about the world (Fine 1986, 156–157; Stanford 2006, 193–194).[8]

3. *Axiological instrumentalism*: The aim of science is to construct a theory or model that is instrumentally reliable (Fine 1986, 157).[9]

This list of familiar positions in the realism debate does not contain any distinctly methodological principles.[10] Their absence is noteworthy because one of the motivations behind the realism debate is a desire to understand how successful science works. If Stein's conjecture were true, it would imply that there are methodological principles of realism and instrumentalism that matter to successful scientific practice and which deserve careful philosophical analysis.

In this chapter I flesh out Stein's conjecture, provide both philosophical and historical arguments for it, and explore the implication of the conjecture for the realism debate. I shall begin by characterizing the realist and instrumentalist methodological principles and the subtle tension that emerges among a scientist's uses of a model over time if the scientist adopts both principles (Section 4.2). I then present a detailed case of successful scientific research—that of Seymour Benzer's in the 1950s and 60s—and argue that various uses of the very same model in this case followed both the realist and instrumentalist methodological principles. I also show that the tension of the sort I characterize in Section 4.2 emerged in Benzer's case (Sections 4.3–4.5). Finally, I argue

[8]The very last clause of this formulation of epistemic instrumentalism faces a serious problem identified by Kyle Stanford: "The central problem is that to use a theory for prediction, intervention, and other practical ends just *is* to believe at least *some* of the things it tells us about the world" (Stanford 2006, 194). Stanford proposes an improved version of epistemic instrumentalism according to which we are to believe what a given theory or model says about the world insofar as we can understand it in terms of *another* theory or model—including our commonsense claims about everyday experience—about which we are epistemic *realists* (Stanford 2006, 197–211). My argument in this chapter works for either formulation of epistemic instrumentalism.

[9]This position assumes that an instrumentally reliable theory or model need not be true of the unobservable parts of the world.

[10]The absence of methodological positions in the above list is not an oversight on my part. None of the most comprehensive lists of realist theses includes a methodological thesis (see, e.g., Boyd 1983, 45; Leplin 1984, 1–2; Fine 1986, 156–157; Kitcher 1993, 127; Psillos 1999, xix; Chakravartty 2013). Ilkka Niiniluoto's (1999, Ch. 6) discussion of methodology as part of his comprehensive account of realism appears to be an exception that proves the rule. For Niiniluoto, methodological rules are realist insofar as they promote the attainment of truth, which is the aim of science identified by axiological realism. To find out which methodological rule promotes the realist aim, we have to be able to measure its effectiveness in attaining truth. Thus, Niiniluoto devotes much of his discussion to various indicators of truth or truthlikeness. His conception of realist methodology is different from mine in that I regard a methodological rule as realist if it tells a scientist to do things that display their commitment to investigating the true nature of the world. More recently, Hasok Chang urges reconception of the realism debate in terms of what he calls active scientific realism, according to which "science should strive to maximize our contact with reality and our learning from it" (Chang 2012, 205). In discussing this axiological realism, Chang discusses controlled experiments or tests and operationalization of concepts as methods for learning from reality (Chang 2012, 221–222).

that epistemic realism or epistemic instrumentalism by itself prohibits a scientist from adopting both the realist and instrumentalist methodological principles. Stein's conjecture thus poses new challenges to realists and instrumentalists, and I briefly suggest possible avenues of response that realists and instrumentalists may take (Section 4.6).

4.2 REALISM AND INSTRUMENTALISM AS METHODOLOGICAL PRINCIPLES

The realism and instrumentalism relevant to Stein's conjecture may be understood as methodological principles about how a scientist should use a model in the ongoing process of research.[11] I formulate these principles, using the concept of scientific model developed in Chapter 2:

> SCIENTIFIC MODELS AS EXTERNAL REPRESENTATIONS: A given external representation is a scientific model if a scientist interprets some part or property of the representation as representing (in the sense of carrying information about) some part or property of the object of an investigation and if the scientist uses the representation as a tool to approach the scientist's specific research questions or aims.

The key part of this concept that becomes important below is that to count as a model, only *some* part or property of an external representation needs to represent some part or property of the object of an investigation. In other words, a model can have other parts or properties that are not or have not yet been interpreted as representing something. This concept applies to a research project in science that has an object of investigation, which I assume is a part or parts of the world—such as entities, events, processes, or phenomena—and which a scientist involved in that project primarily investigates. Her specific research questions and aims have to do with the object of her investigation.[12]

[11] Stein says that his own analysis of Poincaré's instrumentalism "could be recast in *methodological terms*—in effect, that the true issue was not [metaphysical] realism vs. instrumentalism" (Stein 1989, 56; my emphasis).

[12] Our identification of the object of a given investigation should be specific enough to help us understand what exactly a scientist is investigating, what questions she is trying to answer, and what research aims she has. For example, to describe what a biologist running a cell biology lab investigates, it would be uninformative, though true, to say that she studies cells; it would be more informative to say that she studies the process by which functional cells become cancerous in adult human beings, that her question is what causes this process to stop, and that her aim is to develop a technique to control this process. So, assuming that there are no better alternatives, we should say that this process is the object of her investigation.

With this characterization of model in mind and drawing on Stein's remarks on Maxwell, Kelvin, and Poincaré, I formulate two methodological principles:

> THE REALIST METHODOLOGICAL PRINCIPLE: A scientist should use a model to make claims—in the form of hypotheses, descriptions, or explanations—and generate questions about the true nature of the object of an investigation. To do this, whenever possible, the scientist should interpret parts and properties of the model as representing the parts and properties of the object of the investigation.[13]
>
> THE INSTRUMENTALIST METHODOLOGICAL PRINCIPLE: A scientist should use a model to organize or predict observational and experimental results, to design new techniques for intervention, and to do other tasks that address practical needs. To do this, the scientist should interpret parts and properties of the model as representing things other than the parts and properties of the object of their investigation whenever such interpretation would help the scientist perform the task.[14]

The first principle is "realist" because it shares with other forms of realism the emphasis on truth about the world, and the second principle is "instrumentalist" because it shares with other forms of instrumentalism the emphasis on human needs.[15] These principles tell a scientist *how to use a model as a tool* in her research rather than *what model to ultimately aim for*. In this sense, these principles are methodological rather than axiological.

A scientist's use of a model, if guided by both the realist and instrumentalist methodological principles, exhibits a characteristic tension that emerges in the ongoing process of her investigation. I will present an actual case of this tension in the subsequent sections, but to get an idea of what this tension looks like, let us consider a scientist who interprets one of the properties of an external representation as representing a property of the object of her investigation and uses the representation to answer her research question. This external representation is her model. Suppose

[13]This does not tell a scientist to interpret *every* part and property of her model as representing the parts and properties of the object of her investigation. She may simply ignore certain parts or properties of her model because a model often has parts or properties that no scientist is likely to interpret as representing something. For example, a scientist is likely to ignore ink and paper out of which a diagrammatic model is made.

[14]This does not tell a scientist to interpret *every* part and property of her model as representing things other than the parts and properties of the object of her investigation. Doing so implies that the object she is using is not a model of the object of her investigation. Thus, even if she follows the instrumentalist methodological principle, she still interprets at least some part or property of her model as representing some part or property of the object of her investigation.

[15]Here I am following Arthur Fine's discussion of the distinguishing themes of realism and instrumentalism (Fine 1986, 150–157; Fine 2001, 110–113).

that at one point in the course of her investigation she uses her model to describe the true nature of the object of her investigation by interpreting many parts and properties of her model as representing the parts and properties of the object of her investigation. We can say that she follows the realist methodological principle. At some future point, however, she uses her model to organize experimental data by interpreting many of the properties of her model *that she previously interpreted as representing properties of the object of her investigation* as representing things other than the object of her investigation, such as quantities recorded in experiments or patterns of experimental data. We can say that she follows the instrumentalist methodological principle. When we see the temporal course of her research activities in this way, we can see a characteristic tension among her uses of the model: After using her model according to the realist methodological principle, she may *cancel* or *disregard* some of the interpretations she gave to her model when she uses her model according to the instrumentalist methodological principle. Thus, while she is using a model according to one of the methodological principles, she cannot use the same model to do things she could do were she following the other principle. The adoption of both principles becomes apparent when we study how a scientist's interpretations of parts and properties of a model and her uses of the model change over the course of research. Historical studies of scientific practice are thus useful resources for philosophers of science interested in methodology, and below I shall use my own historical study of scientific practice to show how the realist and instrumentalist methodological principles were jointly adopted in actual scientific practice.

Stein's conjecture says, then, that in successful scientific research, a scientist uses a model according to both the realist and instrumentalist methodological principles formulated above despite the tension the adoption of both principles creates among the uses of the model. Let us now turn to a detailed case of successful scientific research that supports this conjecture.[16]

[16]In his *Theory and Truth*, Lawrence Sklar (2000) argues that the historical development of fundamental physics owes much to anti-realist criticisms of realist attitudes towards fundamental theories. If Sklar is right, there appears to be the coexistence of realism and anti-realism that mattered to the successful development of physical theories. This thesis resembles Stein's conjecture although Sklar does not cite Stein's paper, and Sklar's discussion can provide additional support for Stein's conjecture. I am grateful for Kyle Stanford for drawing my attention to the relevance of Sklar's work to Stein's.

4.3 EVALUATING STEIN'S CONJECTURE: THE CASE OF SEYMOUR BENZER'S RESEARCH

A good case to evaluate Stein's conjecture must meet two criteria. First, it must be an instance of successful scientific research. To identify such a case, we can turn to the history of science and identify past research projects that turned out to be uncontroversially successful. Second, detailed historical records for the case must be available so that we can analyze how a scientist used a model (or a family of models) over time. Stein's conjecture becomes plausible to the extent that the scientist's uses of a model followed the realist and instrumentalist methodological principles and exhibited the sort of tension described above. In what follows I present a detailed case that meets these criteria, and in Section 4.4.3, I will return to the general philosophical argument that this case supports.[17]

4.3.1 Background to Benzer's Research

One year after Watson and Crick's (1953a) development of the double helix model of DNA, Seymour Benzer began experimental work that would produce a series of seminal contributions to molecular biology.[18] For example, Benzer showed experimentally that a gene identified by its association with a physiological function has a fine, internal structure because mutation and recombination can occur within the gene. In other words, a gene has internal sites where mutation and recombination can occur. This result was incompatible with the then traditional concept of a gene as an indivisible unit of function, mutation, and recombination. Benzer thus suggested the concepts of cistron, muton, and recon for units of function, mutation, and recombination, respectively. He then went on to estimate the sizes of a cistron, muton, and recon in terms of the number of base pairs in DNA. For instance, he estimated a recon to be no more than two base pairs (see, e.g., Benzer 1957). That is, Benzer provided a thoroughly molecular characterization of what ge-

[17] My choice of a case reflects both my expertise in biology and my prior acquaintance with the case and the detailed historical work by Holmes (2006). It should be possible to find additional cases.

[18] For excellent intellectual biographies of Benzer, see Weiner (1999) and Holmes (2006). Weiner gives a highly accessible account of Benzer's research on genetic fine structure (Weiner 1999, 46–60). Holmes has reconstructed in great detail the first two years (1954–1956) of Benzer's research on genetic fine structure (Holmes 2006, 179–298). Holmes had access to Benzer's personal papers before they were donated to the Caltech Archives, and my analysis of Benzer's research draws on the same personal papers as well as Holmes's pioneering study.

neticists had traditionally called the gene. Moreover, Benzer provided experimental evidence that the structure of DNA is linear at the molecular level, which supported the Watson-Crick model of DNA (see, e.g., Benzer 1960, 1962).[19] Fortunately we can reconstruct Benzer's research activities in detail by using rich archival materials pertaining to his work.[20] Benzer's research is thus a good case for evaluating Stein's conjecture.

Benzer's research began with his discovery of certain properties of bacteriophage T4, a virus that infects and eventually kills *Escherichia coli*. Once a T4 phage infects a bacterium by injecting its genetic material into the host cell, the genetic material is multiplied inside, and new T4 phages develop. At the end of this life cycle, the phages make the host cell lyse (i.e., burst open) and release them into the environment. When lysis occurs, the wild-type phage T4 produces small plaques with rough edges.[21] Some mutants of phage T4 make the host cell lyse rapidly and produce large plaques with sharp edges (Figure 4.1). These mutants are called r mutants (r for rapid lysis) and are results of mutations in certain regions of the genetic material of T4 (Benzer 1955, 346).[22]

Mutations in one region of the genetic material of T4, called the rII region, produce r mutants with another trait that distinguishes them from the wild type: the rII mutants produce plaques on *E. coli* strain B but not on strain K12(λ), whereas the wild type produces plaques on both strains (Figure 4.1; Benzer 1955, 346). These properties of the rII mutants make them a useful experimental system to carry out detailed genetic analysis. One can easily isolate r mutants on strain B, where they produce distinctive plaques, and discriminate two r mutants by recombination experiments in which one infects K12(λ) with the two r mutants and lets them exchange their genetic material. If they are distinct r mutants, a wild-type recombinant will arise by genetic recombination and produce the wild-type plaques on K12(λ). If they are identical mutants, no plaques will form. K12(λ) thus serves as a highly selective host strain, enabling one to detect very rare wild-type recombinants and hence subtly distinct r mutants (Benzer 1955, 346–347).

[19]The importance of Benzer's research can also be gauged by the extent to which it was covered in the early textbooks in molecular biology (see, e.g., Watson 1965, 231–237; Stent 1971, 362–375). For a summary of Benzer's important contributions to molecular biology by a contemporary biologist, see Greenspan (2009, 7–10).

[20]Seymour Benzer Papers, 10242-MS, Caltech Archives, California Institute of Technology. The collection is organized into 126 boxes, each of which is organized into folders. The finding aid is available online (http://www.oac.cdlib.org/findaid/ark:/13030/c8vh5ptc/). Hereafter when I cite Seymour Benzer Papers, I write "SBP" followed by the box and folder numbers.

[21]A plaque is a clear area, that is, area without host cells, in the lawn of bacteria.

[22]Hershey (1946) first isolated r mutants in phage T2, which is related to T4.

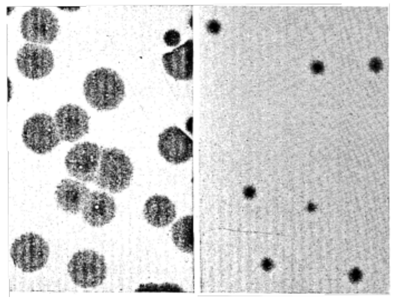


Figure 4.1: rII mutants. *Left*: Compared to the small wild-type plaques, rII mutants produce large plaques on *E. coli* strain B. *Right*: But rII mutants do not produce any plaques on strain K12(λ). Only the wild-type phages produce plaques. Reproduced, with permission, from Benzer (1962, 72). Copyright © (1962) Scientific American, Inc. All rights reserved.

4.3.2 A Genetic Map as a Model of DNA

Recombination experiments can be used to construct what is called "a genetic map." In a typical linear map, each point on a line represents an organism that is genetically mutated from the wild type, and a mutant is treated as a marker of that part of the genetic material which is altered from the wild type. The distance between two points represents the observed frequency among the progeny of the two mutants that exhibit the wild-type phenotype. Before Benzer began mapping the rII region of T4, A. H. Doermann and M. B. Hill produced a linear map of two r mutants named r47 and r51 (Figure 4.2). The resolution of the map depends on the extent to which the experimental system allows the researcher to detect rare wild-type recombinants. In this regard, Benzer's rII system was extremely powerful.[23]

Benzer made his first genetic map of rII mutants on May 10, 1954, displaying Doermann's r_{47} and r_{51} and the new mutant Benzer isolated named "V" (Figure 4.3).[24] Benzer's rules of mapmak-

[23]Benzer discovered the above properties of the rII mutants in the first half of 1954. Because of the lack of space, I will not describe Benzer's early experiments with rII mutants. For a detailed account of his work during this period, see Holmes (2006). Benzer also gave a brief autobiographical account (Benzer 1966).

[24]Doermann had sent Benzer his stocks of previously mapped mutants including r_{47} and r_{51} (A. H. Doermann to Seymour Benzer, April 3, 1954, SBP 67.6).

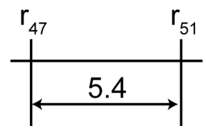


Figure 4.2: A genetic map of two r mutants of phage T4. "r_{47}" and "r_{51}" designate mutants and the distance between them the percentage recombination frequency (5.4% here). Redrawn with modification from Doermann and Hill (1953, 87).

ing are explained in Figure 4.4. According to these rules, every mutant was to be represented as a *point* on the horizontal line.

Benzer consistently interpreted particular parts of a genetic map as representing parts of DNA and their properties. Given a map like Figure 4.3, he interpreted the the point corresponding to a mutant (say, r_{47}) as representing a change in a particular part of DNA as well as that part which had changed. The map point labeled r_{47} represented a part of DNA and its property of being mutated from the wild type. In one paper Benzer described his interpretation of a map as follows:

> A genetic map is an image composed of individual points. Each point represents a mutation [i.e., a change in a part of DNA] which has been localized with respect to other mutations by recombination experiments. The image thus obtained is a highly colored representation of the hereditary material.[25] (Benzer 1957, 72)

A genetic map was thus Benzer's model of DNA.

4.4 ANOMALOUS MUTANTS, USES OF A GENETIC MAP, AND THE INSTRUMENTALIST METHODOLOGICAL PRINCIPLE

By the end of 1954 Benzer was immersed in the project of bringing the resolution of a genetic map down to the molecular level. In particular he was trying to estimate in molecular terms the lengths of the genetic units of recombination, mutation, and physiological function and was trying to do

[25] It is not clear to me what he means by "highly colored" here, but this does not affect the main point I am making.

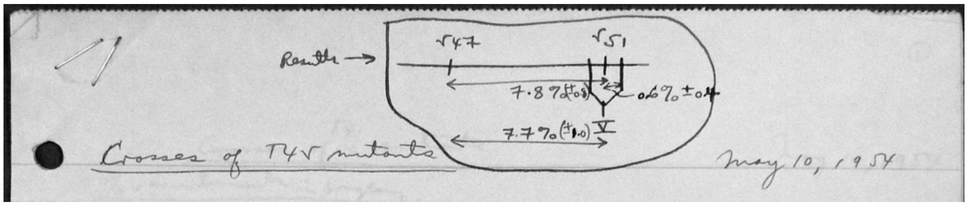


Figure 4.3: Benzer's first genetic map. The recombination frequency between r_{47} and r_{51} that Benzer obtained from his experiments was slightly different from the value Doermann and Hill reported (Figure 4.2). Reproduced from Seymour Benzer Papers, Box 67 Folder 6, 10242-MS, Caltech Archives, California Institute of Technology. Courtesy of the Archives, California Institute of Technology.

so by converting the length on a map, which was in the unit of recombination frequency, into the length of DNA measured by the number of base pairs (Benzer 1955, 345).[26] For this purpose he needed to isolate and map a great number of rII mutants (Holmes 2006, 258).

In this and the next section I follow Benzer's uses of a family of genetic maps in detail. In this section I show that some of Benzer's crucial uses of a genetic map followed the instrumentalist methodological principle and emphasize that, without these uses of his model, Benzer's research would not have enjoyed the success it actually did. To do this, I have to describe Benzer's discovery of the so-called "anomalous mutants" and how Benzer accounted for them in his map. Then in the next section I show that some of Benzer's crucial uses of a genetic map followed the realist methodological principle, and at the end of that section I discuss the characteristic tension among Benzer's uses of a genetic map created by his adoption of both the realist and instrumentalist methodological principles.

[26]Benzer thought about the correspondence between map distance and physical distance as early as May 28, 1954 when he wrote "thoughts on the gene" (SBP 67.6). For discussion of his early thoughts on this topic, see Holmes (2006, 212–220).

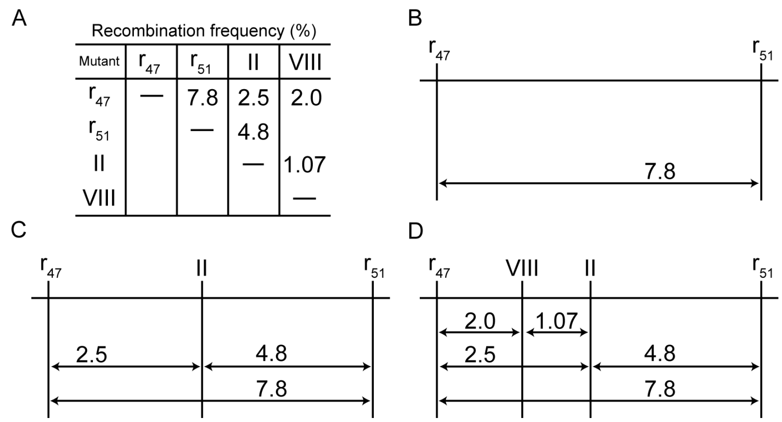


Figure 4.4: Rules of mapmaking. (A) Benzer frequently used a table to summarize recombination data obtained from crosses between mutants on the top row and those on the left column. The table is an abbreviated version of Benzer's table made on May 17, 1954 (SBP 67.6). (B) *Rule 1*: Draw a horizontal line and two vertical lines near the ends. Let the vertical lines represent two mutants that exhibited the greatest recombination frequency on the table, and draw a double-headed arrow between the vertical lines with a label indicating the recombination frequency between the mutants represented. (C) *Rule 2*: Consider other mutants that were crossed with one of the two mutants just put on the map, say, r_{47}, and find the mutant that showed the greatest recombination frequency with r_{47}. In this case, this mutant is the one named "II," and it is to be put on the map as before. Since the recombination frequency between II and r_{51} is also shown on the table, this fact is also on the map. (D) *Rule 3*: Apply Rule 2 to the mutants not yet put on the map and repeat the procedure until all the mutants and data shown on the table are put on the map. Note that the resulting map fails to show strict additivity of distances, but Benzer's practice was not to make the map show additivity and instead to make the map faithful to the data.

4.4.1 Anomalous Mutants

On February 23, 1955, Benzer crossed four rII mutants: 47 and 168 with 295 and 312. These double crosses showed few plaques, making Benzer write: "obviously the four mutants are not all allelic" (SBP 67.8). In other words, these mutants were not identical. On the 24th, Benzer performed "allelism tests" (SBP 67.8). In these tests, he made ten crosses among five rII mutants—47, 168, 295, 312, and 145—and counted the number of wild-type plaques. No plaques were observed in four of the ten crosses, namely, those between 47 and 168, 47 and 295, 47 and 312, as well as 295 and 312. According to the basic rules of mapmaking (Figure 4.4), these four mutants should occupy the same point on a genetic map. But two crosses among these mutants, namely, those between 168 and 295, and 168 and 312, *did* produce plaques: 168, 295, and 312 should occupy different points on a map. These four mutants showed, to use Benzer's own word, "anomalies" (Benzer 1955, 351).[27]

Benzer put these results on a map (Figure 4.5), and it showed nonzero distances within zero distances. For example, in the left part of the map, Benzer placed 168 in between 312 and 47. He drew 312 and 47 as well as 168 and 47 as having zero distances, but he drew 312 and 168 as having a non-zero distance. As noted above, for Benzer, a genetic map was a model of DNA; in particular, the space between two points on a genetic map was a representation of the physical space between two mutational sites of DNA. Thus, a genetic map showing nonzero distances within zero distances was suggesting a physical impossibility.

Apparently realizing this potential problem, Benzer drew a revised map on a piece of paper and stapled it over the original map (Figure 4.5).[28] In this map he drew 47, 312, 295, and 168 as overlapping horizontal bars. He put 47 at the very bottom and drew 295 above it. Leaving some space next to 295, he drew 168, making it overlap with 47. He drew 312 as overlapping with 295 and 47, but not 168.[29]

[27] Anomalous mutants also exhibited little or no tendency to revert to the wild type (Benzer 1955, 351).

[28] This revised map was Benzer's second attempt, and underneath this map we can find his first attempt where he drew r47 as a horizontal bar at the bottom and 312, 295, and 168 as vertical lines above it. He indicated that each of these three mutants had a zero distance to different positions of the horizontal bar that represented 47.

[29] Referring to the revised map, Holmes says: "This critical move appears to mark the point at which Benzer recognized that these four mutants could not be point mutations but must be alterations extending along a portion of the chromosome" (Holmes 2006, 271–272).

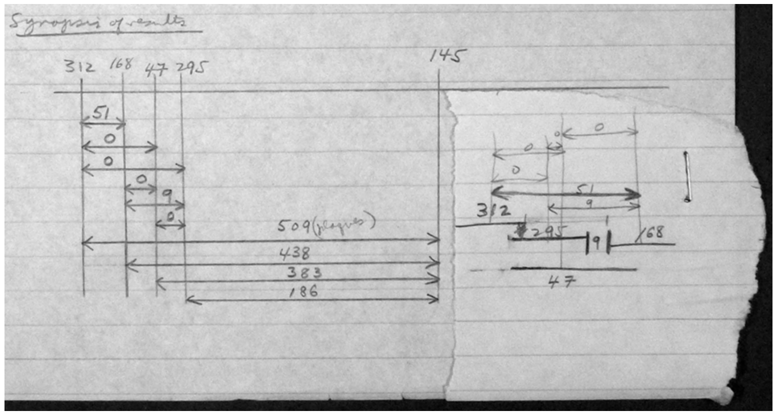


Figure 4.5: Benzer's map of anomalous mutants drawn on February 24, 1955. Reproduced from Seymour Benzer Papers, Box 67 Folder 8, 10242-MS, Caltech Archives, California Institute of Technology. Courtesy of the Archives, California Institute of Technology.

On the 26th, Benzer did the same ten crosses he had done on the 24th, and this time he calculated recombination frequencies. Crosses among the four anomalous mutants—47, 312, 295, and 168—still showed incoherent results. Benzer drew a map on a graph paper (Figure 4.6). He represented anomalous mutants as horizontal bars at the top of the map, the practice he would continue to use. But he also noted that the distance between 295 and 145 "does not fit this scheme" (SBP 67.8), because the map represented the distance of 0.78% between 295 and 145 as longer than that of 0.99% between 47 and 145 as well as 1% between 168 and 145.[30]

4.4.2 Mapping Anomalous Mutants

Benzer's response to the discovery of anomalous mutants was to change the rules of mapmaking so that the experimental results concerning these mutants can be represented differently than those concerning other mutants that he was interested in. That is, he changed the rules for making his model of DNA. The new rules are explained in Figure 4.7; the most important change is Rule 2*,

[30] At the bottom of this map, Benzer provided a tentative physical interpretation of map distances by indicating that 1 cm on this map corresponded to the maximum of 25 nucleotide pairs. Alternatively, Holmes (2006, 272) interpreted that Benzer meant that the distance between 295 and 168 is 25 nucleotide pairs, but Benzer's annotation at the bottom of the page does not correspond to the gap between the horizontal bars representing 295 and 168 at the top of the page.

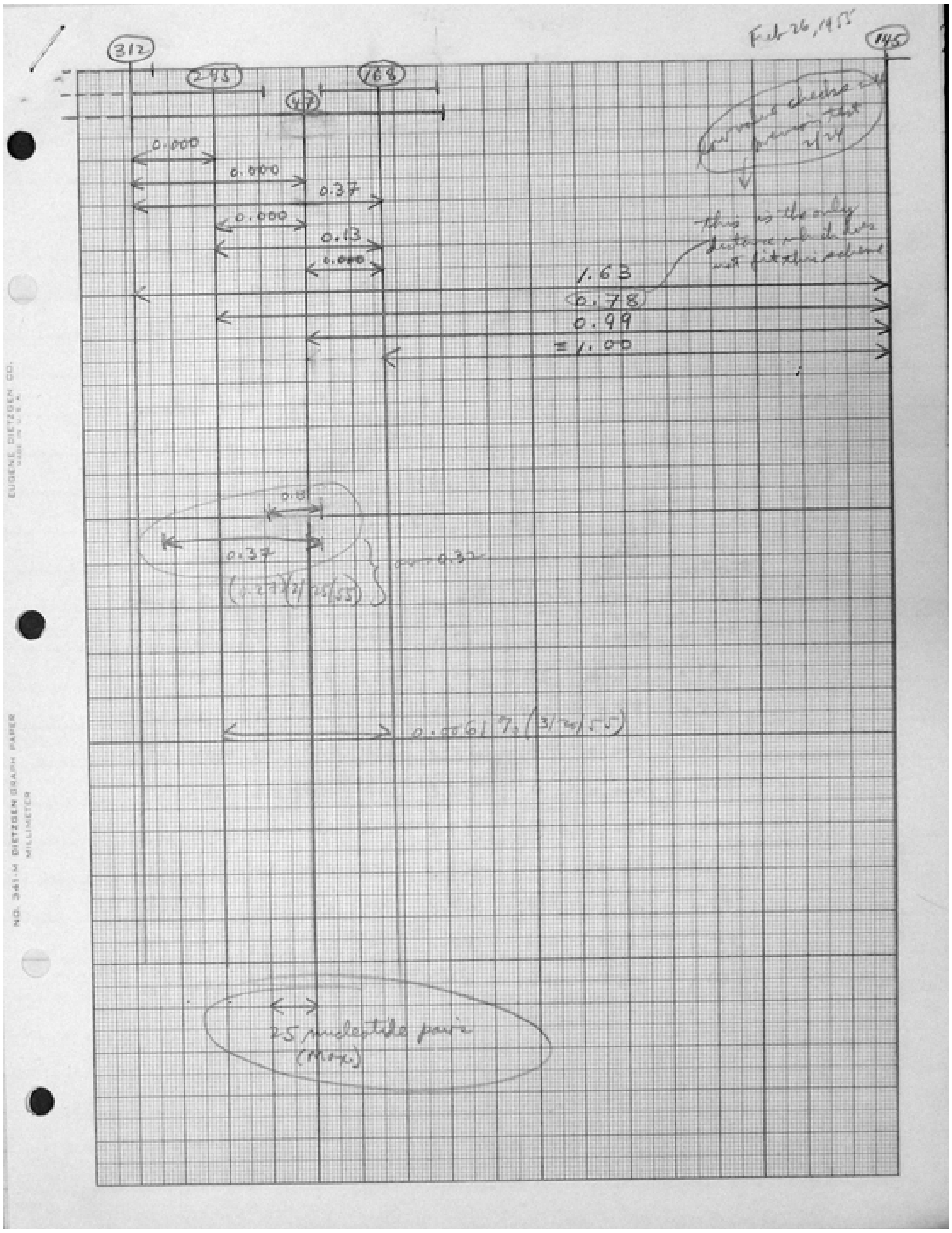


Figure 4.6: Benzer's genetic map dated February 26, 1955. Reproduced from Seymour Benzer Papers, Box 67 Folder 8, 10242-MS, Caltech Archives, California Institute of Technology. Courtesy of the Archives, California Institute of Technology.

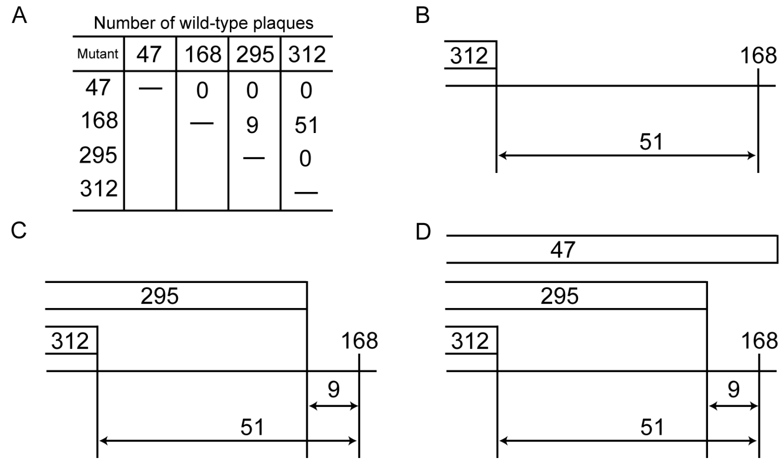


Figure 4.7: Rules of mapmaking for anomalous mutants. (A) A subset of Benzer's experimental results on February 24, 1955 (SBP 67.8). The top row and the leftmost column show the mutants crossed. The anomalous mutants are those that fail to recombine with two or more mutants: 47, 295, and 312. The data are presented (as Benzer did) in the observed number of wild-type plaques rather than percentage recombination frequencies, but the unit of data does not affect the rules. (B) *Rule 1*: Same as before (see Figure 4.4), but if a mutant being drawn is anomalous, then it should be drawn as a bar. Here 168 is on the right of this map because of the results of crosses with 145 shown in Figure 4.5. (C) *Rule 2**: Consider other mutants that produced wild-type plaques with one of the two mutants just put on the map, say, 168, and find the mutant that produced most wild-type recombinants with 168. In this case, the relevant mutant is 295. But 295 does not recombine with the other mutant already mapped, namely, 312. Thus, 295 is to be drawn as a horizontal bar above and covering 312, and the space between the right hand end of 295 and 168 is marked with the number of wild-type plaques found in the cross between these two mutants. (D) *Rule 3*: Apply Rule 2 or Rule 2* to the mutants not yet put on the map and repeat the procedure until all the mutants and data shown on the table are put on the map. The new reading convention is that the mutant represented as a bar does not produce wild-type recombinants with any of the mutants that overlap with it. For Benzer's published version of a map corresponding to (D), see Benzer (1955, 352, Figure 5 A).

which, unlike Rule 2 in Figure 4.4, says that anomalous mutants should be drawn as horizontal bars. Benzer did speculate about the nature of these anomalous mutants (e.g., these mutations extended over a long stretch of DNA). Although he had no evidence regarding any of the hypotheses he came up with, he apparently assumed that anomalous mutants were unlikely to be "true 'point' mutations (i.e., involving an alteration of only one nucleotide pair)" (Benzer 1955, 351). Thus, Benzer said that for the purpose of making a high resolution genetic map, that is, a map that would resolve the distance between two neighboring "point" mutants, "it would seem well advised to employ only mutants for which some reversion is observed [i.e., mutants that are not anomalous]" (Benzer 1955, 351).

Since December of 1955, however, Benzer would find productive uses of the map of anomalous mutants. On December 2, 1955, Benzer made crosses among some anomalous mutants.[31] He drew a diagram entitled "grouping up to date" (Figure 4.8). He found three mutants to be non-overlapping, and six mutants to be overlapping. On December 5, he tested other mutants to see if any of them are "under 196" (SBP 67.9), the anomalous mutant drawn on the left of Figure 4.8. Benzer took twenty mutants, crossed them with 196, and classified mutants that showed low recombination frequencies as "under 196" and others as "not under 196." Based on the surviving notes, this was the first time Benzer used the phrase "under 196." It referred to the way in which mutants were represented on a map: 196 was an anomalous mutant represented as a horizontal bar on top, and certain mutants were placed *under* it. In this sense, mutants were called "under 196."

What Benzer invented was a way to organize anomalous mutants relative to each other on a map. As shown in Figure 4.8, some anomalous mutants were represented as lying next to each other on a map, while others were represented as lying under another mutant. Such map was different from a map like Figure 4.2 in that it did not represent distances between mutants as measured by recombination frequencies. Rather it represented which mutant is next to which one: it represented relative positions of mutants. In the next several years, along with other experiments, Benzer constantly looked for new anomalous mutants and tested them against existing anomalous mutants.

Benzer eventually produced a genetic map with many *segments*, each of which is defined as the stretch of the map that only one of a given set of anomalous mutants covers. Consider, for example,

[31] Benzer's work described in the paragraph below was first reconstructed by Holmes (2006, 289–290).

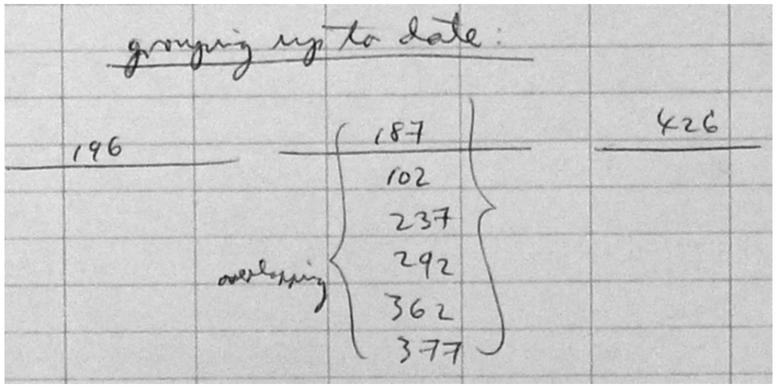


Figure 4.8: Anomalous mutants tested on December 2, 1955. Reproduced from Seymour Benzer Papers, Box 67 Folder 9, 10242-MS, Caltech Archives, California Institute of Technology. Courtesy of the Archives, California Institute of Technology.

the mutants 1272 and 1756 in Figure 4.9, which shows Benzer's working map of new anomalous mutants tested on September 21, 1959. For simplicity, assume that the right hand ends of these mutants occupy the same map position. Given these two mutants, a map segment is defined as the space that begins at the left end of 1272 and ends at the left end of 1756, that is, the space covered only by 1272. This segment can be further divided if there is a mutant whose left end falls in the middle of the segment. Thus, in Figure 4.9, Benzer indicated 1589 and 1605 as "exciting!" because they appeared promising mutants to further divide the segment defined by 1272 and 1756. Near the end of his mapping project Benzer had a map of anomalous mutants that divided the rII region into 80 segments, which is too large to reproduce here (Benzer 1962, 78–79).[32]

[32] In this section I entirely glossed over Benzer's idea that the rII region is divisible into two segments A and B, which function differently during the growth of phages in *E. coli* K12(λ). These segments came to be called A and B cistrons (Benzer 1957), and in Figure 4.9 their boundary was marked by a horizontal dashed line. Here I briefly explain Benzer's idea for the interested reader. Recall that unlike the wild type, the rII mutants do not produce progeny on K12(λ). If a rII mutant and a wild-type phage together infected the same host cell, *both* types of phage were found among the progeny. This suggested that the wild type supplied the necessary function for intracellular growth that the rII mutant could not perform. By the beginning of 1955, Benzer found that some pairs of rII mutants, when they together infected K12(λ), produced a lot of progeny, while other pairs produced little or none (Holmes 2006, 257–258). This suggested that two mutants forming the former type of pair had different functional defects so that they could supply each other, without recombination, the function that was defective in the other. On the other hand, two mutants forming the latter type of pair had the same functional defect and could produce progeny only when recombination occurred. Benzer thus concluded that the rII region has two functional segments. Doermann's r47 is on segment A and r51 on segment B (Benzer 1955).

Benzer used the map of anomalous mutants to design a method for rapid mapping of rII mutants. The new method allowed Benzer to classify a new rII mutant into a short segment of the rII region, thereby eliminating the need to cross every new mutant with all the other mutants in order to determine its position on a genetic map. Given the need to map thousands of mutants, this method was crucial to the success of Benzer's research.[33] To understand this new method, let us look at Figure 4.10, which is a published map of anomalous mutants as of April 1960.[34] Suppose we want to map a new rII mutant. Our first step is to cross the new mutant with the anomalous mutants 1272, 1241, J3, PT1, PB242, A105, and 638—called "big seven" (Benzer 1961, 407)—which were represented as very long bars at the top of Figure 4.10. If we find that the new mutant recombines with all the mutants except 1272, then we classify it as being under 1272: the new mutant belongs to the segment between the left end of 1271 and that of 1241, which is labeled as "A1" at the bottom of Figure 4.10. Our second step is to construct another experiment with 1272, 1364, and EM66, which divide segment A1 into three shorter segments. If we find, say, that the new mutant recombines with EM66 but not with 1272 or 1364, then we say that the new mutant is under 1364: the new mutant belongs to the segment between the left end of 1364 and that of EM66, which is labeled as "A1b1" at the bottom of Figure 4.10 (Benzer 1961, 406–407). If we want to determine a more precise map location of the new mutant, we only need to do recombination experiments with the mutants that also belong to A1b1 and calculate recombination frequencies.

[33] If he had only a handful of mutants to map, it would not have been so important to eliminate the need to cross every mutant with every other. But this was not the case, and Benzer's new method was so important that Gunther Stent later wrote:

> It is fair to say that without this astute exploitation of deletion [i.e., anomalous] mutants for rapid mapping, our knowledge of the genetic fine structure of the phage genome would still be very rudimentary; progress would have been hamstrung by the geometric increase in the number of crosses required for the mapping of an arithmetically increasing number of mutants available for study. (Stent 1971, 368)

[34] The paper in which this map was published appeared in the March 1961 issue of Proceedings of the National Academy of Sciences of USA, but the paper was presented at the Academy on April 27, 1960.

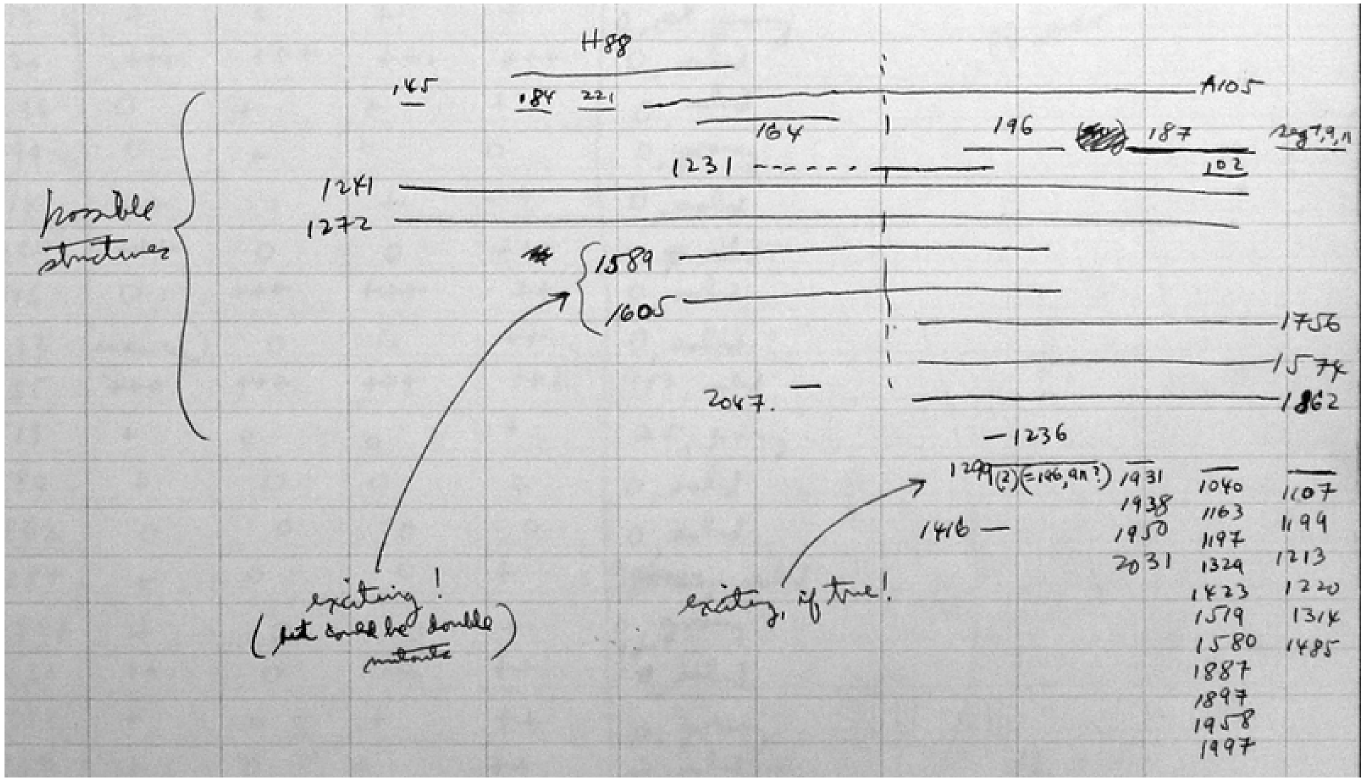


Figure 4.9: A working map of new anomalous mutants drawn on September 21, 1959. Reproduced from Seymour Benzer Papers, Box 70 Folder 2, 10242-MS, Caltech Archives, California Institute of Technology. Courtesy of the Archives, California Institute of Technology.

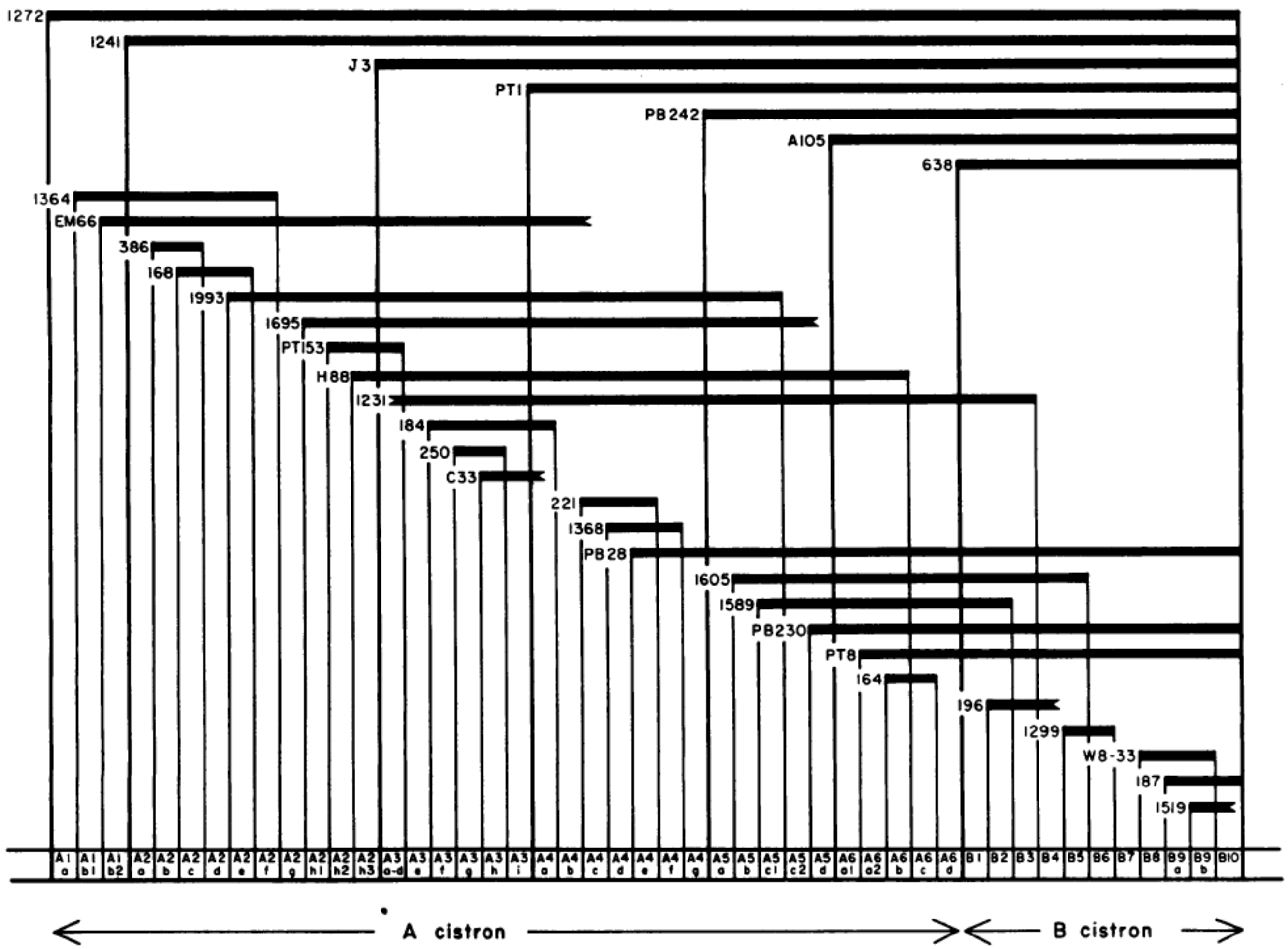


Figure 4.10: A map of anomalous mutants as of April 1960. From Benzer (1961, 406)

4.4.3 Uses of a Genetic Map and the Instrumentalist Methodological Principle

Benzer used a genetic map—his model of DNA—to organize the results of recombination experiments and to design a new method for rapid mapping. I argue that these two crucial uses of a genetic map followed the instrumentalist methodological principle.

To organize the results of recombination experiments, Benzer interpreted some parts and properties of the model as representing things other than the parts and properties of DNA.[35] He interpreted the spatial relations exhibited by points and bars on a map—that one point was *under* a bar or that one bar *overlapped* another bar—as representing the mutant phages' behaviors during experiments (i.e., whether and how much they produce wild-type recombinants), rather than as representing, say, spatial relations of parts of DNA. Horizontal bars, for example, were interpreted as representing mutated parts of DNA, but the length of each bar and overlaps among bars were not interpreted as representing properties of DNA. Benzer's interpretation of spatial properties of a map made the rules of mapmaking efficient (Figure 4.4 and 4.7), because with this interpretation he was able to construct a map simply by reading off numbers from the table of experimental results. Thus, Benzer's use of a genetic map to organize the experimental results followed the instrumentalist methodological principle.

To design a method for rapid mapping, Benzer interpreted a map segment as representing a part of DNA. But again he interpreted the overlaps between bars on a map as representing the phages' behaviors during experiments. This interpretation was crucial for designing a two-step method to classify rII mutants into a segment on a map. Thus, Benzer's use of a genetic map to design a mapping method followed the instrumentalist methodological principle.

[35] While interpreting some parts and properties of his model as representing things other than the parts and properties of DNA, Benzer also interpreted other parts and properties of his map as representing parts and properties of DNA. In particular, he interpreted points and horizontal bars on a map as representing parts of DNA and their property of being mutated from the wild type, and the linearity of a map as representing the linearity of the structure of DNA. And it was because of this interpretation (or preference to give such an interpretation) that Benzer kept his map linear although he knew that linearity was not strictly required by the results of recombination experiments (Benzer 1959). Thus, his genetic map was a model of DNA while it was being used according to the instrumentalist methodological principle.

4.5 THE NATURE OF DNA, USES OF A GENETIC MAP, AND THE REALIST METHODOLOGICAL PRINCIPLE

We have seen how Benzer used his model according to the instrumentalist methodological principle. In this section I argue that Benzer's uses of his model also followed the realist methodological principle. To do this I begin by describing the overarching aim of Benzer's research, which naturally made Benzer use his model according to the realist methodological principle. At the end of this section I show how a subtle tension emerged between his uses of a model because of his joint adoption of the realist and instrumentalist methodological principles.

4.5.1 The Junkman's Problem

The overarching aim of Benzer's research was to gain deeper understanding of DNA. This aim is evident in Benzer's general characterization of his research. On December 4, 1956, in his biophysics seminar, Benzer gave a lecture entitled "Mutations and the Junkman's Problem." He began by describing the processes of mutation and recombination and their relevance to his research on genetic fine structure. He then described his research as concerning "the junkman's problem," whose "objective" was "to determine something about the structure [of DNA] by these operations [mutations and recombinations]" (SBP 82.1). The junk was mutations and recombinations, and the value he wanted to get out of the junk was knowledge about the structure of DNA. His tool was the genetic map, the model of DNA. Compared to other surviving statements of his research aims, what Benzer called the junkman's problem was a fair characterization of the overall aim of his research.[36]

Benzer obtained many important solutions to the junkman's problem (e.g., Benzer 1955, 1956, 1957, 1959, 1961, 1962). I discuss two of them that reveal how Benzer used his model according to the realist methodological principle.

[36] For more detailed statements of his research aims, see Benzer's "Application for Extension of Grant from American Cancer Society" (September 28, 1954, SBP 1.14); Benzer's grant proposal to the National Science Foundation entitled "Genetic Fine Structure and Its Relation to the Molecular Structure of DNA" (February 12, 1955, SBP 9.7); his "Application for Extension of Grant from American Cancer Society" (September 21, 1955, SBP 1.15); and Benzer's NSF progress reports dated April 29, 1957 (SBP 9.7) and March 31, 1958 (SBP 9.8).

4.5.2 Explaining the Nature of Anomalous Mutants

Benzer used his genetic map of anomalous mutants (e.g., Figure 4.5 and 4.6) to develop explanatory hypotheses about the molecular nature of mutation responsible for anomalous mutants. As we saw, in February 1955 when Benzer found anomalous mutants, he changed the rules of mapmaking so that he could represent anomalous mutants together with other mutants on the same map (Figure 4.7). In addition, Benzer asked why anomalous mutants could not be drawn as points on a map and considered the hypothesis that an anomalous mutant had a stretch of DNA deleted or altered. But he did not have any evidence for a specific hypothesis about the nature of mutation involved in anomalous mutants.[37] Only in 1961, Benzer and his then postdoc Masayasu Nomura developed an elegant test of the deletion hypothesis: if a stretch of DNA is physically deleted, then the map distance between mutants outside of that stretch should become shorter. If the stretch is altered but not deleted, the map distance between outside mutants should remain unchanged. Nomura and Benzer's experiments showed that the map distance of two mutants was indeed shorter when anomalous mutation occurred between them (Nomura and Benzer 1961).

In conformity with the realist methodological principle, Benzer used his model to develop a specific hypothesis about the nature of the object of his investigation, and to do so he interpreted a horizontal bar on a map as representing a stretch of DNA with the length proportional to that of the bar and its property of being deleted during the process of mutation. Benzer also interpreted the distance between any two points on a map as representing the physical distance between parts of DNA. As we saw, in making a map of anomalous mutants, Benzer interpreted a horizontal bar as representing only a mutated part of DNA, and this interpretation was sufficient for doing the things he was doing by following the instrumentalist methodological principle. But in formulating the deletion hypothesis, Benzer gave the same part of his model a new interpretation that this part represents a part of DNA and its property. This is evidence that he was following the realist methodological principle.

[37] Benzer often called anomalous mutants "deletions," writing the word in quotation marks to indicate the fact that the term was merely a label (e.g., Benzer 1957, 76).

4.5.3 The Genetic Map and the Watson-Crick Model

Benzer also used his genetic map to make specific inferences about the structure of DNA. By April 1957, he had a map that represented a very large number of rII mutants. On April 29, 1957, Benzer wrote a progress report on his National Science Foundation grant that had begun in September of 1955. In it he said:

> The fact that all mutations of the deletion type [i.e., anomalous mutants] can be ordered in a linear framework has provided a rigorous demonstration that the genetic structure is of linear topology down to its finest details. This is exactly the expectation from the Watson-Crick structure of DNA. Furthermore, the finding that the genetic structure is divisible by mutation and recombination down to a level corresponding to individual nucleotides is also consistent with the Watson-Crick structure. (Seymour Benzer to W. V. Consolazio, April 29, 1957, SBP 9.7)

By "a linear framework" he referred to the fact that in a map like Figure 4.10 all segments were ordered next to each other on a line (Benzer 1959, 1607; Benzer 1961, 415). The above passage contains two arguments for the structure of DNA as described by the Watson-Crick model. The first argument can be reconstructed as follows:

1. If DNA has the linear structure as described by the Watson-Crick model, then it must be possible to arrange all mutants in a linear topological order.

2. All anomalous mutants studied thus far can be arranged in a linear map.

3. Therefore, the map of anomalous mutants strongly supports the idea that the structure of DNA is linear at the molecular level.

Benzer's second argument can be reconstructed as follows:

1. If DNA is made up of a chain of bases as described by the Watson-Crick model, then in principle mutation and recombination can occur at any point in DNA.

2. rII mutants are distributed throughout the genetic map, and the map has many segments detectable by recombination experiments.

3. Therefore, the structure of the genetic map is consistent with the structure of DNA as described by the Watson-Crick model.

As can be seen in the second premise of each argument, Benzer took the properties of his model—linearity and divisibility—as representing the properties of DNA predicted by the Watson-Crick model. This interpretation allowed him to argue that the features of his model of DNA support the Watson-Crick model.[38] As noted above, Benzer consistently interpreted the linearity of a map as representing a property of DNA, and this interpretation affected how he used his model to organize the experimental results: it affected his use of a model according to the instrumentalist methodological principle. Now, in using his model according to the instrumentalist methodological principle, Benzer did not give any interpretation to the divisibility of a map, presumably because no interpretation was necessary to do the particular tasks he was doing. But to make claims about the structure of DNA, Benzer interpreted divisibility of a map as also representing a property of DNA. His attempt to interpret a previously disregarded property of his model as representing a property of the object of his investigation is evidence that he was following the realist methodological principle.

* * *

Having seen how Benzer followed both the instrumentalist and realist methodological principles, we are now in a good position to recognize the characteristic tension among Benzer's uses of his model that resulted from his adoption of these two methodological principles. I describe two places where we can recognize this tension.

One place where we can see the characteristic tension is in his decision to change the rules of mapmaking in response to the discovery of anomalous mutants. Benzer changed the rules of mapmaking to put anomalous mutants on a map (cf. Figure 4.4 and 4.7) although the initial map of anomalous mutants (on the left of Figure 4.5) was adequate for summarizing the results of experiments. What, then, was Benzer's motivation to change his model? Besides using his model to organize the experimental results, Benzer used it to make claims about DNA. The initial genetic map of anomalous mutants had properties, especially the presence of nonzero distances within zero distances, that seemed hard to interpret as representing physical properties of DNA. But since he followed the realist methodological principle, he would have preferred a model that exhibited properties that he *could* interpret as representing properties of DNA. And he found such a model

[38] In later years when he obtained a more saturated map, Benzer would give these arguments with stronger support (Benzer 1960, 17–20; Benzer 1962, 83–84).

by changing the rules of mapmaking. Here we see some tension between using a model according to the instrumentalist methodological principle and using it according to the realist methodological principle. Benzer initially used a model to organize the experimental results even if the model showed nonzero distances within zero distances: he was following the instrumentalist methodological principle. He then modified his model in order to eliminate a property that he could not possibly interpret as representing a property of the object of his investigation: he prepared a model that he could potentially use according to the realist methodological principle.

Another place where we can see the characteristic tension is in Benzer's sustained attempt to construct a saturated map of anomalous mutants (e.g., Figure 4.9 and 4.10). The first map of anomalous mutant was a product of the tension I just described above, and Benzer's adoption of the realist methodological principle made it significant that a genetic map produced by the revised rules of mapmaking had properties that could be interpreted as representing properties of DNA. From 1955 Benzer continued making this model more detailed although it was only in 1961 that he obtained evidence suggesting that he could successfully use his model to make claims about the nature of mutations underlying anomalous mutants. During these years, Benzer could not have sufficiently justified his *continued use* of his model by appealing to the fact that his model *could* be used according to the realist methodological principle. Indeed, the realist methodological principle could have motivated him to look for a radically different model. But he went on with his map of anomalous mutants. What we see here is the characteristic tension: Benzer was not just following the realist methodological principle, and his adoption of the instrumentalist methodological principle allowed him to continue using and developing a map of anomalous mutants, for such a map was crucial for rapidly mapping new rII mutants.

4.6 STEIN'S CONJECTURE AND THE REALISM DEBATE

My case study of Benzer's uses of a model of DNA supports Stein's conjecture. Using a model according to both the realist and instrumentalist methodological principles can be crucial for the success of scientific research. Thus, other things being equal, it is reasonable for a scientist to adopt both methodological principles even if doing so may create tensions among her uses of a model. In my view, philosophers of science trying to understand how successful science works will benefit from examining the realist and instrumentalist methodological principles with the help of detailed historical studies of scientific practice.

Stein's conjecture also matters to the realism debate. Although it is reasonable for a scientist to adopt both the realist and instrumentalist methodological principles, I shall argue that some of the currently most important positions in the realism debate prohibit a scientist from adopting both methodological principles. This thesis, if true, matters to the realism debate because this debate implicitly assumes that a correct position in the debate makes existing successful scientific methods reasonable for a working scientist to follow. For brevity, let us restate the above assumption as follows: a correct position in the realism debate *makes sense of* successful scientific methods. Thus, if a scientist does successful research by following the method that is naturally suggested by a given metaphysical, semantic, epistemic, or axiological position, then this fit constitutes prima facie support for that position. And a mismatch counts against that position.[39]

My thesis is only that *some* positions in the realism debate cannot make sense of the adoption of both the realist and instrumentalist methodological principles. My argument will focus on epistemic realism and instrumentalism, which are some of the most important positions in the realism debate today. I argue as follows:

P1 Epistemic realism and epistemic instrumentalism contradict each other.

P2 Epistemic realism can make sense of the realist methodological principle, but not the instrumentalist methodological principle.

P3 Epistemic instrumentalism can make sense of the instrumentalist methodological principle, but not the realist methodological principle.

[39] To see the above assumption at play, consider, for example, one of Stathis Psillos' arguments against semantic instrumentalism (as formulated in Section 4.1. For Psillos' own formulation, see Psillos (1999, xix)). Following Duhem, he argues that scientific theories are not only means for organizing our experience but also means for understanding the world, and this latter use of theories does not make sense if theories are not truth-conditioned descriptions of the world. Semantic instrumentalism thus contradicts what scientists do with theories (Psillos 1999, 34). Here the feature of successful methodology, in particular, what successful scientists do with theories, is used to criticize semantic instrumentalism. Consider also Bas van Fraassen's defense of constructive empiricism, which can be seen partly as a version of epistemic instrumentalism (see Fine 1986, 157; Fine 2001). Van Fraassen defends constructive empiricism by showing how it can help us make sense of the methodology of scientific experiments (van Fraassen 1980, 70, 73–77). It is not difficult to find other examples: Richard Boyd argues that instrumental reliability of theory-dependent scientific methods can only be explained by scientific realism consisting of the metaphysical, semantic, and epistemic realist positions, which makes it the case that background theories on which scientific methods are dependent are approximately true of the world (Boyd 1983, 64–65).

P4 The conjunction of epistemic realism and epistemic instrumentalism can make sense of both methodological principles, but the conjunction is a contradiction.

C1 Therefore, epistemic realism or epistemic instrumentalism by itself prohibits a scientist from adopting both the realist and instrumentalist methodological principles.

C2 Therefore, some of the currently most important positions in the realism debate cannot show that it is reasonable for a scientist to adopt both the realist and instrumentalist methodological principles.

Below I argue primarily for P2 and P3, assuming that P1 is true and that P4 is obvious given P1, P2, and P3. C1 says that epistemic realism or epistemic instrumentalism alone "prohibits" a scientist from adopting both methodological principles in the sense that holding one of these epistemic positions allows a scientist to adopt one of the methodological principles but not the other principle on pain of contradiction.

To start, then, with P2: For the epistemic realist, the realist methodological principle is a reasonable principle to follow. Her epistemic realism says that she should believe that her model is an approximately accurate representation of the object of her investigation. It is thus reasonable for her to use her model to make claims and generate questions about the object of her investigation and to interpret, whenever possible, parts and properties of her model as representing the parts and properties of the object of her investigation. But for the epistemic realist, the instrumentalist methodological principle is not a reasonable principle to follow. To use her model according to the instrumentalist methodological principle, she may have to *reinterpret* parts and properties of her model that she has already interpreted as representing the parts and properties of the object of her investigation: she may have to interpret some of these parts and properties of her model as representing *not* the parts and properties of the object of her investigation *but* some other things whenever this reinterpretation would help her perform practical tasks. If she reinterpreted parts and properties of her model in this way, she would be violating the realist methodological principle. Thus, epistemic realism can make sense of the realist methodological principle, but not the instrumentalist methodological principle.

Turning to P3, for the epistemic instrumentalist, the instrumentalist methodological principle is a reasonable principle to follow. The epistemic instrumentalist believes that she should make best practical use out of her model without believing what it says about the object of her investigation. Her epistemic instrumentalism thus justifies her use of a model to do tasks that address her practical needs and to interpret parts and properties of her model as representing things other than part and properties of her investigative target whenever such interpretation helps her perform the practical tasks. But for the epistemic instrumentalist, the realist methodological principle is not a reasonable principle to follow because she does not believe what her model says about the object of her investigation. Thus, it is not reasonable for her to use her model to make claims about the true nature of the object of her investigation.[40] Epistemic instrumentalism can make sense of the instrumentalist methodological principle, but not the realist methodological principle.

Now, as P4 says, the conjunction of epistemic realism and epistemic instrumentalism may seem to be able to make sense of both methodological principles. But this amounts to accepting contradictory epistemic positions: it is not a rational option for an individual scientist. Therefore, accepting one of these epistemic positions prohibits a scientist from adopting both the realist and instrumentalist methodological principles. Some of the currently most important positions in the realism debate cannot show that it is reasonable for a scientist to adopt both the realist and instrumentalist methodological principles.

Stein's conjecture thus poses new challenges to epistemic realism and epistemic instrumentalism. The historical evidence presented above suggests that other things being equal it is reasonable for a scientist to use a model according to both the realist and instrumentalist methodological principles. But epistemic realism or epistemic instrumentalism by itself seems to prohibit a scientist from following both these methodological principles.

The epistemic realist might point out that epistemic realism does allow a scientist to use a model according to the instrumentalist methodological principle if doing so is a prerequisite for using the model according to the realist methodological principle. But *that* instrumentalist methodological principle seems to be:

> THE INSTRUMENTALIST METHODOLOGICAL PRINCIPLE*: A scientist should use a model to organize or predict observational and experimental results, to design new techniques for intervention, and to do other tasks that address practical needs *only if the scientist does not have to reinterpret parts and properties of the model—that have been*

[40]But she would interpret parts and properties of her model as representing parts and properties of the object of her investigation if such interpretation helped her to perform practical tasks by providing otherwise unavailable constraints on certain parts or properties of her model.

> *interpreted as representing parts and properties of the object of the investigation—as representing things other than the parts and properties of the object of the investigation.*

Epistemic realism can make sense of this version of the instrumentalist methodological principle. But the italicized part of this principle restricts the range of permissible interpretations in ways that the full-fledged instrumentalist methodological principle that I formulated does not.

The epistemic instrumentalist might point out that there is an important version of epistemic instrumentalism to consider: the form of epistemic instrumentalism developed by Stanford (2006, 197–211; see footnote 8). The epistemic instrumentalist of this stripe would believe at least some of what her model says about the object of her investigation.[41] Her epistemic instrumentalism would then allow her to use the model to make claims about the object of her investigation insofar as doing so would help her do practical tasks. But here the relevant realist methodological principle seems to be:

> THE REALIST METHODOLOGICAL PRINCIPLE*: A scientist should use a model to make claims—in the form of hypotheses, descriptions, or explanations—and generate questions about the true nature of the object of an investigation *only if doing so helps the scientist perform tasks that address practical needs*. To do this, the scientist should interpret parts and properties of the model as representing the parts and properties of the object of the investigation *only if such interpretation help the scientist perform practical tasks*.

Stanford's version of epistemic instrumentalism seems to be able to make sense of this version of the realist methodological principle. But the first italicized part of this principle makes a scientist's adherence to this principle conditional on her successful use of her model according to the instrumentalist methodological principle. Thus, this principle is realist in the sense that it tells her to do the same things that the full-fledged realist methodological principle does, but it is deeply instrumentalist since it ultimately concerns practical needs. In addition, the second italicized part replaces "whenever possible" in the original version of the principle and makes the above version of principle more restrained than the original.

[41] Insofar as she can understand what her model says in terms of another theory or model about which she is an epistemic realist—this constraint is a distinguishing feature of Stanford's version of epistemic instrumentalism (see footnote 8). What I say here on behalf of Stanford's version of epistemic instrumentalism assumes that this constraint is always satisfied.

I end this section by sketching other possible responses to the challenges posed by Stein's conjecture. First, the epistemic realist may respond to the challenges by showing that the instrumentalist methodological principle* is equivalent to the full-fledged instrumentalist methodological principle. To do so, she may use a historical case, such as the Benzer case I presented, to show how following both the realist methodological principle and the instrumentalist methodological principle* can result in the same sort of tension among a scientist's uses of a model that I described above. Second, the epistemic instrumentalist may try to adopt Stanford's version of epistemic instrumentalism and show that the realist methodological principle* is equivalent to the realist methodological principle. She can then use a historical case to show how following both the realist methodological principle* and the instrumentalist methodological principle can result in the same characteristic tension among a scientist's uses of a model. Third, while accepting my thesis that epistemic realism or instrumentalism by itself cannot make sense of a scientist's adoption of both the realist and instrumentalist methodological principles, the epistemic realist and instrumentalist may try to resolve the challenges by looking at a group of scientists instead of an individual scientist. Suppose that some members of the group may adopt the realist methodological principle, while others adopt the instrumentalist methodological principle. If no single member adopts both principles, then Stein's conjecture does not pose challenges to epistemic realism or instrumentalism.[42]

4.7 CONCLUSION

Stein conjectured that in successful scientific research, a scientist uses a model according to the methodological principles of realism and instrumentalism despite the tension they create among the uses of the model. In this chapter I formulated the relevant methodological principles of realism and instrumentalism and argued that a detailed analysis of Benzer's uses of his model of DNA—a genetic map—indeed supports Stein's conjecture. I then argued that epistemic realism or epistemic

[42]It is also possible, though may not be attractive, to simply acknowledge that epistemic realism or epistemic instrumentalism is not in the business of making sense of successful scientific practice. This amounts to the rejection of the implicit assumption of the realism debate that a correct position in the debate makes sense of successful scientific methods.

instrumentalism by itself prohibits a scientist from adopting both the realist and instrumentalist methodological principles. Stein's conjecture thus presents new challenges to some of the currently most important positions in the realism debate, and I suggested some avenues that realists and instrumentalists may pursue in response.

5.0 CONCLUSION

What is a scientific model? I began Chapter 1 with this question, because in order to understand the practice of modeling, we need to say what a model is. I surveyed four answers from the literature: models are (i) mathematical entities, (ii) abstract entities, (iii) abstract or concrete entities, and (iv) imagined concrete entities. I argued that except for (i), all these answers take *the metaphysics based perspective* on models in the sense that they invoke the concepts of abstract and concrete entities developed in metaphysics. The metaphysics based perspective sees models as first and foremost *objects* that scientists use to represent parts of the world. Thus, to say what a model is, it applies the concepts of objects—abstract and concrete entities—to models.

I sketched an alternative perspective that applies the concepts of internal and external representations from cognitive science. This perspective sees models as first and foremost *representations* that scientists use for cognitive and epistemic purposes in research. Thus, to say what a model is, I applied the concepts of representations—internal and external representations—to models. In particular, my own perspective is the cognitive science based perspective with a focus on external representations: I see models as external representations that scientists interact with via perceptual and bodily processes in order to perform cognitive and epistemic tasks.

In Chapter 2, from the cognitive science based perspective, I developed a general account of the practice of modeling. I began the chapter by examining influential accounts of modeling articulated by Giere and Weisberg. They take the metaphysics based perspective on models: for Giere and Weisberg, some models, such as scale models, are concrete, physical entities, while other models, such as mathematical models, are abstract, non-physical entities. Giere and Weisberg then suggest that scientists use external representations like interpreted equations to construct, manipulate, and analyze abstract models. But I argued that the category of abstract models obscures rather than illuminates the practice of modeling. I then developed an alternative account of mathematical modeling in which modeling is understood as a practice of constructing, manipulating, and analyzing external representations in service of cognitive and epistemic aims of research. I concluded that to better understand the practice of modeling, including mathematical modeling, we should see models as external representations.

In Chapter 3, I turned to the dynamics of research involving models. Focusing on the relationship among uses of a model, particular aims of research in which scientists use the model, and criteria of success relevant to a given use of the model, I argued (i) that the relationship between uses of a model and particular aims of research is dynamic in the sense that uses of the same model can shift over the course of scientists' research in response to the shift in aim, and (ii) that criteria of success for one use of a model can be different from those for another use of the same model. I presented a detailed case study of Wright's research to support these claims.

In Chapter 4, developing Stein's idea further, I argued that in successful scientific research, a scientist uses a model according to the methodological principles of realism and instrumentalism despite the tension that they create among the scientist's uses of the model over time. I defended this thesis through a detailed analysis of successful scientific research done by Benzer in the 1950s and 60s. I showed that various uses of the very same model in this case followed both the realist and instrumentalist methodological principles. These uses exhibited a characteristic tension between the realist and instrumentalist uses of the model. I then argued that epistemic realism or epistemic instrumentalism by itself prohibits a scientist from adopting both the realist and instrumentalist methodological principles.

BIBLIOGRAPHY

Achinstein, Peter. 1968. *Concepts of Science: A Philosophical Analysis*. Baltimore, MD: Johns Hopkins Press.

Andronov, A. A., E. A. Leontovich, I. I. Gordon, and A. G. Maier. 1973. *Qualitative Theory of Second-Order Dynamic Systems*. New York: Wiley.

Bailer-Jones, Daniela M. 1999. Tracing the development of models in the philosophy of science. In L. Magnani, N. J. Nersessian, and P. Thagard (eds.), *Model-Based Reasoning in Scientific Discovery*, 23–40. New York: Kluwer.

Bailer-Jones, Daniela M. 2009. *Scientific Models in Philosophy of Science*. Pittsburgh, PA: University of Pittsburgh Press.

Barberousse, Anouk, and Pascal Ludwig. 2009. Models as fictions. In M. Suárez (ed.), *Fictions in Science: Philosophical Essays on Modeling and Idealization*, 56–73. New York: Routledge.

Benzer, Seymour. 1955. Fine structure of a genetic region in bacteriophage. *Proceedings of the National Academy of Sciences of the United States of America* 41:344–354.

———. 1956. Genetic fine structure and its relation to the DNA molecule. In *Mutation: Report of Symposium held June 15 to 17, 1955*, 3–5. Upton, NY: Brookhaven National Laboratory.

———. 1957. The elementary units of heredity. In W. D. McElroy and B. Glass (eds.), *The Chemical Basis of Heredity*, 70–93. Baltimore, MD: Johns Hopkins Press.

———. 1959. On the topology of the genetic fine structure. *Proceedings of the National Academy of Sciences of the United States of America* 45:1607–1620.

———. 1960. Genetic fine structure. *Harvey Lectures* 56:1–21.

———. 1961. On the topography of the genetic fine structure. *Proceedings of the National Academy of Sciences of the United States of America* 47:403–415.

———. 1962. The fine structure of the gene. *Scientific American* 206:70–84.

———. 1966. Adventures in the rII region. In J. Cairns, G. S. Stent, and J. D. Watson (eds.), *Phage and the Origins of Molecular Biology*, 157–165. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Benzer, Seymour, Papers. 10242-MS. Caltech Archives. California Institute of Technology.

Black, Max. 1962. *Models and Metaphors: Studies in Language and Philosophy*. Ithaca, NY: Cornell University Press.

Boon, Mieke, and Tarja Knuuttila. 2009. Models as epistemic tools in engineering sciences. In A. Meijers (ed.), *Philosophy of Technology and Engineering Sciences*, 693–726. Amsterdam: Elsevier.

Boumans, Marcel. 1999. Built-in justification. In M. S. Morgan and M. Morrison (eds.), *Models as Mediators: Perspectives on Natural and Social Science*, 66–96. Cambridge: Cambridge University Press.

Boyd, Richard N. 1983. On the current status of the issue of scientific realism. *Erkenntnis* 19:45–90.

Boyd, Richard N. 2010. Scientific realism. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy (Summer 2010 Edition)*. URL = <http://plato.stanford.edu/archives/sum2010/entries/scientific-realism/>.

Cartwright, Nancy. 1989. *Nature's Capacities and their Measurement*. Oxford: Clarendon Press.

Cartwright, Nancy. 1999. Models and the limits of thoery: Quantum Hamiltonians and the BCS model of superconductivity. In M. S. Morgan and M. Morrison (eds.), *Models as Mediators: Perspectives on Natural and Social Science*, 241–281. Cambridge: Cambridge University Press.

Cat, Jordi. 2001. On understanding: Maxwell on the methods of illustration and scientific metaphor. *Studies in History and Philosophy of Modern Physics* 32:395–441.

Chakravartty, Anjan. 2013. Scientific realism. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy (Summer 2013 Edition)*. URL = <http://plato.stanford.edu/archives/sum2013/entries/scientific-realism/>.

Chang, Hasok. 2012. *Is Water H_2O?: Evidence, Realism and Pluralism*. Dordrecht: Springer.

Clark, Andy. 2008. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford: Oxford University Press.

Clark, Andy, and David Chalmers. 1998. The extended mind. *Analysis* 58:7–19.

da Costa, Newton C. A., and Steven French. 1990. The model-theoretic approach in the philosophy of science. *Philosophy of Science* 57:248–265.

de Chadarevian, Soraya, and Nick Hopwood. (eds.) 2004. *Models: The Third Dimension of Science*. Stanford: Stanford University Press.

De Cruz, Helen, Hansjörg Neth, and Dirk Schlimm. 2010. The cognitive basis of arithmetic. In B. Löwe and T. Müller (eds.), *PhiMSAMP. Philosophy of Mathematics: Sociological Aspects and Mathematical Practice*, 59–106. London: College Publications.

Dobzhansky, Theodosius, and Sewall Wright. 1941. Genetics of natural populations, V: Relations between mutation rate and accumulation of lethals in populations of *Drosophila pseudoobscra*. *Genetics* 26:23–51.

Doermann, A. H., and M. B. Hill. 1953. Genetic structure of bacteriophage T4 as described by recombination studies of factors influencing plaque morphology. *Genetics* 38:79–90.

Downes, Stephen M. 1992. The importance of models in theorizing: A deflationary semantic view. In D. Hull, M. Forbes, and K. Okruhlik. (eds.), *PSA 1992*, Volume 1, 142–153. East Lansing, MI: Philosophy of Science Association.

Dutilh Novaes, Catarina. 2012. *Formal Languages in Logic: A Philosophical and Cognitive Analysis*. Cambridge: Cambridge University Press.

Dutilh Novaes, Catarina. 2013. Mathematical reasoning and external symbolic systems. *Logique et Analyse* 221:45–65.

Edelstein-Keshet, Leah. 2005. *Mathematical Models in Biology*. Philadelphia: SIAM.

Emerson, Sterling. 1938. The genetics of self-incompatibility in *Oenothera organensis*. *Genetics* 23:190–202.

Emerson, Sterling. 1939. A preliminary survey of the oenothera organensis population. *Genetics* 24:524–537.

Epling, Carl, and Theodosius Dobzhansky. 1942. Genetics of natural populations, VI: Microgeographic races in *Linanthus parryae*. *Genetics* 27:317–332.

Epling, Carl, Harlan Lewis, and Francis M. Ball. 1960. The breeding group and seed storage: A study in population dynamics. *Evolution* 14:238–255.

Fine, Arthur. 1986. Unnatural attitudes: Realist and instrumentalist attachments to science. *Mind* 95:149–179.

Fine, Arthur. 2001. The scientific image twenty years later. *Philosophical Studies* 106:107–122.

Franklin, Rosalind E., and R. G. Gosling. 1953a. Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate. *Nature* 172:156–157.

Franklin, Rosalind E., and R. G. Gosling. 1953b. Molecular configuration in sodium thymonucleate. *Nature* 171:740–741.

French, Steven, and James Ladyman. 1999. Reinflating the semantic approach. *International Studies in the Philosophy of Science* 13:103–121.

Frigg, Roman. 2010. Models and fiction. *Synthese* 172:251–268.

Giere, Ronald N. 1988. *Explaining Science: A Cognitive Approach*. Chicago: University of Chicago Press.

———. 1999. Using models to represent reality. In L. Magnani, N. J. Nersessian, and P. Thagard (eds.), *Model-Based Reasoning in Scientific Discovery*, 41–58. New York: Kluwer.

———. 2002a. Discussion note: Distributed cognition in epistemic cultures. *Philosophy of Science* 69:637–644.

———. 2002b. Scientific cognition as distributed cognition. In P. Carruthers, S. P. Stich, and M. Siegal (eds.), *Cognitive Basis of Science*, 285–299. Cambridge: Cambridge University Press.

———. 2004. How models are used to represent reality. *Philosophy of Science* 71:742–752.

———. 2006. *Scientific Perspectivism*. Chicago: University of Chicago Press.

———. 2009. Why scientific models should not be regarded as works of fiction. In M. Suárez (ed.), *Fictions in Science: Philosophical Essays on Modeling and Idealization*, 248–270. New York: Routledge.

———. 2010. An agent-based conception of models and scientific representation. *Synthese* 172:269–281.

———, and Barton Moffatt. 2003. Distributed cognition: Where the cognitive and the social merge. *Social Studies of Science* 33:301–310.

Godfrey-Smith, Peter. 2003. *Theory and Reality: An Introduction to the Philosophy of Science*. Chicago: University of Chicago Press.

———. 2006a. The strategy of model-based science. *Biology and Philosophy* 21:725–740.

———. 2006b. Theories and models in metaphysics. *Harvard Review of Philosophy* 14:4–19.

———. 2009. Models and fictions in science. *Philosophical Studies* 143:101–116.

Goodman, Nelson. 1976. *Languages of Art* (2d. ed.). Indianapolis, IN: Hackett Publishing Company.

Gould, Stephen Jay. 1980. G. G. Simpson, paleontology, and the modern synthesis. In E. Mayr and W. B. Provine (eds.), *The Evolutionary Synthesis: Perspectives on the Unification of Biology*, 153–172. Cambridge, MA: Harvard University Press.

———. 1982. Introduction. In Theodosius Dobzhansky, *Genetics and the Origin of Species*, xvii–xli. New York: Columbia University Press.

———. 1983. The hardening of the modern synthesis. In M. Grene (ed.), *Dimensions of Darwinism: Themes and Counterthemes in Twentieth-Century Evolutionary Theory*, 71–93. Cambridge: Cambridge University Press.

Greenspan, Ralph J. 2009. Seymour Benzer: 1921–2007. *Biographical Memoirs of National Academy of Sciences*:3–18.

Griesemer, James R. 1990. Modeling in the museum: On the role of remnant models in the work of Joseph Grinnell. *Biology and Philosophy* 5:3–36.

———. 1991. Material models in biology. In A. Fine, M. Forbes, and L. Wessels. (eds.), *PSA 1990*, Volume 2, 79–93. East Lansing, MI: Philosophy of Science Association.

———. 2004. Three-dimensional models in philosophical perspective. In S. de Chadarevian and N. Hopwood (eds.), *Models: The Third Dimension of Science*, 433–442. Stanford, CA: Stanford University Press.

Hacking, Ian. 1983. *Representing and Intervening: Introductory Topics in the Philosohy of Natural Science*. Cambridge: Cambridge University Press.

Hershey, Alfred D. 1946. Mutation of bacteriophage with respect to type of plaque. *Genetics* 31:620–640.

Hesse, Mary B. 1966. *Models and Analogies in Science*. Notre Dame, IN: University of Notre Dame Press.

Hollan, James, Edwin Hutchins, and David Kirsh. 2000. Distributed cognition: Toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction* 7:174–196.

Holmes, Frederic Lawrence. 2006. *Reconceiving the Gene: Seymour Benzer's Adventures in Phage Genetics*. New Haven: Yale University Press.

Hutchins, Edwin. 1995. *Cognition in the Wild*. Cambridge, MA: MIT Press.

Huxley, Julian. 1938. Clines: An auxiliary taxonomic principle. *Nature* 142:219–220.

Johnson-Laird, P. N. 1980. Mental models in cognitive science. *Cognitive Science* 4:71–115.

———. 1983. *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Cambridge, MA: Harvard University Press.

Judson, Horace Freeland. 1996. *The Eighth Day of Creation: Makers of the Revolution in Biology* (Expanded ed.). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Kirsh, David. 2006. Distributed cognition: A methodological note. *Pragmatics and Cognition* 14:249–262.

Kitcher, Philip. 1993. *The Advancement of Science: Science without Legend, Objectivity without Illusions*. New York: Oxford University Press.

Knuuttila, Tarja. 2005. Models, representation, and mediation. *Philosophy of Science* 72:1260–1271.

———. 2011. Modelling and representing: An artefactual approach to model-based representation. *Studies in History and Philosophy of Science* 42:262–271.

———, and Mieke Boon. 2011. How do models give us knowledge?: The case of Carnot's ideal heat engine. *European Journal for Philosophy of Science* 1:309–334.

———, and Atro Voutilainen. 2003. A parser as an epistemic artifact: A material view on models. *Philosophy of Science* 70:1484–1495.

Lakoff, George, and Mark Johnson. 1980. *Metaphors We Live By*. Chicago: University of Chicago Press.

———. 1999. *Philosophy in the Flesh: The Embodied Mind and Its Challange to Western Thought*. New York: Bacis Books.

Larkin, Jill H., and Herbert A. Simon. 1987. Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science* 11:65–99.

Laudan, Larry. 1977. *Progress and Its Problems: Towards a Theory of Scientific Growth*. Berkeley: University of California Press.

Leplin, Jarrett. 1984. Introduction. In J. Leplin (ed.), *Scientific Realism*, 1–7. Berkeley: University of California Press.

Levins, Richard. 1966. The strategy of model building in population biology. *American Scientist* 54:421–431.

———. 1993. A response to Orzack and Sober: Formal analysis and the fluidity of science. *The Quarterly Review of Biology* 68:547–555.

Lewis, David. 1986. *On the Plurality of Worlds*. Oxford: Basil Blackwell.

Lloyd, Elisabeth A. 1984. A semantic approach to the structure of population genetics. *Philosophy of Science* 51:242–264.

Lyons, Timothy D. 2005. Toward a purely axiological scientific realism. *Erkenntnis* 63:167–204.

Magnani, Lorenzo, and Nancy J. Nersessian. (eds.) 2002. *Model-Based Reasoning: Science, Technology, Values*. New York: Kluwer.

Magnani, Lorenzo, Nancy J. Nersessian, and Paul Thagard. (eds.) 1999. *Model-Based Reasoning in Scientific Discovery*. New York: Kluwer.

May, Robert M. 1973. *Stability and Complexity in Model Ecosystems*. Princeton, NJ: Princeton University Press.

Morgan, Mary S., and Margaret Morrison. (eds.) 1999. *Models as Mediators: Perspectives on Natural and Social Science*. Cambridge: Cambridge University Press.

Morrison, Margaret. 1998. Modelling nature: Between physics and the physical world. *Philosophia naturalis* 35:65–85.

———, and Mary S. Morgan. 1999. Models as mediating instruments. In M. S. Morgan and M. Morrison (eds.), *Models as Mediators: Perspectives on Natural and Social Science*, 10–37. Cambridge: Cambridge University Press.

Murray, James D. 2002. *Mathematical Biology: I. An Introduction* (3d. ed.). New York: Springer.

Nagel, Ernest. 1961. *The Structure of Science: Problems in the Logic of Scientific Explanation*. New York: Harcourt.

Nersessian, Nancy J. 2005. Interpreting scientific and engineering practices: Integrating the cognitive, social, and cultural dimensions. In M. E. Gorman, R. D. Tweney, D. C. Gooding, and A. P. Kincannon (eds.), *Scientific and Technological Thinking*, 17–56. Mahwah, NJ: Lawrence Erlbaum.

———. 2008. *Creating Scientific Concepts*. Cambridge, MA: MIT Press.

Niiniluoto, Ilkka. 1999. *Critical Scientific Realism*. Oxford: Oxford University Press.

Nomura, Masayasu, and Seymour Benzer. 1961. The nature of the "deletion" mutants in the rII region of phage T4. *Journal of Molecular Biology* 3:684–692.

Norman, Donald A. 1993. *Things That Make Us Smart: Defending Human Attributes in the Age of the Machine*. Reading, MA: Addison-Wesley.

Odell, Garrett M. 1980. Qualitative theory of systems of ordinary differential equations, including phase plane analysis and the use of the Hopf bifurcation theorem. In L. A. Segel (ed.), *Mathematical Models in Molecular and Cellular Biology*, 649–727. Cambridge: Cambridge University Press.

Odenbaugh, Jay. 2003. Complex systems, trade-offs, and theoretical population biology: Richard Levins' "Strategy of model building in population biology" revisited. *Philosophy of Science* 70:1496–1507.

———. 2008. Models. In S. Sarkar and A. Plutynski (eds.), *A Companion to the Philosophy of Biology*, 506–524. Oxford: Blackwell.

Olby, Robert. 1994. *The Path to the Double Helix: The Discovery of DNA* (Revised ed.). New York: Dover.

———. 2009. *Francis Crick: Hunter of Life's Secrets*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Orzack, Steven Hecht, and Elliott Sober. 1993. A critical assessment of Levins's *The strategy of model building in population biology* (1966). *Quarterly Review of Biology* 68:533–546.

Plutynski, Anya. 2004. Explanation in classical population genetics. *Philosophy of Science* 71:1201–1214.

Provine, William B. 1986. *Sewall Wright and Evolutionary Biology*. Chicago: University of Chicago Press.

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

Ramsey, William M. 2007. *Representation Reconsidered*. Cambridge: Cambridge University Press.

Rips, L. 1986. Mental muddles. In H. Brand and R. M. Hernish (eds.), *The Representation of Knowledge and Belief*, 258–286. Tucson, AZ: University of Arizona Press.

Robbins, Philip, and Murat Aydede. 2009. A short primer on situated cognition. In *The Cambridge Handbook of Situated Cognition*, 3–10. Cambridge: Cambridge University Press.

Rosen, Gideon. 2012. Abstract objects. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy (Spring 2012 Edition)*. URL = <http://plato.stanford.edu/archives/spr2012/entries/abstract-objects/>.

Roughgarden, J. 1979. *Theory of Population Genetics and Evolutionary Ecology: An Introduction*. New York: Macmillan.

———. 1998. *Primer of Ecological Theory*. Upper Saddle River, NJ: Prentice Hall.

Schemske, Douglas W., and Paulette Bierzychudek. 2001. Perspective: Evolution of flower color in the desert annual linanthus parryae: Wright revisited. *Evolution* 55:1269–1282.

Sklar, Lawrence. 2000. *Theory and Truth*. Oxford: Oxford University Press.

Smocovitis, Vassiliki Betty. 1996. *Unifying Biology: The Evolutionary Synthesis and Evolutionary Biology*. Princeton, NJ: Princeton University Press.

———. 1999. The 1959 Darwin centennial celebration in America. *Osiris* 14:274–323.

Stanford, P. Kyle. 2006. *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives*. Oxford: Oxford University Press.

Stein, Howard. 1989. Yes, but ... some skeptical remarks on realism and anti-realism. *Dialectica* 43:47–65.

Stent, Gunther S. 1971. *Molecular Genetics: An Introductory Narrative*. San Francisco: W. H. Freeman.

Suchman, Lucy A. 1987. *Plans and Situated Actions: The Problem of Human Machine Communication*. Cambridge: Cambridge University Press.

Suppes, Patrick. 1957. *Introduction to Logic*. New York: Van Nostrand Reinhold Company.

———. 1960. A comparison of the meaning and uses of models in mathematics and the empirical sciences. *Synthese* 12:287–301.

———. 1967. What is a scientific theory? In S. Morgenbesser (ed.), *Philosophy of Science Today*, 55–67. New York: Basic Books.

Tarski, Alfred. 1953. A general method in proofs of undecidability. In A. Tarski, A. Mostowski, and R. M. Robinson (eds.), *Studies in Logic and the Foundations of Mathematics*, 3–35. Amsterdam: North-Holland Publishing Company.

Teller, Paul. 2001. Twilight of the perfect model model. *Erkenntnis* 55:393–415.

Thagard, Paul. 2005. *Mind: Introduction to Cognitive Science* (2d. ed.). Cambridge, MA: MIT Press.

Thomson-Jones, Martin. 2010. Missing systems and the face value practice. *Synthese* 172:283–299.

van Fraassen, Bas C. 1970. On the extension of Beth's semantics of physical theories. *Philosophy of Science* 37:325–339.

———. 1972. A formal approach to the philosophy of science. In R. G. Colodny (ed.), *Paradigms and Paradoxes: The Philosophical Challenge of the Quantum Domain*, Chapter 303–366. Pittsburgh, PA: University of Pittsburgh Press.

———. 1980. *The Scientific Image*. Oxford: Oxford University Press.

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

Varela, Francisco J., Evan Thompson, and Eleanor Rosch. 1991. *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.

Vorms, Marion. 2010. The theoretician's gambits: Scientific representations, their formats and content. In L. Magnani, W. Carnielli, and C. Pizzi (eds.), *Model-Based Reasoning in Science and Technology: Abduction, Logic, and Computational Discovery*, 533–558. Berlin: Springer.

———. 2011. Representing with imaginary models: Formats matter. *Studies in History and Philosophy of Science* 42:287–295.

Watson, James D. 1965. *Molecular Biology of the Gene*. New York: W. A. Benjamin.

———. 1980 [1968]. The double helix: A personal account of the discovery of the structure of DNA. In G. S. Stent (ed.), *The Double Helix: A Personal Account of the Discovery of the Structure of DNA*, 1–133. New York: W. W. Norton.

———, and F. H. C. Crick. 1953a. A structure for deoxyribose nucleic acid. *Nature* 171:737–738.

———, and F. H. C. Crick. 1953b. The structure of DNA. *Cold Spring Harbor Symposia on Quantitative Biology* 18:123–131.

Weiner, Jonathan. 1999. *Time, Love, Memory: A Great Biologist and His Quest for the Origins of Behavior*. New York: Vintage Books.

Weisberg, Michael. 2003. *When Less is More: Tradeoffs and Idealization in Model Building*. Ph. D. thesis, Stanford University.

———. 2007. Who is a modeler? *British Journal for the Philosophy of Science* 58:207–233.

———. 2010. Target directed modeling. *The Modern Schoolman* 87:251–266.

———. 2013. *Simulation and Similarity: Using Models to Understand the World*. New York: Oxford University Press.

Wilkins, M. H. F., A. R. Stokes, and H. R. Wilson. 1953. Molecular structure of deoxypentose nucleic acids. *Nature* 171:738–740.

Wilson, Margaret. 2002. Six views of embodied cognition. *Psychonomic Bulletin and Review* 9:625–636.

Wright, Sewall. 1921. Systems of mating. *Genetics* 6:111–178.

———. 1931. Evolution in Mendelian populations. *Genetics* 16:97–159.

———. 1938. Size of population and breeding structure in relation to evolution. *Science* 87:425–431.

———. 1939. The distribution of self-sterility alleles in populations. *Genetics* 24:538–552.

———. 1940. Breeding structure of populations in relation to speciation. *American Naturalist* 74:232–248.

———. 1943a. An analysis of local variability of flower color in *Linanthus parryae*. *Genetics* 28:139–156.

———. 1943b. Isolation by distance. *Genetics* 28:114–138.

———. 1951. The genetical structure of populations. *Annals of Eugenics* 15:323–354.

———. 1965. The interpretations of population structure by F-statistics with special regard to systems of mating. *Evolution* 19:395–420.

———. 1978. *Evolution and the Genetics of Populations, Vol.4: Variability within and among Natural Populations*. Chicago: University of Chicago Press.

———, Theodosius Dobzhansky, and William Hovanitz. 1942. Genetics of natural populations, VII: The allelism of lethals in the third chromosome of Drosophila pseudoobscura. *Genetics* 27:363–394.

Wright, Sewall, Papers. American Philosophical Society.

Yagisawa, Takashi. 2013. Possible objects. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy (Winter 2013 Edition)*. URL = <http://plato.stanford.edu/archives/win2013/entries/possible-objects/>.

Zhang, Jiajie. 1997. The nature of external representations in problem solving. *Cognitive Science* 21:179–217.

———, and Donald A. Norman. 1994. Representations in distributed cognitive tasks. *Cognitive Science* 18:87–122.