Empirical and Pragmatic Grounds of Scientific Representation

by

Dana Matthiessen

BA in Philosophy, Wesleyan University, 2009

MA in Philosophy, Georgetown University, 2015

Submitted to the Graduate Faculty of

Kenneth P. Dietrich School of Arts & Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2022

UNIVERSITY OF PITTSBURGH

Kenneth P. Dietrich School of Arts & Sciences

This dissertation was presented

by

Dana Matthiessen

It was defended on

September 12, 2022

and approved by

James Woodward, Emeritus Professor, History and Philosophy of Science

Michael Dietrich, Professor, History and Philosophy of Science

Eran Tal, Associate Professor, Philosophy at McGill University

Dissertation Director: Sandra Mitchell, Distinguished Professor, History and Philosophy of Science

Empirical and Pragmatic Grounds of Scientific Representation

Dana Matthiessen, PhD

University of Pittsburgh, 2022

The central thesis of this dissertation is that the ability to reason and learn about the natural world using models can be explained in terms of the practices that warrant researchers to integrate models with accounts of their data-gathering procedures and act on their behalf.

I argue that a model only functions as a representation with respect to a target phenomenon when this phenomenon is a plausible member of its domain of application and when the model can be used to characterize this target from data. I argue that this requires, first, that the model can be compared to data and second, that the model be integrated with an account of the process by which this data was produced from the target phenomenon. I provide an account of the representational accuracy of models based on their integration with a theory of technique and subsequent comparison with data patterns. On the same basis, I provide an account of the pragmatic representational content of models in terms of the set of practical inferences they license as a supplement to the empirical programs within the model's domain of application.

Historically, one often sees a back-and-forth negotiation where a model-based target characterization and a data-gathering practice are iteratively tuned to one another. Models are routinely informed by empirical results in the process of their construction and adjusted in response to them. Conversely, models add depth to target characterizations and fill out theories of technique in ways that alter data-gathering procedures. From this perspective, we can understand how a model's representational content might gradually accrue to it and allow for finer distinctions in data outcomes. I present an extended case that tracks the development of X-ray crystallography and its use for the characterization of the molecular structure of proteins.

Ultimately, what is presented here is intended as a robustly pragmatist account of scientific representation. That is, one that does not only tie model use to purposes, but also to the realm of human action.

Table of Contents

List of Tables

List of Figures

Preface

This dissertation ties together the courses and relationships that shaped my philosophical interests in HPS. First and foremost, a course with Sandy Mitchell and Mazviita Chirimuuta on the epistemology of experimental practice. My final project in that course was a study of different methods of data production and processing in structural biology. It then led to my comprehensive exam essay written under Sandy, where I produced the first writing that would become parts of this dissertation. This was also the beginning of the empirical work that lies at the center of the dissertation—an embedding within the University of Pittsburgh's laboratory of Structural Biology and Molecular Biophysics and to the historical case study running through the last half. Another course, forming the other side of this dissertation's theme, was taught by Jim Woodward and Sandy Mitchell on the theme of models and modeling. Sandy being the constant, it was fitting that she be my dissertation director, and I owe a great deal of its completion to her continual enthusiasm about this project, our weekly meetings at its onset, and her responses to my requests at the later stages. I am indebted to her work on integration and modeling in biology, as the final chapter might suggest, and am delighted to be at a point in where I can provide helpful feedback on her work as well. Woodward taught the first general philosophy of science course that I took as a graduate student and I've learned a great deal through exposure to his philosophical intuitions, in addition to his written work. Above all, I appreciate his ability to cut through the fat in my writing and his skill in very politely bringing up distinctions that threaten to undermine my views. Mike Dietrich has provided the foundations for my understanding of the history of molecular biology, along with a substantial amount of philosophy of biology. His course on the former topic led to the second bit of writing that has found its way into this dissertation. My interest in devoting a large portion of this dissertation to historical themes and my comfort in pursuing them are thanks to Mike (including his recommendation to take a course on historical methods with Mari Webel). Finally, I am grateful for Eran Tal's willingness to join my committee, his careful attention to my writing, and the time he has spent engaging in some very helpful back-and-forth emails. Eran's written comments have led to crucial conceptual clarification of this work in its later stages.

In addition to my committee members, I have to thank Kareem Khalifa, who attended a talk of mine at the Society for the Philosophy of Science conference in Ghent and was kind enough to share his interest afterward. It's been rewarding to develop a philosophical friendship with Kareem over shared drafts. Many thanks are due to Angela Gronenborn for permitting me to spend time shadowing her student Andrew Bellesis, and many to Andrew himself and everyone else at the lab for putting up with my persistent questions. Ying Wu was another lab guide who went so far as to teach me to make my own crystals. Kevin Sapp graciously agreed to an interview regarding his experience working on the ATLAS team. Other helpful advice, comments, and appreciated attention to my work have come from Gal Ben Porath, Nora Boyd, John Bickle, Bill Bechtel, Letitia Meynell, Peter Tan, Shahin Kaveh, Tom Wysocki, Katie Morrow, Dasha Pruss, Jennifer Whyte, William Penn, Rose Novick, and several anonymous reviewers. Beyond my committee members, I am intellectually indebted, more or less subtly, to many other teachers. Among them: Mark Wilson, Bob Batterman, John Norton, Porter Williams, Bryce Huebner, Mark Lance, Quill Kukla, Arne Hoecker, and Joseph Rouse. I'm sure there are others I will regret leaving out. Special thanks to Lindley Darden and James Mattingly for giving me—someone whose first interests were history of philosophy—the encouragement and confidence to pursue work in this sub-field. And to Paolo Palmieri for giving me an opportunity to enjoy life and rediscover my love of scholarship when I was on the verge of burnout. Thanks, of course, to my dear friends and family, whose delightful presence constantly tempted from finishing this work. And a final thanks to Hillary Steffes for her patience and support during my periods of reclusive writing and research.

Introduction

Over the last twenty-five years, a body of literature has emerged in the philosophy of science that seeks to explain how scientists are able to use abstract models as surrogates for reasoning and learning about the natural world. The central thesis of this dissertation is that this capacity for surrogative reasoning can be explained in terms of the practices that warrant researchers to integrate models with accounts of their data-gathering procedures and act on their behalf. This is an alternative to present accounts in the literature, which tend to explain surrogative reasoning by appeal to a relation between model and target, typically in the form of a component-wise mapping between model and target features. This relation is variously presented in more substantive or cognitive terms; that is, as a relation that is either "in the world" or "in the head." The trouble with accounts on both sides of this divide is that they tend to neglect the epistemic standpoint of practicing scientists, including the way scientists develop and evaluate their models. Instead, they tend toward an ontic account of representation. They treat the representation relation as something resides outside of these practices, where the worldly content of a model can simply be "read off" by a competent scientist. I call the fundamental assumptions underlying this approach Ontic Priority and Semantic Imperialism.

Both of these assumptions are mistaken. They presuppose that a well-characterized target is available for surrogative reasoning, whereas this is often not the case. As a result, they are confronted with a basic paradox when it comes to explaining surrogative reasoning and the accuracy of scientific models: scientists often lack the ability to verify that a mapping relation obtains between model and target. So, *a fortiori*, do we all. Yet, scientists working in these conditions are routinely able to reason in productive ways about their targets. Attempts to address this paradox along lines that assume Ontic Priority and Semantic Imperialism fail.

Instead, I argue that we should consider models in terms of the way they are used by scientists. Mine is not the first use-based account of scientific representation, but it differs from most (perhaps alongside van Fraassen's account) in its strong emphasis on the way models are tied to empirical practices.

I argue that a model only functions as a representation with respect to a target phenomenon when this phenomenon is a plausible member of its domain of application and when the model can be used to characterize this target from data. I argue that this requires, first, that the model can be compared to data and second, that the model be integrated with an account of the process by which this data was produced from the target phenomenon. I call such an account a *theory of technique*. I devote a substantial portion of the dissertation to elaborating how theories of technique function as part of a body of empirico-technical knowledge. This knowledge is centered on the network of causal relations that accompany scientists' interactions with a target phenomenon—a kind of "experimental niche." I call this knowledge an *empirical program* and argue that scientists are warranted in treating data as reliably connected to a target phenomenon when their data-gathering procedure is directed by an empirical program.

The reliability of results produced under an empirical program is a historical achievement. It results from the arduous search for points of stability within the causal thickets that surround natural phenomena. Once scientists have an account of the causal relations that tie a phenomenon to its surroundings—what I call a superficial characterization—they can venture claims about the internal structure of a phenomenon. These structures help "fill in" the superficial account of the dependency relations between a phenomenon, a data-gathering apparatus that interacts with it, and the other causal factors that accompany this interaction. When parts of models can be interpreted in terms of these causal factors, they provide researchers with defeasible warrant for performing specific interventions that may not be otherwise suggested by a theory of technique, and for anticipating the resulting data patterns. These are the basic ingredients of the representational use of models: a model is integrated with a theory of technique and used to provide a characterization of a target phenomenon that is responsible for the data gathered with these techniques. I provide an account of the representational accuracy of models based on their integration with a theory of technique and subsequent comparison with data patterns. On the same basis, I provide an account of the pragmatic representational content of models in terms of the set of practical inferences they license as a supplement to the empirical programs within the model's domain of application. I explain this supplementary function in terms of "model-entry" and "model-exit" moves—inferences that carry a user

between the internal inferential rules of the model and the theory of technique, which are licensed by an interpretation of model components in terms of factors within the latter theory. By this means, a model is connected to the practical activities of researchers.

Historically, one often sees a back-and-forth negotiation where a model-based target characterization and a data-gathering practice are iteratively tuned to one another. Models are routinely informed by empirical results in the process of their construction and adjusted in response to them. Conversely, models add depth to target characterizations and fill out theories of technique in ways that alter data-gathering procedures. From this perspective, we can understand how a model's representational content might gradually accrue to it and allow for finer distinctions in data outcomes. Over the long run and with the proper training, scientists come to talk about and use models in ways that blur the distinction between the model and its target qua causal factor in data-gathering setups in a way that suggests a mapping relation between the two. The role of empirical results in establishing a model's applicability to a particular domain is obscured.

I thus appeal to the history of science for two reasons: to give a diachronic account of how certain practices become reliable sources of data with respect to a given target. And to provide evidence for the claim that a model's representational use is indebted to these practices—through their contributions in determining its domain of application and in providing data for comparison to the model that is reliably linked to its target. Because of this, history plays a significant role in the second half of the dissertation. I present an extended case that tracks the development of X-ray crystallography and its use for the characterization of the molecular structure of proteins. I wanted to provide enough of an overview so that one could see the broad developments in target characterizations and data-gathering techniques, but also give sufficient detail so one could trace moments of close interaction and reciprocal influence between these. This is quite the balance to strike in a dissertation that also aims to do some philosophical work, and I've erred on the side of providing more details.

Ultimately, what is presented here is intended as a robustly pragmatist account of scientific representation. That is, one that does not only tie model use to purposes, but also to the realm of human

action. On my view, scientific representation is not only "out there" in the world, nor is it strictly "in the head." Rather it is constituted through broader scientific practices of model construction, reception, and evaluation in ongoing relation to empirical programs; whatever can be said of it qua relation between model and target is given through these practices.

Structure of the dissertation: Chapter 1 will review the relevant literature and present my critique of accounts of scientific representation that assume Ontic Priority and Semantic Imperialism. Chapter 2 will give an overview of the elements of a use-based account of scientific representation that emphasizes the link between models and data-gathering practices. In particular, it will specify two necessary conditions for surrogative reasoning: a target must fall within a model's domain of application and the data generated from this target must be reliably linked to it.

The following chapters begin an excursus on data-gathering practice, Chapter 3 will present the contribution of empirical programs to the reliability of data. Chapter 4 will trace the development of the empirical program of protein X-ray crystallography. Chapter 5 provides a philosophical commentary on the history presented in Chapter 4, emphasizing the increase in reliability of empirical programs over time. Finally, we return fully to the philosophy of scientific representation in Chapter 6, where I defend the claims of Chapter 2 in greater detail and present an account of representational accuracy and representational content in terms of the integration of a model with theories of technique. I argue that this account is sufficient to explain the ability to reason and learn about the world with models and that it does so without presupposing the kind of mapping relation between model and target that creates so much trouble for other accounts. More detailed chapter summaries are provided below.

Chapter one: two basic desiderata of a theory of scientific representation are the following: an explanatory account of surrogative reasoning—i.e., how models allow a user to draw informative inferences from models to targets; and an account of what makes models accurate with respect to their targets. Much of the current literature on scientific representation is divided into two camps. The first takes a component-wise mapping view of the representation relation, motivated by two theses: Ontic Priority and Semantic Imperialism. Attending to common circumstances in science, however, shows that this view leads to a

paradoxical claim: the relation enabling surrogative inference, the proper instantiation of which is the criterion of representational accuracy, is epistemically inaccessible for many users of models. And yet they are able to reason about targets and evaluate their models. The second camp responds to these issues by claiming that a general account of representation cannot be given. Neither gives a satisfactory explanation of surrogative reasoning or representational accuracy from the standpoint of scientific practice. But there is room for a general account to be developed from a shift in focus to scientific activities, in particular those involved in establishing a means of interface between model and target.

Chapter two: An analysis of the notion of informative inferences, taken in light of the failures of accounts reviewed in chapter one, shows that contextual constraints on model use need to be taken into account in order to distinguish the objectivity of scientific representation from the subjectivity of stipulative fiat. To be informative, an inference must exclude alternative possibilities with respect to its subject matter. This requires that we examine the ways scientific representation is practically constrained in ways that establish a model's relevance for a given representational use.

Two spheres of activity are subject to relevance conditions: those on the "model side" and those on the "target side"; that is, application techniques and empirical techniques. To serve as a representation for a target phenomenon, a model must be applicable to the target domain and connected to data that is reliably gathered from this target. These two relevance conditions are manifest in the activities involved in model application and data gathering. Combined with conclusions from Chapter 1, they form a set of conditions for scientific representation that provide a preliminary explanation for the ability to draw informative model-based inferences about a target phenomena. The virtues of this framework are discussed and objections are considered.

Chapter three: Empirical techniques are an important element of the activity-based view. In many cases, a model must interface with the output of an empirical program, which consists of two parts: (i) an established procedure, consisting of activities that carry a researcher from preparation of a data-gathering set-up to a processed result and (ii) a body of knowledge providing a practical rationale for the actions taken in this procedure. Philosophical considerations, along with ethnographic study of an actual procedure,

allows for the identification of basic features of this body of knowledge, the form of practical reasoning and action that it warrants, and its role in producing reliable data.

To be reliable, data must be consistent, coherent, and comparable. The first two conditions are satisfied by data that meets standards of statistical and systematic uncertainty. The former is evaluated through the statistical analysis of data produced by instruments and met by gathering sufficient quantities of data to account for their variance. The latter is evaluated by ensuring a procedure's conformity with a theory of technique—that is, an account of the causal factors within the data-gathering setup that contribute to a data outcome. Reliability is further ensured through general practices of record-keeping, cross-checking, and procedural revision, which make up for gap in the theory of technique. Finally, data is rendered comparable to models by processing techniques that remove artifactual contributions to data that could not be accounted for by procedural means. Once again, the theory of technique plays an important role in this process.

Chapter four: The data resulting from an empirical program were said to be reliable because they result from actions guided by a body of knowledge. The justification of this body of knowledge is a historical process. A detailed case study shows how this results from an accumulation of techniques and second-order concerns that develop dynamically in reciprocal interaction with a conception of the system under study. The reliability of results comes from the cumulative stabilization and integration of empirical techniques into more expansive forms of data generation and model-based inquiry. This chapter serves as the historical prelude to the philosophical work of the following chapter.

Chapter five: The present chapter is devoted to a philosophical distilling of the historical developments covered in Chapter 4. It recasts various episodes within this history in terms of general patterns and uses these to characterize features of empirical programs qua complex developmental systems. These are: the search for invariances within the "parameter space" of experimental conditions; the stabilization of procedures around such invariances; the incorporation of resulting "points of stability" into extended procedures and hence the scaffolding of novel techniques on prior results; the reciprocal influence between model-based target characterization and experimental techniques; and higher-scale developmental tendencies that emerge from these. I discuss how these patterns contribute to the increased reliability of

data production and interpretation procedures and outline a corresponding theory of historical justification for bodies of knowledge that undergo regular alteration.

Chapter six: In this chapter, I will incorporate these results into an account of scientific representation. I use the example of Laue and Bragg's modeling strategies to advance an account of representational accuracy that depends on a model's integration with an empirical program. From an account of these integrative practices, I derive a traditionally pragmatist notion of representational content for scientific models. I argue, in short, that representational content should be understood interactively, in terms of the action-guiding inferences that are licensed in relation to a given empirical program. I show how this elaborates on elements of the use-based account of scientific representation presented in Chapter 2 and highlight several consequences of the view. In particular, I claim that the process of characterizing a target with empirical results is inferentially linked to the process of representing that target with a mode. Throughout, my view is contrasted with the assumptions that underlie Ontic Priority and Semantic Imperialism. I argue that this view offers an explanatory account of surrogative reasoning that avoids these assumptions, and therefore avoids the Paradox of Scientific Representation.

1.0. Toward an Explanatory Account of Surrogative Reasoning

Scientific models are used as means to make claims about nature. By studying their properties and behaviors, scientists draw inferences about the behavior of target phenomena. Models are thereby said to represent these targets. Representation allows model users to reason about target phenomena in an informative way by reasoning about the model. This capacity for *surrogative reasoning* calls for explanation. How is it that scientific models allow users to draw inferences about a target natural system that are informative? Informative inferences typically involve positive or negative attributions of states, properties, or behaviors to a target phenomenon in a way that is specific to this target and goes beyond knowledge explicitly assumed in a model's design. That an inference is informative with respect to a target does not entail that it is accurate; such attributions may lead to stark failures. To be judged accurate, a model must satisfy some further criterion that accounts for the *success* of the inferences it enables in contexts of empirical application. What, then, accounts for the accuracy of those models that allow for informative inferences? Without satisfactory answers to these questions the contribution of representational models to scientific knowledge remains obscure. And yet, predominant accounts of scientific representation fail to explain surrogative reasoning in terms that are adequate to the practice of science.

I will begin this chapter with some opening remarks on scientific representation. I then critique the predominant views on this topic. Accounts that offer an explanatory theory of surrogative reasoning fall into two broad camps, which I call substantive and inferentialist accounts. Substantive accounts treat representation as a naturalized relation between a model and a target, which may depend in part on the cognitive acts of a model user. Viewed from a high level of generality, many substantive accounts can be seen to share basic assumptions in their philosophical methodology that I refer to as *Ontic Priority* and *Semantic Imperialism*. These two theses lead to a view of scientific representation that is deeply misleading. For one, they account for representation in terms of a component-wise mapping relation between the features of a model and the features of its target; model features are said to be similar to those of its target, or the structure of the model is said to be related to that of the target by a type of structural morphism. I

argue that these views are typically motivated by examples that fail to reflect the epistemic situation of much of contemporary science. They are elaborated through examples where a modeler has a description of the target system on hand that has been arrived at independently of their model, as is the case with models of macroscopic objects familiar to everyday life such as pendulums and rail networks. This is not the case with many phenomena studied by the sciences, which problematize the notion that representation functions by mapping a model onto target features, since these features are only known in terms of a model.

I call this issue the *Paradox of Scientific Representation*: scientists routinely lack epistemic access to the substantive relations on which (philosophers assert) surrogative reasoning and representational accuracy depend. And yet, they are able to reason with their models and evaluate them with respect to empirical results. This is a serious puzzle for the philosophy of scientific representation. I argue that attempts to address it with substantive views are failures. In light of these failures, one response found in the literature is to throw one's hands up and refrain from giving an explanatory account of scientific representation at all by adopting a deflationary inferentialist view. Such views hold that models represent in virtue of their enabling surrogative reasoning and that no general account of this capacity can be given. But by adopting this view we simply abandon the project of accounting for how models enable surrogative reasoning, which I take to be central to understanding the epistemic nature of models.

In laying out these issues, this chapter prepares a way forward for a theory of representation rooted in scientific practice. I will present an account of scientific representation that explains how models allow for surrogative reasoning about their targets—and successful reasoning in particular—in terms that give a role to data-gathering practices. In this way, I diverge from deflationary and substantive accounts on offer. What these lack is an adequate focus on the role of empirical procedures in forging a relation between a model and its target.

1.1. Preliminary remarks on scientific representation

Representation is variously described as a relation between two terms, or an activity that relates them—a *significans* or representational vehicle, and a *significandum* or representee. When something is described as a representation, I take this to be shorthand for stating that it stands in some such relation or is related in this way to a (possibly unspecified) representee. In the philosophy of science literature, the representational vehicle is typically a model system and the representee is a target phenomenon. While the subject of scientific modeling has been of interest since at least Hesse's work (1963), the nature of scientific representation has come to the fore among philosophers of science in recent decades.

It is an open question whether all representations serve a similar purpose or should be thought of in similar terms. Some philosophers seek to account for representation *tout court* (as in (Bueno & French, How Theories Represent, 2011)), or even deny that a clear distinction can be drawn between scientific and non-scientific representations (Callendar & Cohen, 2006). My focus here will be narrower, on the use of theoretical scientific models to represent natural systems. I do not attempt to give an exhaustive definition or categorization of theoretical models, considering how these differ substantially across fields (Cf. (O'Conner & Weatherall, 2016)). I generally intend to speak of those models—be they mathematical, computational, mechanistic, or specified through more qualitative means—that students of a scientific discipline have been trained to work with, that researchers routinely construct and manipulate, and that philosophers have routinely regarded as important components of a scientific theory (Suppes, 1962; van Fraassen, 1980; Cartwright, 1983; Lloyd, 1988; Giere, 1988; Suppe, 1989; Morgan & Morrison, 1999). Theoretical models are designed to enable reasoning about some target system without explicit reference to or inclusion of features of a data gathering process. I mean to include such diverse creatures as classical physical models of celestial orbits, quantum mechanical models of the hydrogen atom, models of predator-prey dynamics, models of ecological networks, models of gaseous atoms diffusing in a container, models of molecular structures, models of intra-cellular processes, models of economic agents and macro-dynamics, and so on, while excluding models of particular experiments, data models, or models used to

correct data. I also do not assume that models in science are reducible to interpretations of the logical axioms of a scientific theory. By limiting my considerations to these (still varied) cases, I also refrain from considering whether artistic representations or mental and linguistic representations enable such surrogative reasoning in the same manner as scientific representations.

Schematically, I take a model to be a theoretical object comprised of (i) elements bearing certain properties and (ii) relations between these elements and their properties. In typical cases of mathematical models, elements are described in terms of properties that take on variable or constant values, such as the position, mass, and velocity of a moving particle. Relations between elements may be expressed in terms of specific expression—such as a mass-dependent force law or a condition stating the value of a spatially varying property at a boundary region. When combined with a set of internal inference rules, models allow users to reason about them. This reasoning may pertain to the properties of a particular state of the model system, in which the values of its variable elements are fixed. Many useful model systems also possess internal dynamics, described in terms of transition rules (e.g., differential equations) and constraints (e.g., conservation principles), which allow users to determine how the model transitions from one state (comprised of a determinate set of elemental and relational property values) to another. Often this allows users to derive further implicit properties of the model system. For example, the standard description of the Lotka-Volterra model of predator-prey relations does not specify that it contains an unstable non-zero fixed point. This can be discovered by studying its properties using the appropriate mathematical rules of inference, and determining, for a given choice of parameters, the variable values at which the derivatives of predatory and prey populations equal zero.

A model description, as I conceive of it, goes beyond the general laws or principles of a scientific theory. It lays out a set of assumptions instantiating some collection of theoretical principles and does so in a way that constitutes a closed inferential system or family of systems. This allows users to derive specific claims about the model system's properties or behaviors. For instance, Einstein's field equations for general relativity are not in themselves a model, but a particular solution to these equations such as the Schwarzschild metric is a model. It allows claims to be made about the trajectories of point masses in

spacetime. Newton's second law alone is not a model, but when paired with an equation describing gravitational force and a description of a point particle in free fall, it allows users to infer the force acting on that particle. A philosophical account of scientific representation seeks to understand how reasoning about a model system of this kind allows scientists to draw inferences about a target phenomenon in the world.

1.2. Ontic Priority and scientific practice

1.2.1. Ontic-first approaches in the literature

Philosophical accounts of scientific representation vary in their model ontology, the nature of the objects and relations in terms of which they seek to explain representation, and the extent to which they stress contexts of use. Among those who commit to specific model ontologies, some describe models as fictional objects (Frigg, 2010a; Godfrey-Smith, 2006), others as abstract entities or structures broadly construed (Teller, 2001; Giere, 2004; Weisberg, 2013), and still others as strictly mathematical structures (Bueno, 1997; French & Ladyman, 1999; Pincock, 2011a). These commitments are accompanied by differing accounts of the representation relation between a model and its target. However, many of these accounts endorse a common set of assumptions. This includes the thesis of Ontic Priority: the view that a theory of scientific representation must first possess a particular account of the *what*—the nature of scientific representation qua model-target relation—in order to explain the *how*—that is, the ability to use models to reason about target phenomena. A typical advocate of Ontic Priority specifies how distinct features of a model and a target phenomenon relate to one another and explains the ability to reason with models on this basis. In some cases, this is supplemented with an appeal to the goals or cognitive operations of a model user. In either case, philosophers seek to explain surrogative reasoning and the representational accuracy of models independently from criteria involved in evaluating the model, say, with respect to data generated from that target. That is, they seek to explain representation in terms of "ontic" conditions rather

than "epistemic" ones. Call this an ontic-first approach. A commitment to Ontic Priority thus involves two conjoined claims: first, that a set of ontic conditions are responsible for, and so explain, the outcome of epistemic evaluations of modeling accuracy; second, that an investigation of these conditions is therefore methodologically prior in any effort to establish accuracy criteria.

This commitment can be found throughout the literature on surrogative reasoning and representational accuracy. The author of a thoroughgoing review article on scientific representation,[1] James Nguyen writes, "We still do not have an account of in virtue of what scientific models represent their targets. This is particular worrying given that it is plausible we should answer this question before we investigate representational accuracy" (2016, p. 188). Nguyen cites Suárez (2004), Contessa (2007), and Frigg (2010b) as authors who, despite their differences, "all take the question of representation as conceptually prior to accurate representation," (p. 188). There is a sense of conceptual priority that is trivially satisfied here, insofar as the notion of accurate representation is a species of representation simpliciter. Less trivial is the philosophical methodology this recommends. Such priority for Nguyen appears to call for a specific order of investigation: first establish what grounds scientific representation, then seek to explain surrogative reasoning and accuracy on this basis.

The main concern of this section is those who seek to characterize representation in terms of an ontic relation between model and target.[2] For these authors, Nguyen's worry might be expressed as follows: how can we judge that a model represents its target, and so explain surrogative reasoning or its successes, without an account of the relation that the model stands in with respect to its target? This is precisely how Chakravartty defends "informational" accounts of representation, which appeal to model-target similarities:

> How, one might wonder, could such practices [of model-based interpretation and inference] be facilitated successfully, were it not for some sort of similarity between the representation and the thing it represents—is it a miracle? […] Indeed, it is precisely *because* the informational view is satisfied in this way that cognitive activities such as interpretations and inferences regarding target systems are successful in the first place (2010, pp. 201-203).

[1] Cf. Frigg and Nguyen (2017).
[2] In other words, I am largely putting aside deflationary views, such as Suárez's.

Bueno and French draw directly on this argument to support their partial morphism account of representation: "It is the requirement that such facilitation be accommodated that effectively demands that there be some relation of similarity that holds between the representational device and the target system" (Bueno & French, 2018, p. 65). Here inferential success is taken as a fact that, by a form of no-miracles argument, necessitates the existence of a particular kind of model-target relation. "[W]ithout it, the success of the very functions that functional [e.g., inferentialist] accounts take to be central would appear to be inexplicable" (Chakravartty, 2010, p. 201). In other words, one must first specify the nature of the model-target relation in order to explain the successes of model-based inference.

On the other hand, "scientists *discover* that their model is a misrepresentation (typically) through lack of empirical success," spurring the creation of improved models. A partial morphisms account will accordingly describe "the interrelationships between the increasingly successful models that, because of this success, can be regarded as increasingly better, more informative representations." (Bueno & French, 2018, p. 69). Improvements in accuracy are accounted for in terms of progressive mappings from less to more successful models. Whatever modicum of accuracy is possessed by an earlier model in this sequence is explained by structural similarities to its target. These are imported into more successful offspring by structural similarities between the model and its successors, thereby preserving the kernel of model-target relation responsible for the accuracy of the prior model. Scientists' confirming a model's accuracy may spur the production of improved models, but this process supervenes on the structural relations that underwrite their successes and failures. For Bueno and French, the practice of confirming accuracy—say, by checking it against data—is a separate epistemic matter from the inferential power of the model itself.

Weisberg's *Simulation and Similarity* agrees on this last point: "I will not primarily be focusing on experimentation, data, or confirmation. Before one can develop a theory of confirmation for models, one needs to be clearer about the nature of the model-world relationship" (2013, p. 90). This relationship is given by Weisberg's weighted feature-matching account. On this view, similarities between a model and target are evaluated in terms of the degree of fit between model features and target features, with some points of overlap or divergence given more weight than others. As a whole, this view is meant to

characterize "how scientists represent the world with models" and "how their representational goals and ideals shape the standards of fidelity that they apply" (p. 135). In doing so, it "reflects judgments about the relationship of models to their targets that scientists can actually make, because it draws on resources that are cognitively available: feature sets and weighting functions" (p. 155).

Again, this approach seeks to explain surrogative reasoning and derive criteria of accuracy ("standards of fidelity" in Weisberg's parlance) from an account of the model-world relationship in a way that is independent from—and deemed conceptually prior to—matters concerning experimentation, data, or confirmation. To understand the rationale for separating these topics, we might look to remarks by Frigg, who likewise argues against the notion that a model's relation to data plays a role in scientific representation. Frigg claims that data are restricted to "an evidential function […] for the fact that the model is a (more or less) faithful representation of what is happening in the world" (2010b, p. 111). From this standpoint, representation is a matter of establishing a particular relation between model and target phenomenon, standards of accuracy are understood in terms of whether this relation in fact obtains, and the data gathering process is simply a means of producing evidence for this fact.

This shows that the two claims associated with Ontic Priority are readily found in the literature. Authors routinely seek to define a model-target relation that is independent from a particular class of epistemic activities, such as data collection, and explain surrogative reasoning and representational accuracy therefrom. This relation is the principal factor that explains scientists' reasoning and judgments; the assessment of a model in light of data is thus relegated to a secondary function from the standpoint of philosophical methodology, serving only to confirm or disconfirm a presumptive model-world relation.[3] This is not to claim that authors completely disregard the role of epistemic and pragmatic factors in their accounts of representation. Agreement with practice is a widely shared goal in philosophy of science, and

[3] It should be noted that Weisberg presents his view as one concerned with the epistemology of model, and that he sometimes wavers on this demotion of data collection. He notes that empirically collected data are "used to make inferences about the nature of targets and mathematical representations of targets" (p. 96) and so play a foundational role in characterizing a model's target. At another point he states that "Modeling practice often involves an interaction between the development of the model and the collection of empirical data" (p. 155). Still, there is a general neglect in this literature of both the process of target characterization (though see (Elliot-Graves, 2020)).

many take care to note that aspects of the interpretation of models depend on "the research questions of interest, the context of research, and the community's prior practice" (Weisberg, 2013, p. 149) or on factors "having to do with the use to which we put the relevant models" (Bueno & French, 2018, p. 66). However, these are often incorporated as addenda or specifications within a framework in which the order of philosophical inquiry begins with the nature of the model-world relation, without reference to epistemic practices.

1.2.2. Component-wise mapping and the Paradox of Scientific Representation

What notions of the model-target relation do these accounts propose? It is useful to distinguish between consideration of the products and the activity of science, or what Van Fraassen (2008) calls science viewed from "from above" and "from within."[4] The view from above is retrospective, considering science in terms of its achieved theories and models in abstraction from the historical practices from which they resulted. The view from within seeks to characterize science from the standpoint of these practices themselves, restricting its arguments to epistemic resources available to working scientists. These views are logically compatible, though approaches to an explanatory project for scientific representation will differ depending on which view is emphasized. Over-emphasis in either direction may carry risks: singular focus on practice can leave one swimming in particularities, while an excessively elevated view may lead to obscurities or formulations of questionable relevance to science in practice. This presents a problem, as a basic desideratum for a theory of scientific representation is that it be adequate to the practical realities of science in the making.

For instance, at a high level of abstraction the relation between a model and its target may be described "from above" as a relation between model components and some aspects of a target phenomenon. One can then decompose a model into constituents (be they elements, states, properties, relations, or

[4] Earlier instances of this distinction are found, e.g., in Cartwright (1983), Longino (1990) and Pickering (1992).

transition rules), and define representation as a specific relation between these model constituents and the focal aspects of a target. Candidates may include forms of similarity, structural, or a comparable denotative mapping between components. One could argue that a model user is able to reason about and draw inferences with respect to focal aspects of a target phenomenon in virtue of this relation. Further, the relation could be investigated through a *comparison* of the model and target so characterized, in which the truth of statements linking model and target features is checked. Many accounts of scientific representation agree on such a picture of the representational capacity of models, based on the following assumptions:

1. There is a target phenomenon *T* in the world, described in terms of its structure or features[5] *TF*
2. There is a model system **M** specified in terms of its structure or set of features **MF**
3. **M** allows inferences to be drawn with respect to *T* in virtue of a relation R between **M** and *T* such that R connects (a sub-set of) members of **MF** to (a sub-set of) members of *TF* in a particular way.

For instance, *T* may be a hydrogen atom with its component parts *TF* of a single proton nucleus and orbital electron. **M** may be the Bohr model of this atom with its structure **MF** given by Bohr's postulates describing the constraints on energy levels, angular momenta, and transitions of the orbital electron, along with those inferences licensed through mathematics and background physical assumptions. R would then describe the connection between, e.g., the orbital electron behavior in the model **M** and the actual electron in *T*.

Textual evidence supports the claim that this picture of representation is commonplace. Among fictionalists, models qua fictional objects apply to real-world phenomena in virtue of the resemblance relations between the two. Models allow for an accurate understanding of natural phenomena because the properties of fictions can be compared to the properties of these systems, in the same way that the events in Orwell's *Animal Farm* can be compared to events in Soviet Russia, or the way a friend can be compared to a fictional character:

> Representation of a real-world system involves two distinct relations, the *specification* of a model system and some relevant *similarity* between model system and the world itself

[5] The term 'features' is here left intentionally vague, as it is meant to include various notions of what in the model is connected with what in the target. These items could be characterized as elements, parts, properties, relations, processes, and so on.

> […] Many of the special features of model-based science come from the role played by resemblance relations between model system and target (Godfrey-Smith, 2006, p. 733).
>
> In each case, what we seem to be doing is comparing properties associated with one system with properties associated with another […] Whichever way the details go, these comparisons are guided by mappings between properties of the kind exhibited by ordinary objects (Godfrey-Smith, 2009, p. 105).
>
> [T]he apparent comparison [of a real system] with a nonexistent object eventually comes down [to] the unproblematic comparison of properties, and the statement making this comparison is true iff the statement comparing the properties with each other is true (Frigg, 2010a, pp. 263-264).

On the face of it, this is very close to the views expressed by those who regard models to be *abstracta* that must be related in a particular way to concrete systems. Again, there is a close connection between the similarities or resemblance between the model and target and the ability to compare one to the other:

> One way, perhaps the most important way, but probably not the only way [that scientists use models to represent] is by exploiting *similarities* between a model and that aspect of the world it is being used to represent […] It is the existence of the specified similarities that makes possible the use of the model to represent the real system in this way (Giere, 2004, pp. 747-748).
>
> Modelers intend talk of similarity between a concrete system and a model as an abstract object to be understood as a comparison between the model and the properties—perfectly respectable abstract objects—instantiated by the concrete object being compared (Teller, 2001, p. 399).

Finally, those endorsing a broadly structuralist view according to which models are best viewed in formal, often set-theoretic, terms invoke the notion of a mathematical correspondence between model features and phenomenon features:

> the inter-relationships between physical structures are most appropriately captured by partial isomorphisms, those between mathematical structures and both physical structures and other mathematical structures are best captured via partial homomorphisms (Bueno, French, & Ladyman, 2012, p. 45).
>
> [This account] is meant to grasp some basic feature of our 'intuitive' notion of empirical adequacy, a particular kind of correspondence between the observational level of scientific theories (represented by the empirical substructure) and the empirical phenomena (as they are found in the appearances). It was exactly such a correspondence that the notion of isomorphism was supposed to apprehend (Bueno, 1997, p. 597).

Despite their differences, these characterizations of the model-target relation all suggest that the notion of a particular kind of relation between a model and its target allows one to speak sensibly of scientists drawing

comparisons or correspondences between the features of a model and those of its real-world target on the basis of a representation relation R between them.

For some, such as Teller, Giere, and Weisberg, R will pick out certain similarities between *T* and **M**, defining a resemblance relation between them. This involves specifying which members of **MF** are meant to stand for which members of *TF*. As Giere and Godfrey-Smith are quick to point out, the choice of R is typically context-sensitive, depending on the intentions and purposes of the researchers who are putting **M** to use. Pincock (2011b) states that R is given by a set of propositions regarding the correspondences between parts **MF** of a mathematical entity **M** and the physical features *TF* of a system *T*. For most structuralists, R will take the form of a structure-preserving map between a target structure *TF* and the structure of a model **MF**.[6]

Regardless of particularities, the set of assumptions described above is enough to identify a common view of model-target relations in terms of a component-wise mapping. This encourages us to think of an instance of representation as a matter of independently considering a model on one hand and a target phenomenon on the other, and then "comparing" or mapping the features of one to the other. These acts of comparison are commonly held to be undergirded by a relation of similarity or mathematical morphism between model and target features. Authors claim that it is by means of this relation R, however specified, that a model can be compared to a target system. But claim **1** of the basic picture of representation involves an issue that is not often discussed by advocates of the mapping view: it is routinely assumed that there is a description on hand of the mathematical structure or relevant features of the target system. How did we arrive at such a description of the target system? How do we know which structural or non-structural features the target possesses and which it does not? And most importantly, what basis do we have for

[6] These structures can be defined set-theoretically as sets of elements that can be ordered according to a set of relations, each of which are sub-divided into three parts listing those elements for which the relation holds, those for which it does not, and those for which it is undetermined. When R obtains between a mathematical structure **M** and a physical structure *T*, it is defined as a mapping that takes elements of **M** to elements of *T* such that there is an exact analogue of each relation (with its sub-divisions) in **M** defined over the elements of *T*.

treating this description as something that is separate from a theoretically-informed model such that it allows for a comparison between the two?

One possible response to these questions is to claim that scientists possess means of describing target systems independent of any model. This is straightforward for some toy examples. Authors appeal to cases such as a mathematical model of a pendulum (Giere, 2004) or the use of pens and sheets or paper to represent ships on the sea (Suárez, 2004). Both targets can be described in terms of some observational vocabulary that does not depend on a prior understanding of the features of their candidate models. Similarly, Godfrey-Smith (2009) elaborates his fictionalist account of model representation by analogy to a reader's comprehension of a Sherlock Holmes story: Sherlock Holmes is not a real person, but we can comprehend his actions and compare them to real world behaviors due to our general familiarity with men, pipes, deerstalker hats, crime scene investigations, and the like. Others appeal to maps to illustrate their theories. For example, Contessa (2007) presents his account of representation in terms of an interpretive relationship between elements on a map and the transit network linking different stations in the London Underground.

In each of these cases, the target system can be thoroughly described using common, model-independent concepts such as the *bob* of a pendulum, the *string* attached to it, the *fixture* holding the string. Concepts of this sort can be applied in a wide range of everyday contexts in which their meaning is not determined by any well-defined scientific theory. For this reason, examples of this kind are deeply misleading. Their strength in supporting similarity or morphism accounts of representation comes from the fact that, in these cases, the model targets are ordinary macroscopic objects—objects that are presented to human perception in forms that are easily understood, divisible into sub-components, and ascribed recognizable properties by means of common concepts. In such cases the task of relating these properties to a model appears straightforward. Reliance on examples like these, in conjunction with the component-wise mapping relation, is what has led to the implicit assumption of an independently described target system in much of the modeling literature.

But what of the multitude of scientific models of systems that elude ordinary perceptual capacities and common concepts: models of nuclei, ecosystems, proteins, galaxies, atomic lattices, neural networks, black holes, cognitive modules, and so on? There is no simple analogy between the means of forming descriptions of these systems and descriptions of everyday human-scale macroscopic objects. The latter do not routinely draw on concepts whose meanings are tightly constrained by scientific theories, while the former do so necessarily.[7] Moreover, there are good reasons to doubt that concepts suited to our everyday macroscopic engagement with the world continue to successfully refer to physical features when transported into the entirely different contexts of highly contrived experimental set-ups, including contexts of radically different spatial, temporal, and energy scales. Stebbings (1937), for instance, argued in her commentary on Eddington's table that our everyday notion of "emptiness" fails to retain the same meaning with respect to objects' constitution at the quantum scale. Describing a table as "mostly empty space" is thus, for her, an application of the concept outside of its proper domain. Similarly, our everyday notion of "hardness" breaks down with respect to objects and materials that we cannot probe through ordinary tactile means and splinters into disjoint measurement procedures (Wilson, 2006, p. 338). In such cases, our common concepts do not provide an adequate tool for describing the target system, and scientists lack a rich description of the target prior to model-based inquiry.

If we grant that this is the case in many areas of science, it raises a threat of circularity for ontic-first accounts of model-target comparison. If models are the primary means of characterizing a target, then to what are these models being compared? On what grounds can it be asserted that models are similar or homomorphic to their real-world targets when these models are themselves an important means of characterizing their targets? [8] Lacking such grounds, philosophers owe a satisfactory account of how

[7] Some theoretical models of the former kind may be fruitfully constructed or considered by analogy to macroscopic objects, as Hesse (1963) illustrates with respect to waves and billiard balls, but this does not grant the modeler any more direct access to the features of the target system itself.

[8] Van Fraassen (2008) details how this conceptual problem has arisen repeatedly in 20th century philosophy. With respect to a Putnam-Lewis debate, he writes, "As long as we are not given an independent description of both the domain and range of an interpretation, we do not *have* any such interpretation, nor any way to identify one […] we can grasp an interpretation—i.e. function linking words to parts of THE WORLD—*only if we can identify and describe that function*. But we cannot do *that* unless we can independently describe THE WORLD" (pp. 234-235).

surrogative reasoning works in these cases. This is especially pressing because in such cases practicing scientists have no clear way to apply the accuracy criteria proposed by an ontic-first approach, which requires that a particular model-target relation obtains. These relations appear to be epistemically inaccessible to practicing scientists and yet they are routinely able to reason about targets and evaluate their models in response to empirical results in these circumstances. This conceptual problem confronting mapping accounts is what I call the Paradox of Scientific Representation.

1.2.3. Attempts to address the Paradox

In light of this problem, two strategies for specifying the notion of comparison suggest themselves. Either,

(i) **Posit:** the target description is a hypothetical posit and the model-target relation is justified abductively (or not) through predictive success or failure, or

(ii) **Data-mediation:** the representation relation is checked by comparison to an empirical result, which either provides the best independent characterization of its target or is the basis for any further model-target comparison.[9]

I'll consider each in turn.

Frigg remarks on a disanalogy between representing North London with a street map and representing natural systems with a scientific model:

> Science is not like this; we do not first survey the hydrogen atom and then construct a model to communicate the findings to those not yet familiar with it […] We stipulate that we expect the model to bear this or that relation to its target, and then evaluate this claim against the best available background knowledge and by subjecting it to test using the usual methods of scientific investigation (Frigg, 2010b, p. 129).

[9] By an empirical result, I refer to the output of a processing sequence by which the data initially recorded from some target-apparatus interaction is given an interpretation or structure. I use the term interchangeably with data structure and data pattern in this paper.

Bueno and Colyvan (2011) address a similar case where there is no prior theoretical description of a target's structure. "This is clearly a problem for the mapping account […] When there is no such structure, we might impose some suitable structure or other and let the resulting mathematical model help us fine tune or revise the starting structure" (p. 347). These are appeals to **Posit**. Representation here amounts to a relation between a model and a (possibly more detailed) copy of it stipulated to be in the world.

Why believe that this act of stipulation tells us anything about the actual phenomenon and how our models actually relate to it? The most salient reason is that this is necessary to explain cases of successful scientific representation. That is, the success of a given model in enabling predictions or other inferences about their target is a reason to believe that there is a natural system in the world with features that stand in the appropriate similarity or structural mapping relation to those of the model. The proponent of this view may argue that there is no better way to account for successful scientific practices of drawing inferences from models to observations. Bueno et al. insist that inferences licensed by models

> are not themselves primitive or ungrounded but ride on the back of the relevant relationships between structures, relationships that can then be captured by partial isomorphisms, homomorphisms or whatever […] Suppose the given structures did not possess any of the relevant features that would license a move from one to the other; then obviously no such move could be justified (Bueno, French, & Ladyman, 2012, p. 45).

The target description is simply assumed and the fact that a model then successfully predicts an empirical result (however it might do so) licenses the abductive inference that a specific relation exists between the model and its target. As in Section 2.1, this is simply the familiar "no miracles argument" transplanted to the topic of representation. Like their analogues in the realism debate, these claims fall prey to conventional criticisms.

One such criticism is a variant of the underdetermination argument. Suppose a given model is empirically successful in that it allows users to reliably draw predictive inferences about the behavior of a target system. In other words, the model generates predictions about empirical results that are then satisfied, within some reasonable degree of error, by subsequent measurements or experimental procedures. I agree that this success calls for explanation, but it remains unclear precisely what comparison or mapping relation between model and target would enable this, and how this is meant to warrant belief in features of a target

beyond what is accessible through empirical procedures. The actual procedures that are used to test claims generated by scientific models often appear strikingly different from a simple similarity-based or morphism-based comparison between one thing and another. The series of magnetic relaxation rates and chemical shift data output by an NMR machine has no straightforward mapping relation to the structure of a macromolecule as suggested by comparative notions of similarity or morphism. This data must undergo a series of cuts, corrections, and transformations before it is usefully understood in terms of a structural model. That this data is eventually related to a model in a fruitful way *could* be attributed to the existence of a relation between the model and the target system underlying this data, but the kinds of procedures used to generate and process data do not provide definitive grounds for stating that this stipulated relation is one of similarity rather than structural morphism, or something else altogether.

The stipulated character of this model-target relation and its obscure connection to the actual empirical procedures connecting models to the world thus underdetermine accounts of *the* nature of this relation—whether it consists in model-target similarities, structural mappings, etc.— and corresponding accounts of the constitution of the target system. This being the case, efforts to justify representation relations in terms of stipulation and IBE can be accused of begging the question against an instrumentalist conception of modeling (Laudan, 1981). If the nature of the model-target relation cannot be determined from the empirical procedures that successfully relate models to the world, then it is plausible to think that these successes simply result from a more deflated interpretation of model contents. For example, one could think of model features as mere guides to anticipating certain data, measurement outcomes, or other empirical results. To claim that successful modeling *must* instantiate a deeper representation relation of one kind or another, and to infer the relation from these successes despite this alternative, is not particularly compelling unless it can be shown what this relation is adding to mere empirical success. If the relevant test of a representation relation is merely whether the model can predict empirical results, then the stipulation that the model is "really like" its target (i.e., similar or isomorphic to it) is idle—all the real work in understanding the model-target relation is done here by the claim that the model can successfully predict the relevant results.

Collectively these views function as the representation literature's analogue to the truth-as-correspondence relation between theoretical statements and states of affairs, which traditional scientific realism draws on to ground the reference of theoretical terms. They depend on analogous realist intuitions and are subject to analogous critiques. Lacking clear traction in the actual scientific practices of relating models to the world, these authors turn toward further speculative accounts of the relation between models and their targets decontextualized from the contexts of models' use. Accordingly, an area of the literature that attempts to account for representational capacities within this framework is now referred to as "the metaphysics of models," in which unified theories of the ontological status of models are debated in terms akin to the metaphysics of mathematical objects.

As Sandra Mitchell (2003) has argued, there is a risk to reifying abstract and idealized scientific representations, since these by nature only provide partial and piecemeal descriptions of the complex structures of natural systems. In such cases the kind of "mirroring" relation put forth by correspondence-based views can lead us down the wrong analytical paths. James Woodward provides a relevant diagnosis of what he calls "representationalist" approaches to the philosophy of science, which hold that

> the truth or accuracy of the representation requires that certain objects exist, where this is often accompanied by the assumption that these can be read off in a rather straightforward way from the representation itself […] [This] tend to lead to metaphysics when applied to discourse and practices whose corresponding objects are not obvious (Woodward, 2015, pp. 12-13).

Putnam (1981) likewise recounted the difficulties that theories of reference based on similitude or isomorphism have run into historically, and we can construct equivalent arguments against the representation relation between model and target conceived of as a relation of similarity or homomorphism. Against such views, Putnam recalls Berkeley's claim against Descartes and Locke was that it was senseless to think (as they did) that concepts of primary qualities stand in a direct correspondence with physical properties. We only make determinations of physical properties based on our sensations and resulting mental images of physical things, Berkeley reasoned, and these mental images do not themselves have properties that can be directly compared to physical properties. As Putnam puts it, "Mental images do not have a *physical* length. They cannot be compared with the standard measuring rod in Paris […] *Nothing*

can be similar to a sensation or image except another sensation or image" (Putnam, 1981, p. 59)—a critique echoing Neurath's criticism of correspondence: "*Statements are compared with statements*, certainly not some 'reality', nor with 'things'" (Neurath, 1931/1983, p. 53).

Analogously, we should be suspicious of loose formulations suggesting that researchers have access to theory- and model-independent means of describing the targets of models such that that one can speak of model construction and evaluation in terms of "comparing" or "coordinating" the properties of a model with its target. Just as it is obscure to claim that a mental image is similar to a physical thing, so it is obscure to claim that mathematical model is similar to a physical system. If, on closer analysis, we find we can know the features or structure of many scientific targets only by means of the models used to represent them, then it makes little sense to appeal to activities of direct model-to-target comparison when giving an account of the empirical success of models. In these contexts, such activities would amount to comparing the model to itself, or perhaps a more detailed version of the model. But a chief goal of an account of scientific representation is to explain how models allow researchers to draw inferences about a target phenomenon *in the world*. Component-wise mapping views and their elaboration in terms of a stipulated representation relation do not appear capable of doing this satisfactorily.

One might respond that the forms of relation emphasized by component-wise mapping views are meant to describe a relation between a theoretical model and a data model, rather than a direct relation between model and target. In a footnote, Chakravartty takes it "as given in discussing relations between scientific representations and 'the world'" that authors are referring to the relation between representational models and models of data (2010, p. 203). This is an appeal to **Data-mediation**. This provides a criterion for accuracy that is amenable to scientific practice: a model user can typically check whether the appropriate relation holds between their model and data. Most authors defending semantic views attest to the mediating role of data (Cf. (Bueno, 1997; French & Ladyman, 1999; van Fraassen, 2008).[10] Weisberg also notes that many mathematical models "are compared to mathematical representations of targets […] very similar to

[10] This goes back at least to Suppes (1962).

what Suppes called a *model of data*" (2013, p. 95). Yet at the same time, most authors take models to represent something *more* than mere data.[11] From this initial comparison between model and data, then, one might think scientists can infer the accuracy of the model via the model-target relation. But here too there are difficulties, as there exist a diversity of techniques for comparing models to empirical results. It is not clear these are reducible to a single form of relation, such as similarity or homomorphism. Nor is it clear that the existence of such a relation between model and data entails further claims about the relation between a model and its target phenomenon. I'll consider an example of each relation holding between model results and a data pattern. In each case, this relation between model and data does not warrant inference to a specific relation between model and target.

Goodness of fit and similarity: Consider the likelihood ratio, a common measure of model-data fit used in fields ranging from particle physics to social psychology. This is a statistical measure designed to compare results derived from experimental data and expectations derived from models. It relates observed measurement values to a distribution of expected values conditional on a given modeling hypothesis H_0. We can write this as (L_{H0}/L_{best}), where H_0 is the hypothesis under test.[12] This ratio compares the goodness of fit of recorded data to two statistical models. One, L_{H0}, fits the data with a model based on hypothesis H_0; the other uses the same base model but parameterized to maximize fit. So the full ratio measures how well H_0 can account for data, returning values around 1 when the fit is optimal and 0 when it fails completely.

This provides a simple comparison between a model and data, and there is indeed a sense in which this measure informs us about similarities, but these are only similarities between two distributions of data, which are not necessarily informative about the relation between model and target. One could, for example,

[11] Frigg (2010a) critiques Van Fraassen's view on this point. Weisberg claims that "Targets and their mathematical representations are not the same as empirically collected data" (2013, p. 96). Nguyen writes, "If it is phenomena that are ultimately represented by our scientific models, then it is those that are accurately or inaccurately represented." (2016, p. 176).

[12] More precisely, it is given by $-2\ln(L_{H0}/L_{\text{best}})$. The nature of the statistical models represented by the Ls may vary widely across disciplines. In particle physics, for example, each model is often given by a product of Poisson distributions of the form $P(x|y) = (y^x e^{-y}/x!)$, where the variables are counts of events inferred from recorded data or generated using a theoretical model.

use a simple quantum mechanical model to simulate the frequency, over a large number of trials, with which a spin-½ particle would be found in an "up" state when acted on by a rotation operator restricted to a single axis. Using this model for H_0, the likelihood test would show the expected distribution of states works well to predict the number of "heads" observed in an equivalent number of fair coin tosses. In a very precise and limited sense, the model system and the coin tosses behave similarly; in the long run both generate similar distributions. But this statistical similarity should not warrant an inference to a deeper similarity between the components of a quantum mechanical model and a fair coin. Given what we know about the model and target in this example, we should resist the claim that anything beyond the recorded frequency of 'H's and 'T's has been accounted for by this comparison; no deeper aspect of a fair coin's make-up has been revealed that is *really like* an observable acting on a mixed spin state, and there is reason to object to this inference: one problem with claiming that a standard coin toss has a similar structure to a quantum spin model is that the model must be arbitrarily restricted to generate the right distribution. Excess structure in the model, corresponding to other dimensions in which its states can be rotated, has empirical content that was left out with no principled justification other than to force a "match" with the statistics of a coin toss. Thus, a similarity between a model derivation and an empirical result, as determined by goodness of fit, is insufficient grounds for claiming that the model is accurate with respect to its target, or for explaining success on this basis.

Test responses and homomorphism: The same issues arise when considering a relation like homomorphism. Taking an example from Borsboom (2005), we could construct a four-item test for agreement with the following statements:

1) I have biological parents (yes: 1; no: 0)
2) I am over the age of seven (yes: 1; no: 0)
3) I menstruate (yes: 1; no: 0)
4) I have recently given birth (yes: 1; no: 0)

The test responses from a sufficiently large group of people would result in a triangular data structure, where all subject responses appear as one of the following lines:

Table 1: Data structure for the test

Item 1	Item 2	Item 3	Item 4	Sum score
1	0	0	0	1
1	1	0	0	2
1	1	1	0	3
1	1	1	1	4

This is the same form of data predicted by a Guttman model, by which the strength of a single latent psychological variable (say, distrust of government) is measurable in terms of a series of threshold prompts. The observed data is homomorphic to the measurement scale resulting from a Guttman model. Still, there is no compelling reason to think this data measures a latent variable underlying test subjects' answers. This nesting is not sufficient to think the progression from people with sum score 1 (humans under seven) to those with score 2 (males and non-menstruating females over seven), and so on to new female parents corresponds to a gradually increasing measurand within these populations. The nature of the questions simply forces the triangular nesting structure into the data. A structural relation such as homomorphism between a model-based prediction and an empirical result is thus insufficient by itself to ground judgments about the accuracy of this model with respect to the target generating this data. As Borsboom summarizes, "What matters is not just the structure of the data, but also the question of how these data originated. The relations in the data must not only exist, they must exist because of the right reasons" (2005, p. 105). In other words, there must be good reasons to believe the empirical results are relevant to the model being applied. First and foremost, these results must be able to serve as reliable basis for characterizing the target phenomenon. This requires the data be generated and processed through a set of procedures specifically designed for this purpose.

To summarize: ontic-first approaches produce criteria for the accuracy of models that are not directly applicable to many cases of scientific practice. They claim that surrogative reasoning with a model depends on an assumed relation between model and target, and that a model is accurate if and only if this

relation obtains. It is suggested that this criterion is applied through a comparison of model and target features, where this is understood to verify whether the appropriate relation between model and target obtains. Prima facie, this fails to account for typical cases where a characterization of a target's features is only given by the model under consideration. In such cases, a model's ability to successfully predict an empirical result does not by itself warrant the conclusion that a specific relation obtains between the model and its target. This remains true even if the appropriate relation exists between a model and a data structure. One may still posit that the model successfully predicts results in virtue of the proper relation existing between model and target, though this is best viewed as a metaphysical reconstruction that does not reflect the standpoint of scientific practice. The explanations of surrogative reasoning and representational accuracy provided by these accounts are unsatisfactory.

1.3. Semantic Imperialism

1.3.1. Semantic Imperialism in the literature

Another way that this picture misleads us becomes apparent when we seek to identify the representational content of a scientific model. The confusion stems, in part, from the residual influence of semantic theories in the philosophy of language and mind on the way philosophers pose and investigate questions of scientific representation. In particular, the notion of a component-wise relation between a representational vehicle and its target is often cashed out in terms that implicitly or explicitly borrow from denotational or truth-conditional semantics. I will illustrate this with a pair of examples.

R. I. G. Hughes (1996), in his influential DDI account, borrows directly from the work of Goodman (1968). He does so to claim that scientific representation, like artistic representation, should be understood in terms of denotation. Denotation, according to Goodman, is a relation of one thing "standing for" another that does not depend on resemblance. In thinking of this relation, Goodman leans on the philosophy of language of his time, inspired by the fact that linguistic symbols can be used to label and represent objects

without resembling them. For Goodman, the relation between a picture and what it represents "is thus assimilated to the relation between a predicate and what it applies to" (1968, p. 5). "Representations," he summarizes, "are pictures that function in somewhat the same way as descriptions," where this use of 'description' "covers all predicates from proper names to purple passages." (1968, p. 30) and includes a distinction between representing-as and representing which mirrors the sense/reference distinction of old. Building on this, Hughes conceives of scientific representation as the construction of a denotative labeling of parts of a model based on the partitioning of some target system, with interpretation as the "inverse" of denotation.[13]

For another example, consider Chakravartty's (2010) defense of "informational" views of representation, which posit a substantive similarity relation between model and target. He argues that we can understand how similarity is necessary for reasoning with models by appeal to "the idea of grasping the semantic content of a linguistic expression" and recognizing "how the semantic *content* of a (successful) linguistic description bears some similarity to its target, even though the superficial means by which that content is expressed (likely) does not" (p. 202, his emphasis). Chakravartty suggests we "analyze the idea of grasping content in terms of acquiring some sort of familiarity with the abstract models with which linguistic descriptions are associated" and in turn accept that "the content of linguistic representations in the sciences is informative with respect to their targets, precisely because it bears specifiable relations of similarity to those targets" (p. 202). This argument turns explicitly on the claim that representational content in scientific modeling should be treated on a par with its truth-conditional exposition in other areas of philosophy.

Generally, views on the nature of the representational content of linguistic utterances or intentional states attempt to clarify the relation between 'what is said/thought' and 'what is thereby spoken of/thought

[13] Suárez (2015b) lays out a deflationary version of the account in terms of *purported* denotation and mapping between *claims* about a model and target, rather than their respective properties. He also claims that denotation in DDI does not involve correspondences between partitioned models and targets, which instead comes at the interpretation stage. It is hard to know exactly what Hughes thought, given the brevity of his presentation, but there is some evidence that this is not the case. Consider the claim that "Just as a vertical line in one of Galileo's diagrams denotes a time interval, elements of a scientific model denote elements of its subject" (Hughes, 1996, p. S330).

about'. One widely influential approach proceeds by giving an account of how singular terms stand for objects and predicates stand for properties, typically by specifying an appropriate reference relation.[14] From here, one can describe how the syntactic structure of an utterance or thought is parsed in order to determine its semantic content, regarded as either (i) the set of conditions in the world under which the utterance is true or (ii) a description of what the world would be like if an intentional state were true. The basic idea is that the relations between terms and predicates that comprise 'what is said' (or relations between the concepts comprising 'what is thought') correspond to those between the worldly objects and properties to which they refer. And in this way the content of utterances or thoughts determines their truth conditions.

The ability to "grasp" the content of linguistic expressions or conceptually structured thoughts is thus elaborated in terms of sensitivity to their truth conditions, either in the behavioral responses of a speaker/thinker to their environment or in their normative responses to relevant use cases. Thus someone who properly grasps the concept "square" will reliably apply this predicate to square things in the world. On the dominant view, there is an invariant core to the content of such expressions that is the target of this "grasping," and which directs its proper usage. When competently exercised, grasp of these contents allows for more sophisticated compositional uses of language and concepts, as when I speak or think of a red square. In this way, the notion of representational content is taken to explain basic human capacities, such as the ability to construct novel thoughts and sentences and the ability to communicate with others.

Several aspects of the framework developed in language and mind have been imported into the modeling literature: For one, the notion that there is a correspondence relation—whether based in denotation, similarity, or mathematical transformation—between the components of a model and their relations, on one hand, and those of a target system, on the other. Further, the notion that proper familiarity with a model involves understanding what it purports to show about the world, i.e., what the world would be like were the model true or accurate. And so the notion that the truth or accuracy of a model is primarily evaluated in terms of this correspondence between components. We thus have sufficient elements in place

[14] In a theory of mind, these linguistic notions are assimilated to concepts.

for a parallel notion of representational content for a scientific model, understood as those components of the model and relations between them that provide a determination of the way the world is or may be. Such a notion of representational content can be extracted from (for example) Weisberg's (2013) claim that modelers place significant weight on matching some features between a model and its target, and de-emphasize the need to match others, or Bueno and French's (2011) appeal to positive, negative, and neutral structures in their account of partial homomorphisms between model and target. A more explicit account of content along these lines is found in Pincock (2011b), who writes that representation claims consist of a set of propositions describing how model parts correspond to physical features of a target system. "Taken together with the mathematical entity, these propositions impose a vast array of conditions on the system. These conditions are the content of the model," and so "we will say that a model is accurate with respect to aspect A of the system when its content concerning A is correct" (p. 22).

In all three cases, authors claim that there is some set of core features that, when paired with a relational means of mapping or interpreting, does the representational "work" in accounting for its successful applications, whereas other features are merely instrumentally useful. But this imported notion of representational content suffers from a range of disanalogies and related issues when applied to scientific modeling practice, which I detail here.

1.3.2. Problems for semantic views of representational content for scientific models

First, a discursive point: even if one were to draw inspiration from treatments of representation in the philosophy of language or mind, doing it along these lines is not a foregone conclusion. The clearest parallels for this approach to representational content are found in views that are often called denotationalism or truth-conditional semantics in language, and the representational theory of mind or encoding-based approaches to cognitive neuroscience. In language, these approaches have come under criticism from views such as inferential role semantics (Brandom, 1994) or varieties of contextualism

(Bach, 2005; Recanati, 2010); in theories of mind, they are criticized from the perspective of varieties of enactive and interactive views (Cussins, 1992; Bickhard, 1999). While there is a sense in which this approach to content is a mainstream view, it is misleading to argue for a particular view of scientific representation by appeal to this framework without acknowledging that one is taking up a contested position.

Another discursive point: insofar as the desiderata for a theory of scientific representation are not perfectly analogous to those motivating theorists of content in language and mind, it is unclear which elements it makes sense to import and which it does not. One purported virtue of the independence of semantic or conceptual content from things like context, expressive force, or propositional attitudes is the way this allows the theorist to point to a general, abstract, subject-neutral core within language or thought, which is postulated to serves as an aid to communication between individuals. It is perfectly plausible that scientific models are aids to certain forms of communication between scientists, but this communication is much less direct and much more socially- and institutionally-mediated than bare utterances, so the idea that these models must contain some general, context-independent content is not necessitated by issues of communicability. Similarly, the elements of models do not necessarily display the kind of compositional nature that is often held to apply to semantic content. Concepts are often assumed to possess a stable referential core, whereas it is not obvious what a quantity like Newtonian force represents independent of its application to particular kinds of systems—i.e,, when paired with a law of gravitation, or Hooke's law, or Couloumb's law, etc. Therefore, we should not expect "content" to work in an analogous way in scientific modeling as it does in language. We should be wary of importing too much of the dominant semantic view into thinking about models.

This divergence between models and language can be pushed much further. [15] In particular, the representational content of scientific models is not "transparent" in the same way that a perceptual or imagined experience is, or in the way that the meaning of complex expression is readily understood by a

[15] For the sake of this discussion, I am concerned first and foremost with mathematical and computational models.

competent speaker. That is, the arrangement of worldly objects and properties that would make the model accurate cannot be simply "read off" a model. Some of this has to do with widely recognized features of scientific models. For one, models are abstract, which means not everything in their target phenomenon is represented by them, and it is not always obviously which aspects can be represented and which cannot. Second, models are idealized, meaning not everything in the model is taken to represent something in the target. As a result, when presented with some model, it is an open question which features of it represent which features of its target (and how well) and which do not.

This initial opacity is familiar to anyone who has struggled to get a tractable result from a model. But the issue is exacerbated by other characteristics of modeling in science, all of which put pressure on the claim that the content of a model can be read off from a component-wise mapping relation.[16] First, all scientific models have a restricted domain of application, beyond which they are subject to various forms of breakdown. Continuum models, such as Fourier's heat equation or the Euler-Bernoulli beam equation, have calculable behaviors at any spatial scale. Given the latter equation alone, one could calculate the flexibility of a material at the 10^{-50} m scale. Present-day physicists recognize that it would be a mistake to treat this as a proper application of the model, but there is nothing in the model itself that specifies such a thing. This is different from the standard treatment of a concept like "red," which, if properly grasped, a subject knows how to apply across a wide range of contexts. Unlike these concepts, the rules governing the application of a model are not part of their intrinsic content.

Second, there are mathematical operations that are physical nonsense when performed in some modeling contexts, yet are meaningful in others. For instance, dividing a physical quantity by zero does not generally return a physically meaningful value, yet there are points in the dynamical models of physical systems where these function as singularities that coincide with an important change in the state of the system, such as phase transitions in materials. The question whether singularities are artifacts of a particular

[16] The first two points here are reformulations of remarks made in a talk by Erik Curiel (2021).

mathematical formulation or have physical significance is a non-trivial one, a matter that bedeviled the early history of general relativity, for instance.

Third, there are many different mathematical formulations of the same physical theory. One often finds that there are models with radically different structural elements that are equally capable of modeling a given phenomenon. Consider, for instance, the diverse range of mathematical structures that can be used to implement a theory. Curiel (2021) presents over twenty mathematical models for general relativity, ranging from manifold-and-tensor models, to fiber bundle models, to Einstein algebras. Many of these do an equally good job of modeling established systems of study in general relativity, yet are very different in their structural details. This raises some basic questions for the view that their adequacy is due to component-wise similarities, structural mappings, or the like. Is the form of model-target similarity preserved under some similarity mapping between models? Is there a common structure to all of these? Is one representation more fundamental, and on what basis? Even if there is a way of affirmatively answering these questions in some cases, one must admit that the mathematics are far from transparent with respect to the organization of properties of things in the world.

Even for a well-understood and long celebrated mathematical model, it may be exceedingly difficult to determine which components are doing the representational work and which are dispensable. Rice (2019), for instance, argues that many models depend on a set of idealizing assumptions to be mathematically tractable, as with the series of assumptions that impose a Maxwell-Boltzmann distribution on the velocity of particles in an ideal gas. Removing any of these "pervasive distortions" would undermine the mathematical framework that makes the model applicable in the first place. Thus, according to Rice, there is no good way to decompose the model piece by piece to determine which truly hold of their target. Problems with determining the representational contribution of model components are likewise illustrated by Kaveh's (2021) analysis of the Bohr model. Bohr's model was and is regarded as empirically accurate with respect to single-electron system such as hydrogen, ionized helium, and twice-ionized lithium. And yet it was a clear failure with respect to neutral helium. The central puzzle in this case is that the best candidate for the component of Bohr's models that might represent or refer to a feature of single-electron

atoms and thereby account for the empirical success—the principal quantum number *n* used in assigning energy states to orbital electrons—is no different in the neutral helium model. Likewise, Kaveh provides historical evidence that this difference in success and reception is not attributable to a neglect to include particular features of the helium system, such as electron-electron interactions or additional quantum numbers such as those due to angular momentum or spin. The single-electron models neglected the same things and were subject to the same kind of empirical inaccuracies as a result.[17] Thus, component-wise similarities and dissimilarities, or some comparable form of mapping, cannot explain why the models were successful in single-electron cases, but not for neutral helium.

One strategic response to issues such as these is a turn to fictionalism. That is, one might argue that models are, in the first instance, tools for reasoning about fictional systems. The claims one makes on the basis of a model are not strictly speaking about its real-world target, but rather a hypothetical, fictionalized referent; modelers represent a phenomenon P "as if" it were a phenomenon P*. Translating these claims to ones about real-world targets is then a second inferential step that takes the form of a counterfactual: "were the target like P*, then it would it behave in this way." Insofar as we take this second step to be part of what it is for a model to represent a worldly phenomenon, and insofar as this step is meant to account for the effective use of the model, this strategy merely pushes back the question of representational content. Presumably some claims about this fictional system are warranted with respect to a target phenomenon (i.e., they render the counterfactual antecedent true), and others are not. One is then obliged to specify what features in the hypothetical system (which largely replicates the structure of the model) correspond to this target and which do not. But this brings up the same issues with reading off the content of a model that were raised in the previous paragraphs. Moreover, this view creates additional semantic quandaries. One might think that fictionalism is a fitting response to the fact that some models involve impossible assumptions. But if the features of these fictional systems are meant to ground counterfactual claims about

[17] It was known, for example, that all of these phenomena produced spectral line patterns that were not accounted for by Bohr's model (corresponding to what is now called fine structure), and that even in the case of hydrogen the model was somehow lumping together empirically distinguishable states, yet as Kragh (1985) tells it this was not taken as fatal for the model.

a target, they then involve conditionals with impossible antecedents, and such claims have been criticized as vacuous (Williamson, 2018). Conceptual problems of this sort arise from the assumption that the use of scientific models as representations involves a commitment to the existence of a corresponding system, each state of which is mapped from those of a model—including the states recognized as scientifically impossible.

Despite all of this, the fact remains that scientists are able to draw informative inferences about target phenomena from models. Theories of scientific representation are right to treat this as a key explanandum. It is valuable to consider the role of representational content here because it makes explicit the need to specify what it is that does the "work" in grounding model-to-target inferences, especially in light of the issues raised in recent decades with respect to abstraction, idealization, and their ilk. It is increasingly apparent, however, that received theories of representation are unlikely to provide this explanation. The notions of content they entail compel us to seek out the parts of models that are doing the representational work (alongside a prescribed relation), but we have good reason to think that many models in science are simply too opaque for this task to prove fruitful. At this, we may just throw our hands up and declare it a mystery that models are such effective reasoning tools. Or we might question the assumption that whatever does the work is located strictly *within* the model, or describable in terms that are internal to the model-target relation. This assumption begins to give way with references to users' purposes and their research context, but these concepts tend to remain unanalyzed.

1.4. Inferentialist accounts of scientific representation

Another strain in the scientific representation literature makes room for the agents who employ models. Suárez (2003) has criticized variants of the similarity and morphism views by pointing out that they fail to satisfy the irreflexivity and asymmetry of representation relations, that they define representation in such a way that they cannot account for the phenomenon of *mis*representation or representation of non-existent objects (defining away all such cases as non-representation), and that they neglect the possibility

of a variety of styles of representation. Authors have responded to these critiques by including the role of an agent who interprets model features as denoting worldly features (thereby introducing asymmetry) and determines which model features and worldly features are relevant for comparison (thereby taking care of overly-inclusive notions of similarity and morphism). Giere (2004) employs the more expansive "S uses X to represent W for purposes P." Likewise, van Fraassen (2008) uses "A represents B as C" and makes explicit appeal to the role of purposes.

While some of these approaches bring in model users as a supplement to substantive accounts of the representation relation, other user-based positions in the philosophical literature have developed out of inferentialism. This is the view that the representational capacity of models should be explicated in terms of the way models allow their users to draw inferences about real-world systems. Suárez (2004), the leading exponent of an inferentialist account of scientific representation, coined the term surrogative reasoning for this inferential capacity. For Suárez, the fact that a competent and informed user can engage in surrogative reasoning is one of two minimally necessary conditions for something to serve as a scientific representation. The other condition is that the model has representational force, which is its capacity to "point to" a real-world system or, more precisely, "to lead a competent and informed user to a consideration of the target." This is not enough to make something a scientific representation because, as Suárez notes, such force can be established by stipulative convention. Representational force is thus a highly relational and context-dependent property determined entirely by the intentions of agents within a community. This is because there is no in-principle limit on the stipulation that one thing stands for another. The condition of representational force thereby places very few limits on what can count as a representation.

Scientific representations, Suárez argues, are different from representations established by stipulative fiat. They are objective. The objectivity of scientific representation must depend entirely on the surrogative reasoning requirement. For a representational vehicle to inform a user about its target, it must be appropriately structured such that an informed and competent agent could use it to gain this information. Suárez illustrates this with a simple example of pens and a piece of paper standing for a ship on the sea: while taking the pens to stand for ships and the paper for the sea would permit certain inferences to be

drawn about ships on the sea from the pen-paper system, this would be much more difficult if the paper were taken to represent the ships and the pens the sea. Yet Suárez refrains from specifying what this structural relationship should be in general. This is because he takes himself to be providing a deflationary account of scientific representation, one that does not purport to provide a definition of this concept in terms of necessary and sufficient condition, and which does not assume any specific form of representation relation. The two necessary conditions he provides are merely a "schema" to be filled in on a case-by-case basis, with the help of the skills and background knowledge of model users.

Suárez's focus on the practices that allow for surrogative reasoning is a valuable addition to the representation discourse. But there are aspects of his deflationary approach that are unsatisfying. On his deflationary account, surrogative reasoning is a crucial feature of representation that is left unexplained. It thus remains mysterious how (un)successful inferences are generated on his account. This is a consequence that Suárez wholly accepts, claiming that an explanation of surrogative reasoning can only be given on a case-by-case basis and that "nothing can explain why those singular properties (or relations), which are in fact used, are the appropriate ones for representation" (Suárez, 2015a, p. 4). Thus "representation is itself a redundant concept […] it expresses some honorific value that accrues to a particular use that agents make of some model" (p. 9). But this forfeits too much too quickly. Suárez has been appropriately criticized for providing insufficient argumentation against projects that seek to account for the objectivity of scientific representation in greater detail, i.e., those that seek to explain how models allow for effective surrogative reasoning. As Contessa (2007) points out, Suárez's argument that available accounts of scientific representation do not succeed at explaining scientific representation in terms of necessary and sufficient conditions does not warrant the conclusion that *no* explanation of scientific representation is possible.

Contessa attempts to flesh out a practical approach to inference-based modeling by giving a general theory of the kind of activity that enables model users to represent. Contessa presents this use in a theory of *interpretation*. For him, the representational capacity of models is accounted for by "the fact that the user interprets the model in terms of the system" (Contessa, 2007, p. 51). The theory of interpretation is meant to describe the activity that underlies (un)successful inference, and so explains surrogative reasoning.

Unfortunately, Contessa's approach reinflates surrogative reasoning with elements of a component-wise mapping, spawning familiar problems.

The centerpiece of his theory is a formal presentation of the activity of interpretation. While he takes care to note that this formalism may not strictly apply in every case, he also claims it can be found "[i]n the overwhelming majority of prototypical cases of epistemic representation." (p. 58). This process of interpretation involves identification of...

- A set of relevant objects in the vehicle and a set of relevant objects in the target
- A set of relevant properties and relations among objects in the vehicle and among objects in the target
- A set of relevant functions from an *n*-times Cartesian product of the vehicle objects to a set of *n*-ary functions on the vehicle
- A set of relevant functions from an *n*-times Cartesian product of the vehicle objects to a set of *n*-ary functions on the target

Which provide a proper interpretation if

- The user takes the vehicle to denote the target
- Each object in the vehicle is taken to denote a single object in the target and vice versa
- Similar denotation relations hold for properties and functions in the vehicle and target

This interpretation grounds inference rules by which a user can infer the existence of objects, relations, and values of functions in the target using their analogues in the model.

This solution is prima facie unworkable, given prior discussion, as it relies on the assumption of an independently described target system that I have previously criticized. Like other explanatory accounts of representation, Contessa appears to assume that a model-independent account of the components of the target system is available such that a "denotation" relation can be easily set up between these and the parts of a model. I do not believe we can simply assume this account exists, and so the theory of what "grounds" successful inference is based purely on the stipulation that the thing in the world's features mirror those of the model. As with prior authors, Contessa leans heavily on a case of representing a familiar macroscopic object to motivate his claims (in this case, drawing inferences about the London subway system from a map).

Authors who emphasize the role of agents in their accounts of representation may argue that this misconstrues their aims. It may be argued that these accounts do not seek to reify representation in terms of a relation between a model and its target, and so do not assume that an independently sourced description of the target is available. Rather, they are making a weaker assumption: for a model to represent a target the representing agent need only *assume* such a description of the target. This is an empirical claim about the psychological states of those engaged in surrogative reasoning. It is possible that this is in fact what some users of models do, but I do not see why practitioners *must* interpret their models realistically in order to reason with them. More instrumental interpretations of models are equally psychologically plausible: some scientists may reason effectively with models while viewing them as nothing more than tools for predicting some empirical result.

Beyond whatever mental state the modeler needs to be in, there is a question of the larger context in which this interpretation takes place. This context may provide background knowledge that is crucial for effective interpretations and so is relevant to determining what kind of relationship models need to have to the world in order to yield reliable inferences. If we want to understand the representational capacity of models through an examination of their actual use, then closer attention needs to be paid to the actual procedures by which models are brought into relation with the world. I do not believe these are adequately accounted for in terms of interpretation as presented by Contessa. On the contrary, we see in practice that models are never related in such a straightforward way to a partitioned real-world system, but are aligned with other models by means of localized practices. Unlike Suárez, I believe a general account of the way in which the local "background" knowledge accompanying measurement and experimentation connects models to the world is worth pursuing. This is not unprecedented—attempts at such a general account showing how experiment and modeling are related go back at least as far as Suppes (1962).

The inferential account can be better fleshed out when we understand representation as a relationship between a model, *an empirical program*, and the world. On the view I will put forward, models function as representations of natural phenomena only when integrated with an empirical program. The work of tying a theoretical model to such phenomena is not reducible to a simple cognitive act of

interpretation qua mapping—"interpretation" (if we want to use that term) involves integrating a model within a technical-procedural complex. For example, when properly aligned with an empirical program, the inferential rules internal to such models can enable practical inferences that license researchers to anticipate and identify a particular feature within their data and attribute this feature to a causal factor that mediates dependencies between more directly accessible factors within the experimental setup.

1.5. Prior data-based views: epistemic vs ontic features of the model-world relation

The suggestion that we examine the relationships between modeling and experiment is not foreign to the philosophy of science literature. One of the inaugural papers on modeling in science, Suppes (1962), deals with the construction of data models as part of a hierarchy of structures comprising an empirically tractable scientific theory. Other works (Giere, 1988; Suppe, 1989; French & Ladyman, 1999) propose a view of theory-world relations that is evidently inspired by Suppes's claims.

However, that these works are firmly rooted in the semantic view of theories and reflect the theory-oriented perspective of this tradition. In particular, they fail to carry out any detailed study of scientific procedures of data generation and processing in order to account for the manner in which a comparison between theoretical contents and data is achieved in practice. Instead, most treatments of the relationship between measurement procedures, data, and models are presented through toy examples (if any) that are readily adapted to formal treatment. Suppe (1989), for instance, purports to lay out the theory of experimental design for cloud chambers in terms of sets of propositions derived from physical theories of heat, electromagnetism, and other general theories, along with facts and additional information that allow for the derivation of special generalizations pertinent to the use of cloud chambers:

> How, then, do we obtain the theory of experimental design? In cases such as the cloud chamber observation, formulations $J_1, \ldots, J_n$ of a number of theories $T_1, \ldots, T_n$ are used together with statements F of facts and collateral information I to obtain the special generalization E covering the cloud chamber experiment. In order to do so, $J_1, \ldots, J_n$, F, and I must be compatible in the sense that they use the same terms to refer to any parameters common to several theories, the facts, or the collateral information. Let J^* the set of

> propositions from compatible J_1, ..., J_n, F, and I used to obtain E in the sort of derivation described earlier; such a derivation requires that F^* contain statements of laws from J_1, …, J_n. If T_1, …, T_n are empirically true, and if F and I are factually true, then it follows that the propositions in J^* are consistent; as such, J^* will have a model. In particular, J^* will have a model whose domain of discourse contains states whose defining parameters are all those parameters designated by parameter terms in the proposition J^* (p. 140).

This does not reflect scientific practice. Philosophers and historians of experiment have argued that the techniques and knowledge base involved in experimentation often develops and evolves independently of general theories, and that the design of experiments depends on a familiarity with such instrumental traditions (Hacking, Representing and Intervening, 1983; Galison, 1987). Hacking (1981) notes that an adequate theoretical explanation of optical microscopes was not established until centuries after their first scientific use. Bogen and Woodward (1988) draw on such views in arguing against the notion that general theories are ever meant to explain data obtained through measurement and experimentation. They point out that many techniques involved in reliably generating data in the lab do not require appeal to general theory in the way suggested by Suppe's notion of experimental design. In the case of bubble chambers, the electrostatic repulsion theory of bubble formation motivating Donald Glaser, the chambers' inventor, turned out to be incorrect, and yet the chambers were used to produce laboratory data for years before the currently accepted theory was developed (Galison, 1985).

Leonelli (2019) has recently indicated further ways that the formal approach following Suppes "fails to tackle critical questions around the source of the epistemic value of data, and the relation between data and models" (p. 7). She notes that the interpretation of data models in terms of statistically tractable models deals with only a subset of those objects scientists refer to as data, that it is hard to apply to exploratory research, and makes uncritical assumptions about what constitutes "raw data," noting that in the case of image-based data production there are often extensive stages of reduction and processing that occur before data are quantified and represented in ways that might correspond to Suppes's notion of data modeling. Tal (2021) has recently driven this point home with respect to a foundational component of traditional structuralist accounts of representation: the representational theory of measurement. These theories assume that it is possible to infer the qualitative structure of objects from empirical data

independently of any prior attribution of quantitative structure to these objects. This would allow for a straightforward test of whether some data satisfies the axioms for a given quantitative measurement structure. This ignores the fact that the measuring instruments used to generate data are often subject to nonlinear systematic errors, which can only be corrected by imposing a preconceived quantitative structure onto data. This undermines the possibility of an independent test.

I believe we need to re-examine the notion of model-target comparison with an eye to procedures involved in gathering the data that can be reliably connected to representational models. But there is a basic obstacle to pursuing an inquiry along these lines: Ontic Priority. Many philosophers of scientific representation claim that matters of data acquisition and processing are subordinate to an understanding of the representational relation between a model and its target. They may dismiss these as merely epistemic questions regarding how models are confirmed, the success or failure of which can be explained in terms of an underlying relation between model and target. But this approach is riddled with conceptual problems, as I have reviewed in Sections 3 and 4.

I believe there are distinct benefits to following the reverse course—that is, seeking to grasp the nature of the model-world relationship through a prior analysis of the way models are actually connected to the measurement-taking and experimental practices of working scientists. Proceeding in this manner mitigates the risk of imposing a framework affected by prejudices regarding philosophically loaded notions like "representation," including those I have ascribed to Semantic Imperialism. Nor does this approach reduce representation to questions of confirmation and evidence. At the very least, philosophers of science should want our epistemology and metaphysics to fit together. If we hold that there is an ontic relationship between model and target, but that this bears no relation to the patently epistemic activities of drawing inferences from models and connecting these to empirical results, then this ontic relation seems to make no connection at all to how people go about establishing an actual relationship between model and world. A conception of science that severs the tie between strictly epistemic confirmatory procedures and strictly ontic representation relations leaves us with an empty notion of model-world "comparisons"—a process that is either left completely undescribed, only described in highly abstract terms, or reduced to a bare

stipulation that leaves the use of the model in empirical inquiry mysterious. The criteria for representational accuracy proposed on this basis depend on the existence of a relation that is epistemically inaccessible to practicing scientists. The notion of representational content they favor is largely inscrutable. But if the inferential utility of models is to be understood in terms of the practices by which scientists connect them to their targets, then it is *required* that we study their relation to data generating and processing procedures. After all, this is the only way that the world becomes a subject for real comparisons.[18]

1.6. Summary of critiques and outline of an alternative approach

In sum, current theories of scientific representation have not satisfactorily accounted for surrogative reasoning or representational accuracy. That is, they do not provide an account of how scientific models allow their users to draw informative inferences about their targets, and they fail to explain what distinguishes empirically successful cases of modeling from those in which inferences are unreliable in terms that are adequate to scientific practice.[19] The explanatory inadequacy of these accounts is rooted in their appeal to a component-wise mapping between features of a model and features of a target system. Teller, Hughes, Bueno, French, Ladyman, Pincock, Godfrey-Smith, Frigg, Contessa, and sometimes Giere and Suárez present representation in these terms, some of them under the influence of Semantic Imperialism. These authors' use of illustrative examples based on models of macroscopic systems reinforces an implicit assumption that an independent description of target phenomena exists such that it makes sense to speak of an independently verifiable mapping between them and their model. This assumption does not reflect the epistemic situation of science, in which the best available descriptions of a

[18] I will address further objections to this approach from the standpoint of Ontic Priority in Chapter 6.
[19] Contessa (2007, 2013) acknowledges this difference in his distinction between epistemic representation and faithful epistemic representation, but his account of how a model achieves the latter is definitional. He thus presents a way to understand what empirically successful modeling consists in (and does so with a correspondence-based view), but he does not effectively explain how it occurs.

target system are dependent on models. Nevertheless, scientists can use and evaluate these models. A commitment to Ontic Priority deprives us of the ability to address this paradox.

The wholly deflationary attitude of Suárez, according to which no deeper explanation can be given for the representational capacity of models, strikes me as a preemptive sacrifice of our ability to understand the roots of surrogative reasoning. On the other hand, I sympathize with the general orientation of his work, according to which our account of scientific representation should have a tight conceptual link to the norms and techniques that guide actual modeling practices. But I disagree with Suárez's contention that there is no way to elaborate the workings of scientific representation beyond case-by-case studies of the different contexts in which models are employed. Instead, I will present an explanatory account of surrogative reasoning and representational accuracy

By forgoing the assumptions of both a model-independent target description and component-wise mappings, we are in a better position to recognize that models are not first and foremost vehicles for inferences about real-world targets, but about other models—both data models and related representations of shared target systems. This is the insight found in authors such as Suppes, Suppe, and van Fraassen, which I broadly adopt. We can approach an understanding of the representational utility of models by investigating the way models are inferentially related to practices of data gathering.

References

Bach, K. (2005). Context ex Machina. In Z. G. Szabo (Ed.), *Semantics Versus Pragmatics* (pp. 15-44). New York: Oxford University Press.

Bickhard, M. (1999). Interaction and Representation. *Theory & Psychology, 9*, 435-458.

Bogen, J., & Woodward, J. (1988). Saving the Phenomena. *The Philosophical Review, 97*, 303-352.

Boorsboom, D. (2005). *Measuring the Mind.* New York: Cambridge University Press.

Brandom, R. (1994). *Making It Explicit.* Cambridge, MA: Harvard University Press.

Bueno, O. (1997). Empirical Adequacy: A Partial Structures Approach. *Studies in the History and Philosophy of Science, 28*, 585-610.

Bueno, O., & Colyvan. (2011). An Inferential Conception of the Application of Mathematics. *Nous, 45*, 345-374.

Bueno, O., & French, S. (2011). How Theories Represent. *British Journal for the Philosophy of Science, 62*, 857-894.

Bueno, O., & French, S. (2018). *Applying Mathematics.* New York: Oxford University Press.

Bueno, O., French, S., & Ladyman, J. (2012). Models and structures: Phenomenological and partial. *Studies in History and Philosophy of Modern Physics, 43*, 43-46.

Callendar, C., & Cohen, J. (2006). There Is No Special Problem About Scientific Representation. *Theoria, 55*, 67-85.

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

Chakravartty, A. (2010). Informational versus functional accounts of scientific representation. *Synthese, 172*, 197-213.

Contessa, G. (2007). Scientific Representation, Interpretation, and Surrogative Reasoning. *Philosophy of Science, 74*, 46-68.

Curiel, E. (2021). *Math Does Not Represent.* Retrieved from Center for Philosophy of Science YouTube: https://www.youtube.com/watch?v=aA_T20HAzyY

Cussins, A. (1992). Content, embodiment, and objectivity: The theory of cognitive trails. *Mind, 101*, 651-688.

Elliot-Graves, A. (2020). What is a target system? *Biology & Philosophy, 35*, https://link.springer.com/article/10.1007/s10539-020-09745-3.

French, S., & Ladyman, J. (1999). Reinflating the semantic approach. *International Studies in the Philosophy of Science, 13*, 103-121.

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

Frigg, R. (2010b). Fiction and Scientific Representation. In M. C. Hunter (Ed.), *Beyond Mimesis and Convention* (pp. 97-138). New York: Springer.

Galison, P. (1985). How the first neutral-current experiments ended. *Reviews of Modern Physics, 55*, 477-509.

Galison, P. (1987). *How Experiments End.* Chicago: University of Chicago Press.

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

Giere, R. (2004). How Models Are Used to Represent Reality. *Philosophy of Science, 71*, 742-752.

Godfrey-Smith, P. (2006). The strategy of model-based science. *Biology and Philosophy, 21*, 725-740.

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

Goodman, N. (1968). *Languages of Art.* New York: Bobbs-Merrill.

Hacking, I. (1981). Do We See Through a Microscope? *Pacific Philosophy Quarterly, 62*, 305-322.

Hacking, I. (1983). *Representing and Intervening.* New York: Cambridge University Press.

Hesse, M. (1963). *Models and Analogies in Science.* New York: Sheed and Ward.

Hughes, R. I. (1996). Models and Representation. *Philosophy of Science, 64*, S325-S336.

Kaveh, S. (2021). *A Prescriptivist Account of Physical Theories.* Dissertation: University of Pittsburgh.

Laudan, L. (1981). A Confutation of Convergent Realism. *Philosophy of Science, 48*, 19-49.

Leonelli, S. (2019). What distinguishes data from models. *European Journal for Philosophy of Science, 9*, Access online, DOI: https://doi.org/10.1007/s13194-018-0246-0.

Lloyd, E. (1988). *The Structure and Confirmation of Evolutionary Biology.* Princeton: Princeton University Press.

Longino, H. (1990). *Science as Social Knowledge.* Princeton: Princeton University Press.

Mitchell, S. (2003). *Biological Complexity and Integrative Pluralism.* Princeton: Princeton University Press.

Morgan, M., & Morrison, M. (Eds.). (1999). *Models as Mediators: Perspectives on Natural and Social Science.* New York: Cambridge University Press.

Neurath, O. (1931/1983). Physicalism. In O. Neurath, R. S. Cohen, & M. Neurath (Eds.), *Philosophical Papers 1913-1946* (pp. 52-57). Dordrecht: Reidel.

Nguyen, J. (2016). On the Pragmatic Equivalence Between Representing Data and Phenomena. *83*, 171-191.

O'Conner, C., & Weatherall, J. (2016). Black Holes, Black-Scholes, and Prairie Voles. *Philosophy of Science, 83*, 613-626.

Pickering, A. (1992). From Science as Knowledge to Science as Practice. In A. Pickering (Ed.), *Science as Practice and Culture* (pp. 1-28). Chicago: University of Chicago Press.

Pincock, C. (2011a). *Mathematics and Scientific Representation.* New York: Oxford University Press.

Pincock, C. (2011b). Modeling reality. *Synthese, 180*, 19-32.

Putnam, H. (1981). *Reason, Truth, and History.* New York: Cambridge University Press.

Recanati, F. (2010). *Truth-Conditional Pragmatics.* New York: Oxford University Press.

Rice, C. (2019). Models Don't Decompose That Way: A Holistic View of Idealized Models. *British Journal for the Philosophy of Science, 70*, 179-208.

Stebbings, S. (1937). *Philosophy and the Physicists.* London: Methuen and Co.

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

Suárez, M. (2004). An Inferential Conception of Scientific Representation. *Philosophy of Science, 71*, 767-779.

Suárez, M. (2015a). Representation in Science. In P. Humphreys (Ed.), *The Oxford Handbook of Philosophy of Science* (pp. Accessed online, DOI: 10.1093/oxfordhb/9780199368815.013.25). New York: Oxford University Press.

Suárez, M. (2015b). Deflationary representation, inference, and practice. *Studies in History and Philosophy of Science, 49*, 36-47.

Suppe, F. (1989). *The Semantic Conception of Scientific Theories and Scientific Realism.* Chicago: University of Chicago Press.

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

Tal, E. (2021). Two Myths of Representational Measurement. *Perspectives on Science, 29*, 701-741.

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

van Fraassen, B. (1980). *The Scientific Image.* New York: Oxford University Press.

van Fraassen, B. (2008). *Scientific Representation: Paradoxes of Perspective.* New York: Oxford University Press.

Weisberg, M. (2013). *Simulation and Similarity: Using Models to Understand the World.* New York: Oxford University Press.

Williamson, T. (2018). Counterpossibles. *Topoi, 37*, 357-368.

Wilson, M. (2006). *Wandering Significance: An essay on conceptual behavior.* New York: Oxford University Press.

Woodward, J. (2015). Sketch of some themes for a pragmatic philosophy of science. *Draft*.

2.0. A New Use-Based Account: Justification and Constraints

Chapter 1 identified surrogative reasoning, or the use of models to draw informative inferences about a target system, as a basic feature of scientific representation. There I criticized the inferentialist view of representation for preemptively abandoning the project of giving an explanatory account of surrogative reasoning. I agree, however, with its claim that enabling surrogative reasoning is a minimal necessary condition for a model to serve as a representation, and I share the inferentialists' aim of understanding representation by connecting it to the norms and techniques that guide actual modeling practices. Here I embark from this point to further analyze the relationship between theoretical models and data gathering practices in science. I take this to be an instance of what Waters (2019) calls a "broad-practice-centered" philosophy of science, aimed at explaining how the scientific investigation of phenomena integrates a diverse range of activities.[20]

Other authors in the philosophy of science have studied representation with an explicit focus on matters of scientific practice. Andrea Woody's work exemplifies this perspective:

> If there is something *distinctive* about representation in science, my bet is firmly on it being located in the sophistication of the representational and inferential practices of the relevant communities. It is not the representations themselves, but how scientists develop, use, and build upon them that should draw our scrutiny (2014, p. 147).

Her account of representations of periodic laws in the 19th century shows how these offered different "resources" to the chemistry community in terms of the kind of inferences they could support. Remarks and philosophical studies such as these,[21] though suggestive, have not to my knowledge been worked into philosophical theories that aim to account for surrogative reasoning more generally.[22] The same can be said

[20] As Waters notes, this approach is more broadly aligned with proposals among historians and philosophers of science (notably (Longino, 1990) and (Pickering, 1992)) to conceive of science as an "activity," rather than mere "product" or body of theoretical knowledge.

[21] See also select essays in (Morgan & Morrison, 1999); (Knuuttila & Loettgers, 2011; 2013; 2016); (Perini, 2012); (Burnston, Sheredos, Abrahamson, & Bechtel, 2014); (Woody, 2004).

[22] Woody's work, for instance, is primarily based on connecting the use of representations to the long-standing topic of scientific explanation. Recent work by Boesch (2017a; 2017b) is more promising on this point, presenting a view according to which the communal "licensing" of scientific representations entails that "we cannot pull the question of the constitution of representation away from questions of practice" (2017a, p. 979)..

for writing on scientific practices of representation under the banner of "science and technology studies" (see, e.g., the edited volumes by Lynch and Woolgar (1990) and Coopmans et al. (2014)). These tend to sideline philosophical debate in favor of studies of a pluralistic range of representational techniques and methods. Authors cover a wide variety of detailed cases but stick to predominantly behavioristic characterizations of scientists' activities. Many studies prioritize descriptive accounts of the production and manipulation of visual imagery (as noted by Tibbetts (1990)), and do not aim to provide a fully clear account of the kinds of inferences made available by these.

While valuable in other respects, such studies often do not clarify model-based inferences at a level of philosophical generality. The central aim of this chapter is to follow the implications of a general claim: to be able to draw informative inferences about a model's target, the model must be related to an empirical procedure that generates data about that target. I defend this claim by analyzing the notion of informative inference and then lay out conditions that must obtain for a given model and empirical procedure to permit such inferences about a target phenomenon. On the modeling side, I claim this requires a grasp of the conditions of the model's applicability. For empirical procedures, this requires the data they generate be properly connected to the target phenomenon and allow for the extraction of patterns that can be connected to features of the relevant model. Without such data, researchers lack a means for connecting a given model to information about its target.

This chapter presents a general account of the practical underpinnings of scientific representation. Chapters 3 through 5 delve further into the role of data gathering practices in providing a reliable characterization of target phenomena and chart the historical interplay between data gathering practices and representational models. After this detour, the points from the current chapter are reprised in Chapter 6. There I discuss the integration of models and data gathering practices in terms of specific historical cases and present an account of representational accuracy and representational content on this basis.

2.1 What makes a model-based inference informative?

First, I want to preempt confusion by repeating a point from Chapter 1: in tying representation to data gathering, I do not mean to claim that models represent data (as in (Van Fraassen, 2008)). Such a view focuses on data to the exclusion of the model's target phenomenon, understood as a component of the setup that generates data. Conflating data and target muddles scientists' conception of their activities, such as when they talk of making interventions on a system based on a model. In such cases, the model appears to be motivating inferences about more than the data itself. On the other hand, it is unclear how a model might be informative with respect to a target phenomenon in the absence of techniques for gathering data from this system. As Barry Barnes notes with reference to anatomical schemas:

> Like all scientific representations, [this figure] is reliably applicable only to aid particular kinds of procedures [of recognition and naming]. In this case the procedures, together with directly associated instrumental interests, are embodied in the role of the anatomist and his student audience. Those who make practical use of such representations are generally well aware that their reliability and applicability is restricted; this awareness is automatically generated in learning to use the representations (Barnes, 1977, p. 8).

If scientific representations are designed to be procedural aids, as Barnes suggests, then researchers who cannot draw any inferential link between their procedural interactions with a phenomenon and a given model have no reason to think this model represents that phenomenon. This motivates the following claim: for a model to be informative about a target phenomenon, it must be possible to relate it to empirical procedures connected to that target.

This claim turns on what it means for a model-based inference to be informative. I begin from the minimal inferentialist condition for scientific representation: to represent a target system a model must enable a competent user to draw informative inferences about that target.[23] In Chapter 1, informative inferences were glossed as inferences that involve positive or negative attributions of states, properties, or behaviors to a target phenomenon in a way that is specific to this target and goes beyond knowledge

[23] Competence involves understanding (i) the circumstances in which it makes sense to apply the model and (ii) the consequences of the application of the model, including its evaluation. These aspects of competence will be discussed later in this chapter and Chapter 6.

explicitly assumed in a model's design. This captures one aspect of what it means for an inference to be informative: its consequence is a statement or proposition that is not already assumed knowledge. This can be put more precisely as follows: an inference is informative about a subject S if and only if its consequence excludes viable alternative possibilities with respect to S.

Model-based inference may be informative with respect to models themselves. Modeling assumptions combined with a set of internal inference rules entail consequences that are not transparent in the design of the model. I call these model derivations or modeling results. They contribute to a wide range of scientific uses for models: studying their internal properties, exploring the consequences of particular theoretical principles or assumptions, comparing one set of models to another, and so on. Inferences that only produce consequences concerning these subjects (i.e., finding an unstable non-zero fixed point) are among the non-representational use of models. Representational uses of models, on the other hand, involve external inference rules resulting in consequences for a target phenomenon. In Chapter 1, I argued that it is unsatisfactory to account for these external inference rules in terms of a component-wise mapping between model and target. To see why, we need only suppose a scenario where scientists do not have a rich model-independent characterization of the target. In this case, the target stipulated by a component-wise mapping would be a copy of the model itself (albeit under an alternate description or "interpretation"). The use of "external" inferences rules would then be equivalent to a mapping from the model to itself (an automorphism, say). Drawing inferences about a target in this case would be no different from drawing inferences about the model from the model. This is a non-representational use of a model disguised as a representational use and thus cannot be used to account for informative inference.

If component-wise mappings will not do, then what kind of rules do account for informative external inferences from model to target? Framing the issue at this level of abstraction is an oversimplification. It presents the inference from model features to target features as a single unmediated step and elides the possibility that reasoning from a scientific model to its target follows a more circuitous path than a simple inference from the model to the target. If model-independent target features are not directly accessible to scientists, then we should consider how model-based inference relates to features of

the world scientists *can* reason about more directly. These features of the world ought to be connected to the phenomena targeted by their models and should also be epistemically accessible to practicing scientists. The clearest candidate are features of the procedures of experimentation and measurement through which scientists interact with target phenomena. It is plausible, then, that informative inferences from model to target are mediated by these data gathering practices and scientists' understanding of them.

I will construct an account of scientific representation that pursues this intuition: in order to draw informative inferences about a target phenomenon from a model, it must be possible to connect the model to empirical procedures connected to this target. Equivalently, model derivations that cannot be connected in any way to possible modes of interacting with a given target are not informative with respect to that target. In a sense, this is an operationalist constraint on what it means for a model-based inference to be informative, as it explicitly links informativeness to data gathering practices. However, I am not making the traditional operationalist claim that the meanings of theoretical terms are, or ought to be, reducible to operations. Rather, I am articulating one of the basic ways that we should think of scientific knowledge—and representation—as empirical. To understand how models enable informative inferences, then, we might focus on the way these inferential activities are constrained by their practical context, i.e., their relation to procedures and techniques involved in data gathering practices. This offers novel strategies for understanding model-based inference in terms of the constraints imposed through the coordination of "top-down" model application with "bottom-up" empirical techniques.[24]

To begin, I will spell out a set of practical constraints on scientific representation. I argue that these provide a set of necessary conditions for effective model-based inference. Where these are satisfied and properly coordinated, scientists are able to use models to draw informative inferences about natural phenomena in the sense I have proposed. The presentation here will be largely schematic, laying out a

[24] If this requirement strikes some readers as too stringent for discussing "representation," I am happy to call a representation *inert* when it involves the use of models to reason about a system for which there are no further means of investigation. By contrast, I am interested in *effective* representation, in which inferences drawn from a model have some practical import for ongoing scientific investigation. This is not to say there aren't a variety of noteworthy modeling practices that occur at the level of inert representations, but these do not take center stage in an account of how models ultimately inform us about their targets.

framework whose details will be filled in through the remainder of the dissertation. The constraints take two general forms: those concerning model application and those concerning data's reliability. Taken together, they establish the relevance of a scientific model for a particular use.

2.2 Relevance conditions for the representational use of models

Sandra Mitchell (2020) distinguishes two norms of scientific modeling: accuracy with respect to existing empirical data and adequacy "judged by how well [models] serve some epistemic or non-epistemic goals" (p. 31). These are logically independent: a model may be adequate for a target but fail to meet some standard of accuracy with respect to empirical results. In such cases we speak of the model being unsuccessful, unfaithful, or misrepresenting its target. In other cases, as in the likelihood and test response thought experiments from Chapter 1, a model fits with data but is inadequate for inferring properties of its target. These are not instances of misrepresentation per se, since the relevant standards for accuracy are prima facie satisfied; instead, these forms of representational failure consist in a model's *misuse* in contexts outside its proper domain of application. In other words, the model is deemed irrelevant for researchers' purposes. An understanding of scientific representation as a practice should include an account of the relevance conditions demarcating model use and misuse.

2.2.1. Domains of application

Chapter 1, Section 4.2 noted that models do not contain any intrinsic specification of the targets to which they apply or the range of conditions under which their derivations are trustworthy. Mitchell and others have noted that this depends on a scientist's purposes or goals, but which purposes are legitimate or properly "scientific"? Boesch (2017a) has offered one answer. He argues that the representational function of a model is constrained by communal licensing activities. These activities distinguish a scientific model

from merely stipulative forms of representation by highlighting how "the construction of the model […] has been responsive to certain theoretical and empirical aims" adopted by a broader community of practicing scientists (2017a, p. 976). The central insight is this: scientists do not simply come across a mathematical structure and then ask themselves, "to what natural system can this be applied?" Scientific models do not arise ex nihilo, but are always designed with specific applications in mind, and in ways meant to accommodate principles and facts accepted by those working within a given field of inquiry. For instance, Lotka's differential equations were not developed as pure abstractions, but were taken from prior work on chemical reaction rates and explicitly tweaked to describe a theoretical predator-prey system. As Knuuttila and Loettgers (2017) put it, Lotka's model-building process was oriented from the beginning toward the specific outcome of explaining ecological dynamics by means of a general equation. Boesch takes this to show how

> the representational nature of the model is constructed through the scientific activities performed by Lotka during the development of the model. Lotka does not merely stipulate that his model targets predator-prey relationships. Instead, he builds this ability into the model during the development of the general equation and further constructs this ability in his application of the question to specific targets (2017, p. 977).

As new theoretical or empirical aims arise, a model is expanded or altered in response, licensing new representational targets. In this way, communal license is taken to determine the intended uses of the model, delimiting the targets about which inferences may be informatively drawn, in a way that is responsive to the received principles of a scientific community.

Several sources of inadequacy are found in the coin toss example from Chapter 1, for example, which indicate the kinds of norms involved in licensing: first, the possible states of a model system ought to accord with background knowledge of its target, if available. The model used to simulate the pattern is based on the assumptions that the angular momentum of quantum systems is quantized and can exist in mixed states. Neither of these assumptions have clear analogues in a coin toss or appear otherwise justified. Second, model states ought not be arbitrarily restricted, e.g., in a way that violates received principles or otherwise lacks theoretical justification, in an effort to "cook up" some desired result. Models for the

quantum mechanics of spin were not developed or reworked with the intention to contribute to the knowledge of coin flips, so restricting a model for the purpose of matching coin results is not licensed.[25]

Scientist's understanding of domains of application is manifest in various activities that contribute to the representational use of models. Contra Contessa (2007), interpretation of model components in terms of target features is not enough for a model to function as a representation. In addition, there must be some prospective means for determining these features through interaction with the target phenomenon. That is, it must be possible to operationalize components of the model by specifying a process or set of processes that yield data relevant to this component. This gives representational significance (or to use van Fraassen's term, empirical grounding) to the modeling results.[26]

The ATLAS experiment at CERN was motivated by an effort to determine the Higgs mass, which functioned as a free parameter in quantum field theory (QFT) models. To be specific, they sought to measure the value of μ in the Higgs sector of the Standard Model with Higgs Lagrangian:

$$L = -\frac{1}{4}\left(F_{\alpha\beta}\right)^2 + |D_\alpha \varphi|^2 + \mu^2 \varphi^* \varphi - \frac{\lambda}{2}(\varphi^* \varphi)^2$$

The process of measuring this quantity followed an established tradition of event counting, background estimation, and statistical analysis within accelerator physics (Cf. Galison (1997)). In other cases, such as the study of psychological constructs or genuinely novel theoretical ideas, an operationalization may not be immediately available, or could be subject to extensive debate. All that matters here is that candidate interactions between an apparatus and a target system are on hand, which render a term such as μ determinable through comparison to data. Without these, the model remains a strictly theoretical object.

[25] While communal standards play an important role in the reception of models, this view does not *reduce* representation to sociological factors. As Knuuttila (2005) point out, the outcome-oriented nature of model-building encourages the use of specific representational features, which can afford scientific reasoning suited to certain purposes *and* limit it in other respects. In this way, a model's adequacy for representing a given target is not solely a matter of community judgment; it is also built into aspects of the model itself. This might explain why some simplifying assumptions or other ways of tweaking a model are apt for its purpose, while others are not.

[26] To be specific, van Fraassen lists this as the "determinability" criterion of empirically grounding a theory. He writes, "any theoretically significant parameter must be such that there are conditions under which its value can be determined on the basis of measurement" (Van Fraassen, 2012, p. 783). I take this aspect of empirical grounding to be necessary for effective reasoning from modeling results to a target system in the world. I am joined in this conviction by Khalifa, Millson, and Risjord, who in a recent work list operationalization as a criterion for scientific representation, understood as "justified surrogative consequence" drawn from a theoretical model (2020).

Another activity that demonstrates scientists' sensitivity to domains of application can be called concretization. As authors—perhaps most notoriously Levins (1966) and Cartwright (1983; 1999)—have pointed out, many modeling assumptions sacrifice realism for their broad scope. This allows for variance in representational uses of models: the same theoretical model can represent many different phenomena and different models can represent different aspects of the same phenomenon. But this also raises hard questions about the relevance of a given model to a given phenomenon. Scientists' sense of a model's domain of application is often honed by prior empirical investigation and training. Where available, they may draw on background knowledge of target behavior or knowledge of the procedures accompanying a particular operationalization in order to further specify or constrain aspects of the model.

Concretization involves adapting a theoretical model to a particular domain of application. This may include supplementing its base structure with further assumptions, empirical laws, or intermediary models.[27] Such supplementation is found in the ATLAS example. Here the base QFT model leaves the mass of the Higgs boson as a free parameter, so physicists are left to treat each mass value as a separate "hypothesis." For a given mass hypothesis, an arsenal of phenomenological methods is used to determine what Higgs production mechanisms and decay modes would dominate at different energies (Fig. 1). These concretize the base theoretical model by building further details and assumptions into simulations. From these simulations, researchers can derive results that are more readily comparable to data.

[27] The notion that theoretical models/mathematical characterizations of phenomena require some further work or specification in order to be applied is acknowledged in various other areas of philosophy of science. In recent philosophy of physics, for instance, this is seen in the distinction between "framework theories" and "concrete theories" (Wallace, 2020), or the claim that physical theories consist of general "fixed formulae" and local, empirical "open-ended formulae" (Kaveh, 2021).

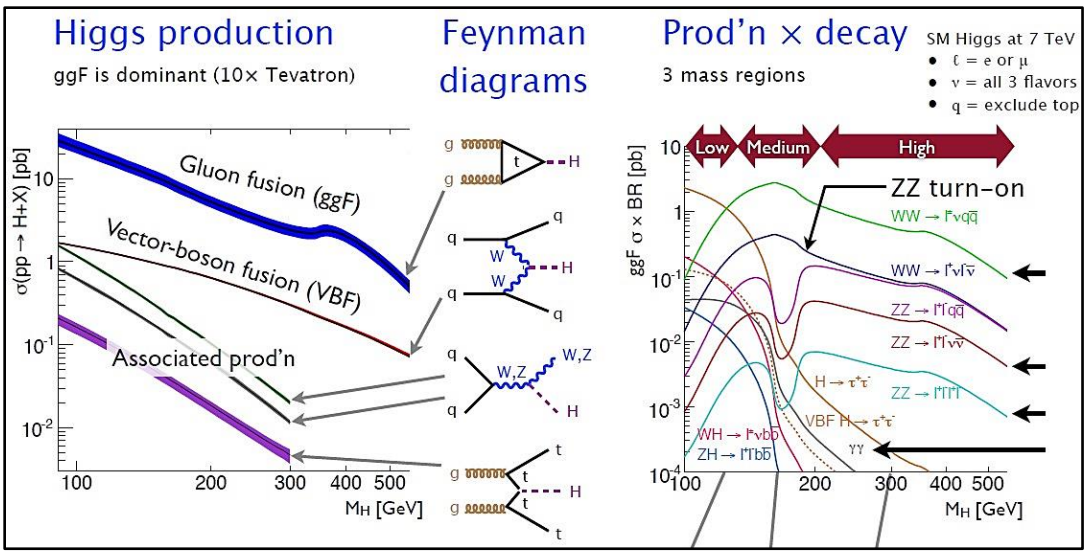


Figure 1: Graphs showing the results of particle phenomenology calculations. On the left, a measure of the probability of four theoretical Higgs production mechanisms (expressed in picobarns) is given for different mass values. On the right, the probability of different decay paths. Images courtesy of Reisaburo Tanaka of the LHC Cross Section Working Group, CERN.

Both operationalization and concretization depend on some grasp of the circumstances in which the model is applicable. This is typically a product of empirical investigations carried out prior to or alongside the development of the model, which establish its communal license. Licensing determines the intended use of the model, delimiting the targets about which inferences may be drawn. Licensing, however, can only be part of the story. In addition to relevance conditions that constrain the scope of a model, there are distinct conditions concerning the nature of the data to which a model is related.

2.2.2. The target-data relation

In Chapter 1, I argued that in many cases of scientific inquiry the only plausible object offered for direct comparison to a model is an empirical result, such as a data structure. The test response homomorphism example from Section 3.3 of that chapter showed that representational adequacy of a model depends not only on features of the model itself, but also on the nature of the data to which a model is related. Without some grounds for claiming the data is linked to an appropriate target, there is nothing to which a model user can draw inferences. This topic has not been routinely addressed in philosophical

discussions of representation, yet it is pertinent for explaining how representation can fail when the patterns in a data structure are not there for "the right reasons." The ability to draw inferences about a target phenomenon from a relation between model and data thus depends not only on model-building activities, but also on the relation between these data and the target phenomenon. As Van Fraassen (2008) puts it,

> There is nothing in an abstract structure itself that can determine that it is the relevant data model, to be matched by the theory. A particular data model is relevant because it was constructed on the basis of results gathered in a certain way, selected by specific criteria of relevance, on certain occasions, in a practical experiment or observational setting, designed for that purpose" (p. 253).

In other words, procedures for generating and processing data must lead to empirical results that serve as a reliable basis for characterizing the model's target. This is a second relevance condition for the representational use of models, reflected in various activities pertaining to data gathering practice.

To interpret data with a model, the data need to be properly tied to the model's target. A justification for such a target-data link is often provided by an understanding of the techniques or instrumentation used in data gathering. For example, researchers at ATLAS collect data produced by detecting the downstream results of proton-proton collisions. Long-standing experimental traditions understand these collisions to result in a shower of particles produced through "decay processes" that will interact in predictable ways with detector materials such as silicon sensors and argon gas. In some cases, the reasons for asserting such a target-data link may depend on an understanding of more basic and general characteristics of the target system, as in nuclear magnetic resonance spectroscopy, which uses a physical theory of the interactions of nuclei with magnetic fields to investigate the structure of molecules. In other cases, these reasons could come from a model lacking independent verification. This may be the case in initial detection experiments, such as searches for particles like the neutrino, where the motivation for collecting data from a target by a particular method comes from interactions anticipated on the basis of that target's model.

This condition alone is not sufficient for claiming that data can be effectively related to a model, nor does it settle the question over whether the model successfully represents its target; it may turn out that the data does not yield any patterns that can be connected to model derivations. Determining that some data falls within a model's domain of applicability is a complex affair. The boundaries of the domain may shift

over time as modelers' grasp on operationalization, applicability constraints, and instrumental techniques develop. Nevertheless, if the representational use of a model is mediated by the relation between models and data, then it is dependent on a proper link existing between its target class and the domain in which it is applied.

A model cannot be used to draw informative inferences about some target system if the data within its domain of application cannot be effectively related to the model. In some cases, there may be good grounds for thinking the data are appropriately connected to the target system, but the resulting data are inadequate for scrutinizing any inferences drawn from the model. There are several ways that data can fail in this regard. Perhaps the most familiar case is when data are irreducibly noisy. The DeBye-Scherrer powder method of X-ray diffraction was found to have this problem with respect to macromolecules such as DNA. Any attempt to produce data by this means resulted in a completely black image that yielded to no methods of analysis and allowed for no interpretations in terms of molecular structure. More generally, the representational use of a model is stymied when data lacks discriminatory power with respect to the model. This means that data lack enough fineness of grain to allow users to distinguish among the aspects of a system to which a model-based inference pertains. This is not necessarily the same as noisy data. For instance, light microscopy is unable to resolve distinct structures below 200 nm, so any inference regarding aspects of a target that can only be resolved below this limit cannot be applied to results from this technique. In other cases this is not intrinsic to a technique, but results from further technical issues such as an inadequately prepared setup. As again in x-ray crystallography, when a crystal diffracts poorly and fails to produce sufficient data to determine the details of a molecule's structure (Fig. 2).

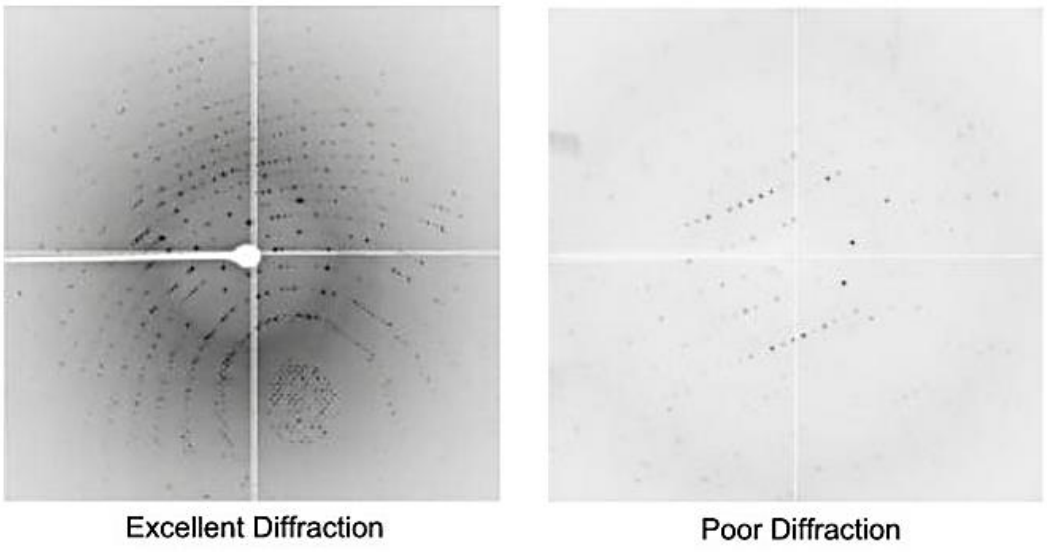


Figure 2: Examples of informative and uninformative diffraction patterns, from Berntson et al. (2003).

Here the problem is not an overabundance of signal, as with noisy data, but a lack of it.

Data processing techniques are the means for extracting patterns from data that are comparable to models. They involve sequences of data cleaning, reduction, analysis (statistical and otherwise), corrections, and other such transformations warranted by an understanding of the experimental context. These data processing techniques progressively alter the initial recordings into a new form, which allows for the extraction of a pattern of interest. But if the generated data is affected by a myriad of unwanted causal factors whose influence is too poorly understood to be removed, warranted pattern extraction may be impossible. As the saying goes: "garbage in, garbage out." The collection of data points or values resulting from data processing is what I call a pattern.[28]

[28] Schematically, the information content of some data set, I(D), can be conceived of as a large collection of recorded marks $\{m_1, m_2, \ldots, m_n\}$. A given technique T involves a series of systematic alterations of this collection $(t_1, t_2, \ldots, t_n)$. Reduction eliminates some sub-set of marks, correction may replace some marks with others, analysis may aggregate a sub-set of marks and provide further data corresponding to collective properties $t_a(m_1, m_2, \ldots, m_n)$. Schematically, these can be written as follows:

- Reduction: $T_r(\{m_1, m_2, m_3, m_4, m_5, m_6\}) \rightarrow \{m_1, m_2, m_3, m_6\}$
- Correction: $T_c(\{m_1, m_2, m_3, m_6\}) \rightarrow \{m_1{*}, m_2, m_3{*}, m_6\}$
- Analysis: $T_a(\{m_1{*}, m_2\}) \rightarrow x_1$ $T_a(\{m_3{*}, m_6\}) \rightarrow x_2$

The output from a data processing technique is what I call a pattern, such as (x_1, x_2) in the schema above. Typically some error values are appended to the output of such a process to reflect uncertainties in the measurement process and processing techniques.

Even when suitably noiseless results arise from a model's domain of application, clear connections between a model and a given data set may not be forthcoming. Data may be, as Nora Boyd puts it, "maladapted" to a model. This may be the case when empirical results presuppose "concepts, parameters, or other such vehicles that are not found in the theory" or presuppose concepts from competitor theories in a way that makes a difference for their evidential impact on a theory in question (Boyd, 2018, p. 408). Establishing connections at the point of data-model interface may be a delicate matter that depends on a distinct suite of techniques.

The ATLAS case presents a particularly complex network of model-data interfaces. Here I'll draw attention to the crucial point of comparison in the paper announcing the evidence for a Higgs boson with mass of approximately 126 GeV. The chief result is summarized in this visualization (Fig. 3):

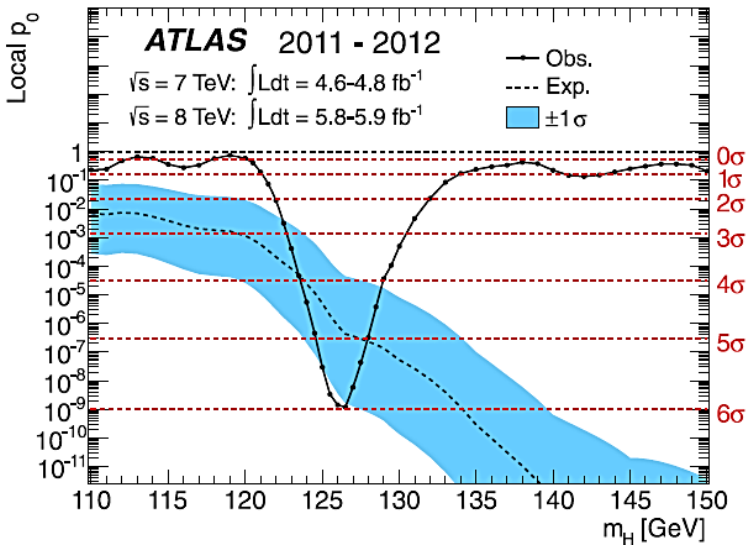


Figure 3: Data visualization from ATLAS (2012).

The solid line in the graph shows the probability p_0 that data from the detector reflect *only* background processes as a function of the hypothesized Higgs boson mass. Background processes are all the decay processes recorded in the detector which do not involve the Higgs. The smaller the value of p_0, the less probable it is that the recorded data consists of only these background processes—those modeled by Standard Model of particle physics without Higgs interactions. The dip in the line down to 5.9σ indicates a probability of nearly 10^{-9} for obtaining the recorded data at this mass value, conditional on expectations for

data derived from a no-Higgs model. Put otherwise, this line indicates a one in 555 million chance that the recorded data was produced in the absence of Higgs interactions. On the other hand, this calculated probability falls largely within one standard deviation (the blue shaded region) of that expected from a Standard Model supplemented with Higgs interactions (dotted curve).

The information presented in this graph—and by the solid line in particular—depicts a form of model-data interface. It shows the outcome of a statistical measure that is specially designed to integrate and compare results derived from experimental data and expectations derived from modeling. More precisely, the probability p_0 results from a measure that relates an observed value to a distribution of expected values conditional on a given hypothesis H_0. We can write this as:

$$\int_x^\infty f(X|H_0)dX$$

where $f(X|H)$ is the probability distribution function of X modeled under the assumption of H_0 and x is the value of X derived from data. Not every form of model-data interface takes this form, nor do they only occur at the final step in the testing of a model. As I will discuss below, a key virtue of viewing scientific representation as an empirical practice lies in the attention it brings to the various ways models are connected to their targets.

2.2.3. Scientific Representation, Practically Constrained

These activities constrain the relevance of models to particular domains of phenomena and data gathering. They can be summarized in two relevance conditions:

Applicability: a model is relevant for representing a target phenomenon *iff* this target falls within its intended scope, as determined by processes of model construction, reception, and modification.

Reliability: an empirical result is relevant for characterizing a target phenomenon *iff* it is the outcome of a reliable sequence of data generation and processing with respect to this target.

Satisfaction of both conditions, along with the assumption that models are only directly comparable to empirical results, can provide a preliminary explanation of the ability to draw informative inferences about a target. Further, representational failures of the type found in the coin toss and response test examples can be explained by one or the other condition not being satisfied.

To put it more formally, I am claiming that the proposition

Agent X uses model M to represent target T

depends on the following:

1) T falls within M's domain of application
2) X competently derives result D from M
3) There is a technique for reliably producing an empirical result E from T
4) The accuracy of D is evaluable with respect to E

Condition (1) ensures that there are ways to operationalize and concretize M so that it can be adapted and applied to the specific circumstances in which scientists interact with T. Condition (2) ensures that the internal inference rules of M are correctly followed. Condition (3) ensures that the data connected to M is linked to T and (4) ensures that this data is processed and comparable to M by means of the relation between D and E. When all of these conditions are in place, a model user is able to derive a result from their model, compare it to an empirical result, and interpret the behavior of T in terms of M on this basis.[29] In this way, the model is informative with respect to T.

2.3 Virtues of this framework

Some work in the philosophy of science literature tends to present modeling and data-oriented activities as operating in isolated domains, where in fact there can exist a dense network of connections. Though a simplification, the above schema of the model-target connection has the benefit of explicitly

[29] This will be discussed in more detail in Chapter 6.

recognizing the substantial feedback that may arise between activities pertaining to data generation and those involving modeling applications and inference. Further consideration of the ATLAS collaboration demonstrates this.

On its face, the statistical measure in Section 3.2 compares experimental data and modeling expectations, both plausibly derived through separate processes. As a matter of fact, both the observed and expected data in ATLAS result from a series of overlapping methods and models. This is not to say "experimental" and "theoretical" components are indistinguishable. There are, on the one hand, the layered materials of the detector itself, the custom electronics and hardware that register the earliest signals, and the computerized "trigger" systems designed to carry out high-speed data reduction and processing algorithms; and one the other hand there are the equations, matrix elements, and Feynman rules, and other calculational techniques of quantum field theory. However, as these poles are brought to meet, one encounters a meshwork of empirical practice in which elements of theory and experiment are each altered and iteratively tuned to one another. A few examples will illustrate this.[30]

Much of the work involved in turning "high theory" particle physics into an empirically tractable form involves the use of simulations. Monte Carlo event generators implement the probability amplitudes and rules of QFT to model the decay modes of particle collisions and derive their kinematics. This includes simulations of processes that lack a theoretical model derived from first principles, such as the confined particle showers called hadronization jets. As one author writes, "There is no unique way to construct a jet: jets are *defined by a jet clustering algorithm*," that is, by reference to a data processing procedure applied to detector results (Nachman, 2016, p. 46). Simulations modeling jets are thus based on an operational definition, which can vary with the preferred algorithm. "String" or "cluster" models each represent jets according to different mathematical structures, with various parameters that may be tuned to data to best accord with an accepted definition of jets. As a result, they may return different results under similar conditions.

[30] See also (Perovic, 2017), who discusses similar features in terms of a "dynamic dependence between the measurement outcomes and the calibration" of the LHC (p. 332).

Just as event simulations are tuned to data, data collection may draw on simulations. A simple case of this is the choice of detector materials and technologies, which were selected and calibrated on the basis of physical modeling and optimization in simulation (Elvira, 2017). But simulations also inform the design of trigger systems, a crucial component of data collection at the LHC. These systems are designed to cut down data by applying criteria that select for specific particles and properties, and reject data that reflect uninteresting background processes. An example based on the electron selection trigger in ATLAS will illustrate the extent to which data and simulation can be interwoven: the cut-off values for variables analyzed by this trigger were initially optimized using electron signal and background distributions based on Monte Carlo simulations. These simulations were based on a detailed model of detector elements—itself calibrated and "realigned" using detector data—meant to show how the detector records signals from a hypothetical electron. When data collection began in 2009, scientists used incoming detector data to tweak inaccurate selection criteria, such as the variable identifying the "shower shape" of an electron interacting with layered silicon. Where clean data was hard to come by, as was the case for identifying and rejecting the background processes associated with electrons, scientists settled for a hybrid data-corrected Monte Carlo approach (Alison, 2012, pp. 136-137).

Accounts of representation are often based on simple predictive uses of models: model results derived on one side, empirical results on the other. Yet the ATLAS case shows that this is not the only way in which data and models may interface. Here simulations used to concretize and derive modeling results are configured using data from the same instruments that produce the observations to which modeling results are compared; and these instruments shape observed data in ways informed by these same simulations. To some it may appear that particle physics has worked itself into a vicious confirmatory circle, and the reasons this is not the case can enrich our understanding of science. I will explore these issues further in the second half of the dissertation, but here it is worth noting we would not think to explore them if philosophical problems were restricted to a picture where "theory proposes and nature disposes."

Consider, in addition, the case of "data-laden models" such as those employed in climate science (Edwards, 1999). Here model parameters and components, such as the gaseous composition of the

atmosphere, may be chosen or designed with significant input from empirical data. In such "semi-empirical" cases the tie to data gathering procedures is to some degree built into the model from the start. This is even more evident in modeling practices found in chemistry and molecular biology, where molecular models are constructed directly from experimental or observational data, albeit with significant additional input through correction and optimization techniques. In such cases, the connection to at least one data generating procedure is straightforward, and the challenge may instead lie in using these models to infer more complex, less empirically accessible properties and behaviors. We can thus distinguish between strictly theoretical models, semi-empirical models, and empirical models. Each is characterized by a model-data interface that is differently situated with respect to model application activities. Other theoretical models derived from first principles may require mediation and concretization with the help of models that are heavily tuned to data, as in the case at ATLAS. As philosophers of science continue to uncover more complex relationships between data and models, the search for *the* representation relation that explains a model's inferential utility seems increasingly quixotic.

The framework proposed here provides a way for bringing the philosophy of scientific representation, which seeks to explain the representational utility of models, further into alignment with the increasingly recognized complexity of model-data interfaces.[31] It does so by shifting the explanans from the precise character of the representation relation to the empirical practices that connect models to their targets. While the relation between model and target may shift from practice to practice, many general characteristics of these practices have been identified by philosophers of science. The activities discussed in Section 3 provide analytical categories for investigating how a particular model-target relation is constituted in each case. Such an investigation can be carried out by attending to how these activities are coordinated in a way that enables informative model-based inferences. Viewing representation as empirical practice thus has the virtue of retaining the explanatory ambitions of the philosophy of scientific representation while accommodating the plurality of model-data interfaces found in contemporary science.

[31] For work complicating the relation between data and models see, e.g., Tal (2012) and Bokulich (2022).

Note that there may still be a role for relations of similarity, morphisms, and the like. Explicit cases of model-target comparison in these terms are sometimes found at the point where a model derivation and processed data are related to one another. But the insight here is limited; it informs us of how modeling results are brought together with data only to the extent that we already grasp the preceding details. Moreover, as I argued in Chapter 1, finding a similarity here does not license an inference to deeper similarities between the initial theoretical model and experimental apparatus that contributed to these downstream results. Such comparisons occur too late in the process to yield a rich explanation of how a model is informative about its target. If, on the other hand, we attend to empirical practices, I believe richer explanations of scientific representation are available. By showing how model-based inferences are connected to the coordinated activities described earlier, we can give an account of the representational utility of models while avoiding the conceptual and explanatory mire of alternative approaches.

2.4 Objections and replies

At this point I anticipate three basic objections. First, that this notion of scientific representation is well and good for established scientific successes, but can't account for models that *mis*represent. Scientists often construct models that turn out to be deeply mistaken when viewed from the perspective of successful theories. Thompson's plum pudding model of the atom and Descartes's vortex model of celestial motion are as demonstrably wrong as any models can get, and yet can allow a competent user to draw some inferences about their purported targets. These inferences may be false, but they are novel inferences nonetheless! They should thus count as examples of scientific representation.

So far I have been stressing the idea that representation depends on activities that can effectively relate data to a model. Am I then defining representation in terms of representational *success*, and conflating misrepresentation with a failure to represent? Is the notion of representation proposed here restricted to cases in which the model-based inferences are true? This objection is easily answered: no. The account

presented above has all the required resources to account for misrepresentation. Effectively relating data to a model is not the same as successful representation. The activities outlined above may result in a model-data interface that shows a prediction to be incorrect. Consider a model of a material's properties (say, its specific heat) that contains deeply erroneous assumptions. If it is a usable model, it still ought to allow for predictions about the behavior of the material, which can be measured against empirical results and found to exceed the threshold for agreement between model and data. In this case the model allows for informative inferences, can be tied to an empirical practice, but does not successfully represent. If no such connection can be drawn between this model and any such data, only then would it fail to represent.[32]

Another challenge to my view is as follows: I have shown that models that do not represent their target successfully can still be regarded as scientific representations, but what of models that are only found to successfully represent their targets well after their initial creation? Two famous examples in recent physics come to mind: gravitational waves, modeled by Einstein in 1918 and detected by LIGO in 2016, and the Higgs boson, first modeled in 1964 and detected by ATLAS and CMS in 2012. Am I committed to the view that these models did not function as representations of their targets prior to the onset of these experiments? Surely not, since prior to these successes there were other attempts to investigate these targets. Gravitational wave detectors have been constructed since the 1960s and Higgs searches occurred at particle accelerators such as LEP before their later discovery. However my view does minimally require that there be *some* understanding on hand for how a data generating procedure could produce results that can be related to model-based inferences. If a model is constructed for which no such account is available, even in terms of prospective technologies, then I am willing to claim that this model cannot serve a representational use. This should not be a cause for concern; there is no requirement that we regard every confabulation with some theoretical backing as a viable representation of worldly phenomena. However, insofar as they aim to make claims of scientific knowledge, it is incumbent on advocates of such models to develop means of connecting them to empirical practice. While it may sound strange to claim that gravitational wave models

[32] Note that this allows me to claim that models whose targets are ultimately believed not to exist can still serve as representations.

could not function as representations at first, I am not the only philosopher of science to take such a view. As van Fraassen writes, "when first introduced, a model or theory may involve theoretically postulated physical quantities for which there is as yet no measurement procedure available." In such a case, "development, simultaneously strengthening the theory and introducing new measurement procedures, is not adventitious or optional: it is a fundamental demand in the empirical sciences" (Van Fraassen, 2012, p. 782).

A final related objection concerns the use of models to derive modal claims about a general phenomenon or class of phenomena. Schelling's model of segregation, for instance, is not designed for comparison to any specific patterns of segregation within actual cities, but rather to illustrate one possible causal factor for the general phenomenon of segregation. Similarly, certain economic or evolutionary models are not clearly designed to interface with any specific form of data gathering and are instead employed for the purpose of making claims about the theoretical processes or principles that might have produced a generally observed patterns (such as the predominance of sexual dimorphism or the tendency of bad money to drive out the good). These models are used in scientific explanations of such phenomena and in this way appear to function as representations, but my emphasis on the relation between models and empirical practice would appear to deny them this role.

In response, it should first be noted again that not all instances of modeling are acts of scientific representation. Models may be used in strictly theoretical ways, as a means of testing the consistency of principles, providing a proof of concept, or exploring a theoretical space of possibilities without regard for their relation to empirical practice. It is also true that the relations between models and data are not always as clear-cut as that presented in Section 3.2. Still, I hold that for a model to be informative of a natural system, there ought to be some basis for thinking it applies. In the case of application to a general class of phenomena, this requirement may be satisfied by procedures that employ a very loose notion of "fit" between modeling results and summary data. For some purposes it may be satisfactory that a model can roughly reproduce a pattern that meshes with some low-resolution description of segregation, say. But we must also attend to how such a model can generate informative inferences about its target. Cases like the

Schelling model are not used to generate new patterns of segregation that can be anticipated from further observations; they are used to identify possible causal factors underlying established results. Informative inferences drawn from such models primarily involve the relationship between target phenomena and the model elements built in as *assumptions* rather than the phenomena and model *results*. They are further investigated through procedures that probe the empirical grounds for such assumptions. Indeed, this is precisely how researchers have responded to the initial Schelling model (Aydinonat, 2007). The inferential utility of models can thus depend on the empirical validity of their constitutive assumptions as much as (or even more than) any specific derivations.[33] These cases are not a refutation of my position so much as they signal the need for a more fine-grained analysis of how models are connected to empirical practice.

2.5 Looking ahead

The remainder of this dissertation will examine in further detail the relationship between modeling inference and licensing, on the one hand, and data gathering practices on the other. Chapter 3 will examine the notion of data reliability in greater detail, introducing the notion of an empirical program. Chapter 4 will chart the development of empirical programs over time using an extended historical case. Chapter 5 will analyze the relationships between models and empirical programs illustrated in Chapter 4. Finally, Chapter 6 will present novel accounts of representational accuracy and representational content based on the preceding study.

References

Alison, J. (2012). *The Road to Discovery: Detector Alignment, Electron Identification, Particle Misidentifcation, WW Physics, and the Discovery of the Higgs Boson.* Dissertation: University of Pennsylvania.

[33] The manner in which empirical results inform the construction of models is explored further in Chapter 5 and 6.

Aydinonat, N. (2007). Models, conjectures and exploration: an analysis of Schelling's checkerboard model of residential segregation. *Journal of Economic Methodology, 14*, 429-454.

Barnes, B. (1977). *Interests and the Growth of Knowledge.* New York: Routledge.

Berntson, A., Stojanoff, V., & Takai, H. (2003). Application of a neural network in high-throughput protein crystallography. *Journal of Synchrotron Radiation, 10*, 445-449.

Boesch, B. (2017a). There is a Special Problem of Scientific Representation. *Philosophy of Science, 84*, 970-981.

Boesch, B. (2017b). The Means-End Account of Scientific, Representational Actions. *Synthese*, published online.

Bokulich, A. (2022). Toward a Taxonomy of the Model-Ladenness of Data. *Philosophy of Science, 87*, 793-806.

Boyd, N. (2018). Evidence Enriched. *Philosophy of Science, 85*, 403-421.

Burnston, D., Sheredos, B., Abrahamson, A., & Bechtel, W. (2014). Scientists' Use of Diagrams in Developing Mechanistic Explanations: A Case Study from Chronobiology. *Pragmatics and Cognition, 22*, 224-243.

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

Cartwright, N. (1999). *The Dappled World: A Study of the Boundaries of Science.* New York: Cambridge University Press.

Contessa, G. (2007). Scientific Representation, Interpretation, and Surrogative Reasoning. *Philosophy of Science, 74*, 46-68.

Coopmans, C., Vertesi, J., Lynch, M., & Woolgar, S. (Eds.). (2014). *Representation in Scientific Practice Revisited.* Cambridge, MA: MIT Press.

Edwards, P. (1999). Global Climate Science, Uncertainty and Politics: Data-laden Models, Model-filtered Data. *Science as Culture, 8*, 437-472.

Elvira, V. (2017). Impact of Detector Simulation in Particle Physics Collider Experiments. *Physics Reports, 695*, 1-54.

Galison, P. (1997). *Image and Logic: A Material Culture of Microphysics.* Chicago: University of Chicago Press.

Kaveh, S. (2021). *A Prescriptivist Account of Physical Theories.* Dissertation: University of Pittsburgh.

Khalifa, K., Millson, J., & Risjord, M. (2020). Inquisitive Inferentialism About Scientific Representation. *Draft*.

Knuuttila, T., & Loettgers, A. (2011). The Productive Tension: Mechanisms Vs. Templates in Modeling the Phenomena. In P. Humphreys, & C. Imbert (Eds.), *Models, Simulations, and Representation.* New York: Routledge.

Knuuttila, T., & Loettgers, A. (2013). Basic Science Through Engineering? Synthetic Modeling the Idea of Biology-Inspired Engineering. *Studies in History and Philosophy of Science Part C, 44*, 158-169.

Knuuttila, T., & Loettgers, A. (2016). Model Templates Within and Between Disciplines: From Magnets to Gases--and Socio-economic Systems. *European Journal for Philosophy of Science, 6*, 377-400.

Levins, R. (1966). The Strategy of Model Building in Population Biology. *American Scientist, 54*, 421-431.

Longino, H. (1990). *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry.* Princeton: Princeton University Press.

Lynch, M., & Woolgar, S. (Eds.). (1990). *Representation in Scientific Practice.* Cambridge, MA: MIT Press.

Mitchell, S. (2020). Instrumental Perspectivism: Is AI Machine Learning like NMR Spectroscopy. In M. Bertolaso, & F. Sterpetti (Eds.), *A Critical Reflection on Automated Science* (pp. 27-42). New York: Springer.

Morgan, M., & Morrison, M. (Eds.). (1999). *Models as Mediators: Perspectives on Natural and Social Science.* New York: Cambridge University Press.

Nachman, B. (2016). *Investigating the Quantum Properties of Jets and the Search for a Supersymmetric Top Quark Partner with the ATLAS Detector.* Dissertation: Stanford University.

Perini, L. (2012). Image Interpretation: Bridging the Gap from Mechanically Produced Image to Representation. *International Studies in the Philosophy of Science, 26*, 153-170.

Perovic, S. (2017). Experimenter's regress argument, empiricism, and the calibration of the large hadron collider. *Synthese, 194*, 313-332.

Pickering, A. (1992). From Science as Knowledge to Science as Practice. In A. Pickering (Ed.), *Science as Practice and Culture* (pp. 1-28). Chicago: University of Chicago Press.

Tal, E. (2012). *The Epistemology of Measurement: A Model-based Account.* Dissertation: University of Toronto.

Tibbets, P. (1990). Representation and the realist-constructivist controversy. In M. Lynch, & S. Woolgar (Eds.), *Representation in Scientific Practice* (pp. 69-84). Cambridge, MA: MIT Press.

Van Fraassen, B. (2008). *Scientific Representation: Paradoxes of Perspective.* New York: Oxford University Press.

Van Fraassen, B. (2012). Modeling and Measurement: The Criterion of Empirical Grounding. *Philosophy of Science, 86*, 773-784.

Wallace, D. (2020). On the Plurality of Quantum Theories: Quantum theory as a framework, and its implications for the quantum measurement problem. In S. French, & J. Saatsi (Eds.), *Scientific Realism and the Quantum.* New York: Oxford University Press.

Waters, K. (2019). An Epistemology for Scientific Practice. *Philosophy of Science, 86*, 585-611.

Woody, A. (2004). More Telltale Signs: What Attention to Representation Reveals About Scientific Explanation. *Philosophy of Science, 71*, 780-793.

Woody, A. (2014). Chemistry's Periodic Law: Rethinking Representation and Explanation After the Turn to Practice. In L. Soler, S. Zwart, M. Lynch, & V. Israel-Jost (Eds.), *Science After the Practice Turn in the Philosophy, History, and Social Studies of Science* (pp. 123-150). New York: Routledge.

3.0. Reliable Data-Gathering: The Role of Empirical Programs

Chapter 2 argued that data must reliably carry information about a target phenomenon to enable surrogative reasoning, understood as informative model-based inference. Here I aim to show how data's reliability depends on the specific organization of knowledge and activities that characterize *empirical programs*: distinct bodies of empirico-technical knowledge organized around a corresponding data-gathering procedure.[34] Since effective representation requires reliable data, it follows that features of these programs which help ensure the reliability of empirical results likewise serve as a constitutive element in the representational use of theoretical models. This chapter thus explicates a basic requirement on data for them to enable scientific representation: simply put, the data must be produced in the right way. Of course, "the right way" depends on the specificities of the data-gathering procedure at hand. Since we cannot catalogue what must be done in every possible procedure, I will present a general account of the epistemic resources that scientists draw on in such circumstances, which I characterize in terms of the *theories of technique* and *in situ strategies* that comprise empirical programs. I argue that these inform the practical reasoning that guides the individual activities of researchers, and that these activities, when combined with other hallmarks of experimental practice, ensure that data is reliably produced.

While acknowledging their interconnection, I want to keep two aspects of reliability distinct. I call *comparability* the property of data that enables their comparison to modeling results. Making data comparable in this way typically depends on the techniques that clean up unprocessed data and remove unwanted features. This requires that the initial recordings contain useful information about the target phenomenon being modeled, a requirement I describe below in terms of data's *consistency and coherence*. Generally speaking, data depends on the realization of all three of these conditions to be reliable. But in all

[34] I've chosen the phrase "data-gathering practice" to be neutral between measurement (or more classically, observation) and experimentation. While observation and experimentation were not always differentiated at the early stages of modern science (Cf. (Anstey, 2014)), they have been viewed as practically and epistemically distinct since at least the 19th century (as in (Herschel, 1831, p. 76). I believe the positive contributions in this chapter apply to both forms, although I'm sympathetic to work that questions the extent to which techniques of intervention are epistemically distinct from or superior to 'non-intervention' (Cf. (Mikhalevich, 2016)).

but the penultimate section of this chapter, I will focus my discussions of reliability on the procedures leading to an initial recording that ensure it is consistent and coherent.

What makes a data-gathering procedure reliable? Bogen and Woodward (1992) give the example of evaluating reliability in terms of the long-run error characteristics of a detection process—the degree to which the variance in a repeated measurement corresponds, or fails to correspond, to the variance of a natural state.[35] Not all judgments of reliability need take this form, but I believe Bogen and Woodward are right to emphasize the importance of scientists' understanding of uncertainty, broadly construed, and the techniques that might account for it. Data's reliability depends on the consistency and fidelity of the outputs of experimentation and measurement practice. Determinations of reliability require, on the one hand, establishing that a procedure carried out numerous times would produce recordings consistent with one another within some margin of variability—call this the *consistency* of data. On the other hand, researchers need some reason to believe that their data bear some systematic connection with the system of interest—that is, that the data are *coherent* with the best available understanding and analysis of techniques being employed. If data are consistent and coherent, then they are a reliable source for further processing. Securing the consistency and coherence of data is typically a matter of addressing statistical and systematic uncertainties. The former result from variations due to innate imprecisions of the instruments in use and, in some cases, natural fluctuations in the system of interest. These are reduced by collecting a sufficient quantity of data and applying basic statistical analyses, such as averaging a large, normally distributed dataset. Systematic uncertainties refer to factors in the experimental arrangement that systematically bias results in a particular direction. For example, thermal activity within a CCD detector may kick out free

[35] This link between the state of a system and the output of an instrument is a commonsense feature of measurement, which Van Fraassen (2008) formalizes into a basic criterion. Chang (1995) likewise cites this link as an intuitive aspect of a measurement's reliability, but notes that it becomes vexing in the case of theoretical concepts that are not precisely defined independent of available measurement techniques. Such concepts, as with the case of energy in the early history of quantum mechanics, offer poor means for estimating the true value of a measured quantity, and open judgments of reliability to the threat of circularity. Chang discusses a possible solution of this problem in the form of the cross-calibration of multiple measurement procedures, a topic examined in greater depth in his (2004) and further by Tal (2012; 2017). See (Hacking, Representing and Intervening, 1983) and (Franklin, 1986) for earlier discussion of this method.

electrons which are then captured, producing a layer of background noise that systematically increases the instrument's registered signal. Data that are within an accepted range of statistical uncertainty are consistent and data that are within an accepted range of systematic uncertainty are coherent.

If producing reliable data is necessary for scientific representation, then a fully explanatory account of representation should provide an account of data's consistency and coherence. Accounting for the consistency of data is typically done by statistical analyses. Accounting for the coherence of data is more subtle and will be the primary focus of this chapter. The account that follows will show how key features of empirical programs enable researchers to identify factors in the data-gathering setup that influence the recording of data, theorize the dependency relationships between these factors and the resulting data, and navigate local contingencies. When combined with practices of record-keeping, cross-checking, and revision, they result in the control of uncertainties and the production of coherent data. In this way, I uncover one layer of the scientific work that enables models to be effectively connected with the world.

3.1. Scientific practice and empirical programs

A social practice is generally defined as a stable pattern of coordinated activities (Rouse, 2006).[36] Philosophers often distinguish actions or activities from mere acts by claiming actions possess intentionality. Intentional actions are characterized in terms of the purpose or aim they fulfill (Gangloff & Allamel-Raffin, 2014). Thus, despite outward appearances, an involuntary spasm that causes one's arm to be raised is not analytically the same as the act of raising one's hand in class.[37] This gives actions normative content; they can be performed better or worse depending on how effectively they fulfill their purpose. The evaluability of actions is in turn responsible for their rule-bound appearance. Someone who 'raises' their

[36] Major theories of social practice have their roots in the existential phenomenology of Martin Heidegger (1963)—most famously promoted in Anglo-American circles by Dreyfus (1991)—and Ludwig Wittgenstein's writing on rule-following (1953)—further popularized by Kripke (1982). Among social theorists, Pierre Bourdieu (1977) has developed a highly influential theory of practice.

[37] This appeal to the purposiveness of action has a history going back at least to Austin's (1962) analysis of speech acts in terms of their felicity conditions.

hand in class by pointing it at the nearest wall may get the attention of their teacher, but the action will not receive the appropriate uptake (to borrow Austin's term); the teacher is likely to be confused and either call on someone else or interrupt the class to ask what the student is doing. In this sense, the student was performing the action incorrectly—they will have failed to fulfill the purpose of signaling readiness to contribute.

This example helps illustrate two further features of social activities: first, their rule-bound character may not be explicitly articulated, either verbally or in the conscious thought of the agent. While the initial learning process for a skilled action may be laborious and require a significant degree of instruction and conscious attention, habit and training can render its execution "automatic."[38] After a short length of time in formal educational settings, students engage in hand-raising without any need to reflect on what they're doing; they tacitly know how to perform the correct actions within the context of classroom discussion.[39] A comparable activity in the laboratory environment is pipetting. The basic use of a pipette can be summarized in a simple series of steps: attach a clean tip, set the quantity of solution to draw, depress the button until you meet resistance, submerge the tip end in solution, release the button to draw up solution, and then depress the button fully to eject. And yet, someone learning to pipette quickly becomes familiar with mechanical subtleties: air can get trapped in the tip and form bubbles that cause the pipette to eject poorly; the device will draw up solution in different amounts depending on the viscosity of the solution and how quickly the button is pressed and released; the springs can give out if used too roughly; the tips seem to draw up solution more effectively when held against the side of a vial; for especially viscous substances, it may help to widen the tip mouth with a pair of scissors; and so on.[40] Such lessons are rarely given upfront

[38] Empirical study of the phenomenon of automatization through learning are widespread in psychology and cognitive science; early citations include Shiffrin and Schneider (1977) and Logan (1988).
[39] The better-known example here is bike riding. Classic references on this point are Ryle's (1949) discussion of "know-how" and Polanyi's (1966) notion of tacit knowledge. Philosophers, historians, and sociologists of science have discussed the role of training and "background" understanding in the observations of scientists (Hanson, 1958; Kuhn, 1962), in their judgments (Daston & Galison, Objectivity, 2007), and in the transmission of scientific knowledge (Collins, 2010).
[40] These subtleties are likely responsible for the poor performance of automated devices. At the lab I visited, the automatic pipetter was treated with disdain.

but are absorbed through a series of attempts, failures, and corrections in response to piecemeal advice. The learning process is embodied in the sense that it occurs in tandem with the fine-tuning of sensorimotor capacities in response to events that frustrate fulfillment of the activity's purpose. Eventually what began as a painstaking affair turns into fluid motion and finally part of the monotony of bench work.

The failed hand-raising example highlights a second aspect of social activities: their purpose is determined by their practical context. Raising a hand fulfills the aim of signaling readiness because the action has an established meaning within the practice of classroom discussion. In this case, training and reinforcement preserve the conventional meaning of such a gesture. A skill such as pipetting cannot be replaced with any other convention to accomplish the same goal. Nevertheless, the purpose of pipetting is also conditioned by practical context. The skill is useful because, when properly executed, it allows for the consistent transfer of a precise quantity of solution, and consistency and precision in experimental conditions are basic desiderata of scientific practice. Analysis of a given practice may thus reveal different activities located within a nested series of aims—dispensing fluid with a pipette, using the technique of pipetting to measure precisely, measuring precisely to ensure consistent conditions, and so on. This flexibility in scope and application also raises worries that the practice concept is too vague and variform to serve much explanatory use (Turner, 1994).[41] I'm sympathetic to this critique; if an account of data-gathering practices is going to explain the reliability of their data, then work needs to be done in clearly specifying the elements of such practices and their connection to this explanandum.

Viewing a technique like pipetting in terms of a social practice does not make it a mere convention like hand-raising. Unlike the latter, there is not a near-limitless variety of alternative actions equally suited to the same purpose. This is because scientific practices are comprised of what Chang (2012) calls *epistemic* activities, "a more or less coherent set of mental or physical operations that are intended to contribute to the production or improvement of knowledge" of the natural world (p. 15). One foundational assumption of scientific practice is that the natural world is the way it is—that there is a fact of the matter—and so any

[41] Similar criticisms have been made of structural social theories, such as that of Giddens (1984).

activity that sets knowledge of this world as its standard is beholden to something irreducible to human social conventions. Over the last decade, Chang has sought to develop a theory of scientific practices by first developing a "grammar" of epistemic activities that can be applied in the analysis of concrete cases.[42] I share this aim, but rather than directly follow his approach, I'll be proceeding from a few detailed examples to generic features responsible for the consistency and coherence of data. The key component here will be the notion of an empirical program, which can be viewed as a specially developed body of local knowledge designed to elicit the "rules" that guide activities within an experimental practice. By calling the knowledge involved in empirical programs local I refer to the fact that it is limited in its applicability and focus to the conditions within which specific data-gathering techniques are employed. When presented in the form of seminars or textbooks, this knowledge does not bear the indefinite scope of a course on the theory of electromagnetism or Keynesian macroeconomics. The knowledge characteristic of empirical programs is thus distinguishable from so-called "high theory" provided by such sources. Philosophical and historical inquiry into the methods of contemporary scientific experimentation reveals a body of knowledge that may incorporate many elements of these theories but is not reducible to them. Such a view was pithily expressed by Hacking's slogan, "experiment has a life of its own," and has received further historical and philosophical elaboration in writing such as Galison's (1987; 1997) on instrumental and experimental traditions, Baird and Faust's (1990) on instrumental progress, and Waters's (2004) on "pools of knowledge" in the practice of classical genetics.

In sum, studies of scientific experimentation show that there exists a body of procedural knowledge and resources for investigative reasoning that accumulate around experimental and instrumental design, specific sets of techniques for bringing about a desired effect and analyzing the outcome, and an attendant understanding of the operations of the experimental apparatus and the relevant conditions in which an experiment is being carried out, all in the interest of producing reliable results. A proper understanding of modern empirical programs calls for recognition of the way that these elements are coordinated in practice.

[42] He cites Gooding (1992) and Hacking (1992) as predecessors.

Take any collection of instruments and techniques that has undergone a historical process of stabilization and standardization resulting in an established method or procedure for gathering and refining data from a target phenomenon. As they approach stability, these practices carve out a particular causal niche based on the form interaction between an instrument and a system of interest. A specific body of knowledge about this interaction is simultaneously developed.[43] These form an aggregate body of technique-based, locally-oriented knowledge that I refer to as the empirical program—a patchwork of pertinent theories, technical know-how, and practical considerations that together provide the core concepts, strategies, and concerns by means of which researchers make the specific decisions and perform the particular actions that take them from the initial preparation of a system to a final result.

I think of empirical programs in terms that extend beyond the notion of a protocol. Whereas protocols typically designate the series of steps taken to carry out single operations and experiments, empirical programs encompass the conceptual and material resources used for a range of measurements and experiments; the same program may direct a scientist to two different protocols if local contingencies differ.[44] The following section will describe how empirical programs direct scientific activity in this way. First, I outline the structure of a data-gathering practice as a nested sequence of activities. I then describe the two main features of empirical programs: the theories of techniques characterizing the interaction between apparatus and target phenomenon, and the in situ strategies that weigh the aptitude of different techniques for a given circumstance. Finally, I explain how these aspects of an empirical program are incorporated into the practical reasoning of scientists, which culminate in imperatives to carry out specific actions.

[43] A number of philosophers emphasize the interaction of instrument and object as the core of experimentation. Hans Radder writes, "Experimentation involves the realization of a number of manipulations of the object and the equipment, brought into mutual interaction, and the theoretical description (or interpretation) of these manipulations and their results." (Radder, 2012, p. 53). Similarly, Rom Harré has described laboratory phenomena as formed from "a novel kind of entity, an indissoluble union of apparatus and world" (2003, p. 31). Prior to this, Rheinberger had described 'epistemic things' and 'technical objects' as "two different yet inseparable elements" of an experimental system (1997, p. 28).

[44] In this sense, they are better compared to Lange's (2003) notion of an "experimental instruction," which provides a 'recipe' for setting up and reproducing a kind of experiment independent from the aims of its original designer.

3.2. How empirical programs guide data-gathering practices

3.2.1. The structure of data-gathering practices

I begin with a schematic outline of the structure of data-gathering practices in their procession from target preparation to an empirical result. A simple linear description gives their bare structure, In a course on experimental methods at the LHC (Hong, 2018), the instructor wrote this as:

"MAKE → SAVE → SEARCH"

This refers to the major steps carried out in the Higgs detection efforts of the ATLAS experiment: conditions are arranged for the potential production of the Higgs, the data are recorded, and then processed. These roughly align with three generic stages of data-gathering: (1) preparation of the proper materials and conditions, (2) execution and recording of the system-apparatus interaction, and (3) processing of the resulting data.

In any actual case a more complex structure unfolds from this framework. Consider a generic description of the procedure for analyzing proteins by nuclear magnetic resonance spectroscopy, or NMR (Fig. 1):

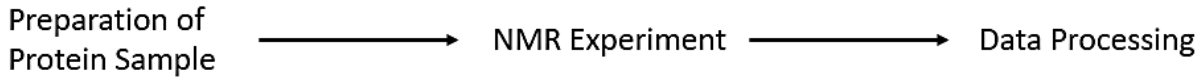


Figure 1: NMR spectroscopy procedure

Each stage shown here can be decomposed into a separate sub-sequence. In the lab I visited, sample preparation involved ordering specially prepared (i.e., genetically engineered) bacterial cells, growing them in a proper medium, inducing protein expression, extracting and purifying the protein, and finally preparing it to undergo NMR (Fig. 2).

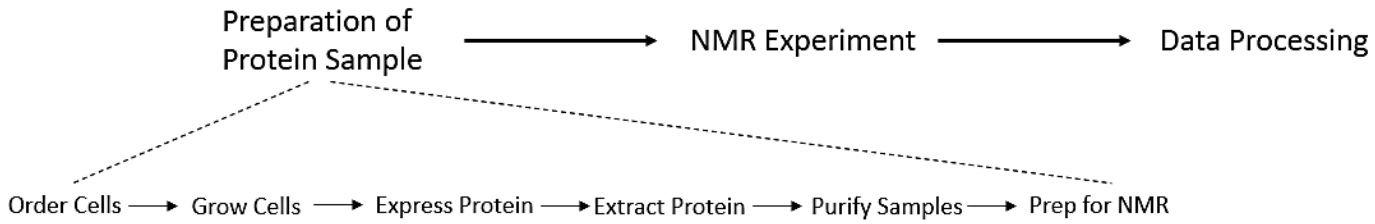


Figure 2: The NMR procedure expanded once.

Each stage in a sub-sequence can unfold into a further series of steps. Protein extraction, which begins when the cells pass a threshold of optical density (measured with a photospectrometer), requires extracting the cells from solution (growth "broth") by centrifugation, decanting and resuspending them in affinity buffer, breaking them open with a microfluidizer, and removing insoluble cell parts (again by centrifugation). And so on (Fig. 3).

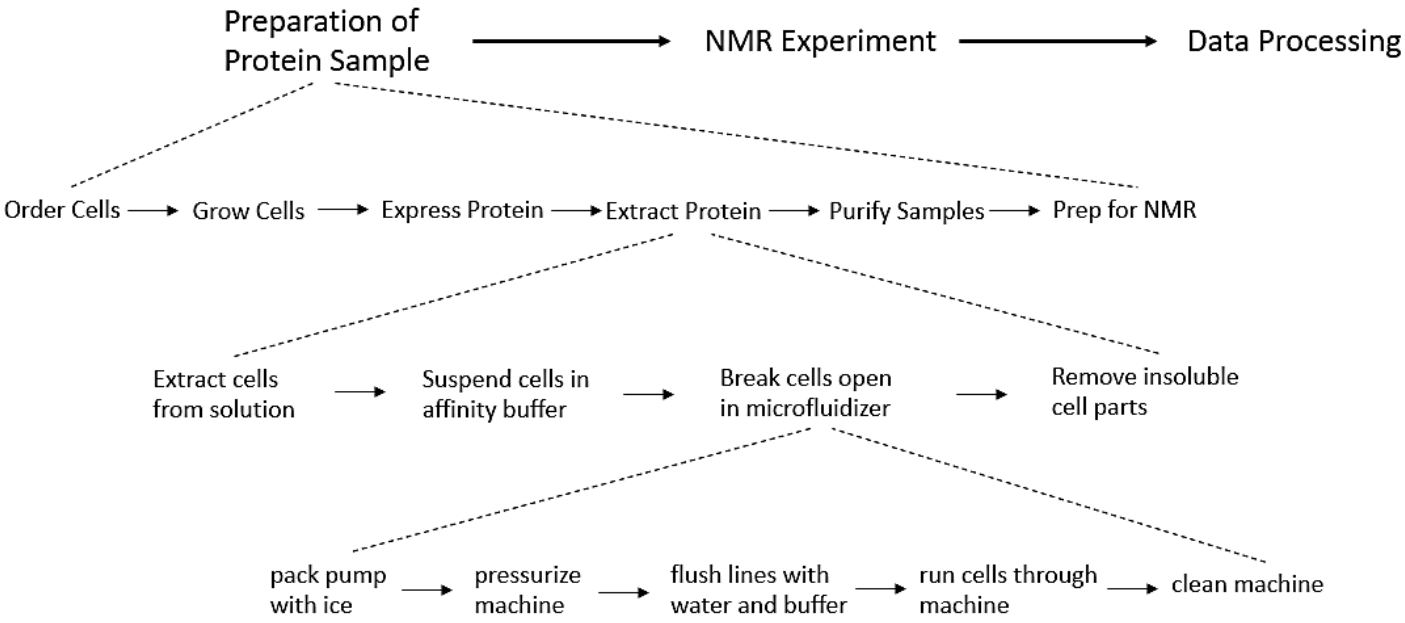


Figure 3: The NMR procedure expanded twice more.

For my purposes the decomposition need not be strictly unique, but it is meant to distinguish stages in a manner that aligns with a practicing scientist's planning.

It is also meant to depict data-gathering practices in a way that is useful for highlighting certain features. Here, then, they are shown in terms of a serial structure in which sequences of activities occurring

at smaller temporal scales can be hierarchically nested within larger-scale procedural descriptions.[45] As in the above paragraph, each stage calls for a particular set of (sometimes overlapping) materials, instruments, and corresponding techniques—buffers, beakers, the photospectrometer, the centrifuge, the microfluidizer, decanting, centrifugation, etc. At each stage these materials and techniques are combined and oriented toward a specific goal, such as extracting cells from a solution, which itself serves a higher-order goal specified in the parent stage of the upper rung of a hierarchy (e.g., extracting protein). Moving upward, one approaches the broadest encompassing goals of a practice. Moving downward, as one approaches the lower rungs of the hierarchy, the sequences are given in terms of operations that are increasingly concrete, albeit laden with the specialized vocabulary of laboratory instruments and procedures. These bottom out in acts whose order and execution are informed by practitioners' procedural knowledge and causal understanding of their experimental context.

This structure illustrates the localized nature of knowledge constituting empirical programs: for standard uses, much of the skills and information required for successfully operating a microfluidizer do not depend on those involved in other steps, or vice versa. Much of this knowledge is therefore relegated to a specific activity within a larger sequence. This in turn affects scientific problem-solving in experimental contexts. That is, localization of knowledge around particular technical applications within a hierarchy can "contain" the impact of anomalies arising within a specific sub-sequence. It does so by prompting investigation of the activities of this sub-sequence before doubt is directed at other activities on the same rung or higher up in a procedural hierarchy. An anomaly that can be attributed to technical error within this sequence ought not disturb a practitioners' understanding of other techniques, as this can be addressed by correcting the execution of the technique in focus. Even when a persistent anomaly indicates a failure to properly understand a technique, this need not shake researcher confidence in other stages, as these typically

[45] This accords with some recent accounts of the cognitive representation of small-scale actions and action planning (Land, Volchenkov, Blasing, & Schack, 2013; Seegelke & Schack, 2016).

employ non-overlapping background knowledge. This segmentation and technique-based localization of knowledge can thus contribute to the stability of a data gathering practice.[46]

At each step of a data gathering practice, scientists may draw upon a varying collection of instruments, calculational and instrumental techniques, and prior information in ways that are tailored to the task at hand. Some instruments are designed for a highly specific task. Others are general purpose and may take on a recurring role. Knowing what techniques to employ in what circumstances and to what end depends on understanding the uses of a wide variety of materials, techniques, and instruments. This empirico-technical knowledge is what comprises an empirical program. I focus on two chief components, which I call *theories of technique* and *in situ strategies*. These provide researchers with the "what" and "how" of a practice—that is, a conception of the way their apparatus interacts with the system of interest and an understanding of how this interaction is best carried out in light of overarching research goals and local contingencies. Ultimately I aim to show that the better one understands the influence of experimental conditions, the more reliable the data that results. Or as Mitchell puts it, "As the causal structure underlying the generation of [target] behaviors is better understood, a methodology tuned to detecting the types of contingencies and interacting factors is more likely to yield useful knowledge" (2009, p. 109).

3.2.2. Theories of technique

A theory of technique combines (i) an account of the target system under a particular theoretical description and mode of preparation with (ii) an account of the workings of the elements of an apparatus and its effect on a system so described. For advanced methods, a significant share of the theory is acquired through textbooks and courses designed to transmit this knowledge within a prescribed range of specialization. Titles of textbooks thus vary from the highly general (*Understanding NMR Spectroscopy*) to the more specific (*Protein NMR Techniques)* to those dealing with a single stage of a data generating

[46] Hacking (1992) presents an explanation of the stability of laboratory sciences based on the mutual adjustment of fifteen taxonomic items (divided among the broader categories of ideas, things, and marks).

practice (*Signal Treatment and Signal Analysis in NMR*). Still, the broad contours of a theory of technique can be found in summary texts. For example, a work on protein X-ray crystallography clearly lays out the key factors for understanding this form of experiment in its initial chapters: "Crystallizing a Protein," presents the principles and techniques of protein crystallization and preparation for X-ray exposure; "X-ray Sources and Detectors," covers the two central components of the experimental instrument; "Crystals," gives the physical and geometrical theory of crystalline materials; and "Theory of X-ray Diffraction by a Crystal," is a theoretical account of the scattering interaction between waves and atoms arranged in a crystal lattice (Drenth, 2007). The preparation, the instrument, the system under preparation, and the system-apparatus interaction are each laid out for the reader, followed by an account of the resulting data ("Average Reflection Intensity and Distribution of Structure Factor Data") and further means of analyzing and improving results. Again, three stages of the practice—preparation, recording, and analysis—are distinguished. Each is informed by its own collection of pertinent theories (Fig. 4).[47]

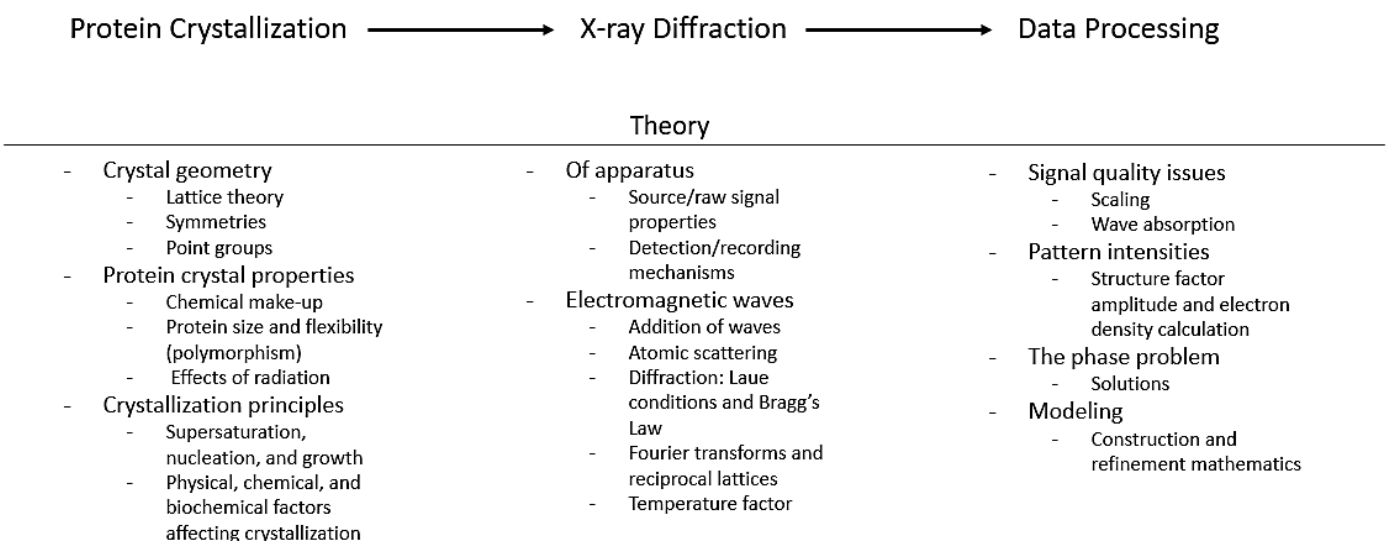


Figure 4: Elements of theory for different stages.

[47] The restriction of certain theoretical knowledge to a single stage here is not strictly accurate. For example, the wave theory of diffraction is relevant to the design of data processing techniques designed to solve the protein crystal, and this diffraction theory itself draws on crystal geometry. The location of the theories is here meant to emphasize those stages in the practice that they best serve to inform or explain.

A theory of technique equips scientists with an understanding of the kinds of systems a given instrument can fruitfully investigate, how the instrumental apparatus interacts with such a system, the way data are recorded, and how to best interpret and alter these recordings. It does this by packaging an account of core principles underlying a target-apparatus interaction and the standard procedure for its execution with an account of the common conditions that affect the quality of its outcome. The theory thus serves to demarcate the general bounds of applicability of a data generating practice and provides practitioners with a guide to those factors contributing to data outcomes. Generally speaking, guidance comes in the form of causal generalizations pertaining to the context at hand, that is, statements that specify general dependency relations between different factors present in the data gathering set-up. Here I have in mind claims such as the following:

- "In any form of spectroscopy the presence of electro-magnetic radiation induces transitions from the ground to the excited state and vice versa, consequently the *net* absorption depends on the population difference between the two" (Rule & Hitchens, 2006, p. 7).
- "[Spherical aberration] is caused by the lens field acting inhomogeneously on the off-axis rays" (Michler, 2008, p. 23), or
- "The *read noise* arises in the process of converting a charge generated in the sensor well into voltage and digitization" (Dobrucki, 2013, p. 127).

Each of these name a factor that is operative at a specific point in the data gathering process and the effect it has on a recorded signal. In some cases they also state dependencies between two such factors, as with radiation and energy states in the first quote. Collectively, statements of this form constitute a theory of the data gathering techniques in use.[48]

[48] As noted earlier, this theory may employ vocabulary that is less familiar to scientists working on related topics but which are relegated to "high theory" or in areas that do not share the techniques of this practice. Terms for phenomena like "photobleaching" in microscopy or "jets" in high energy physics primarily gain their meaning from their uses in experimental contexts and sometimes, as in the case of jets, do not have well-defined theoretical meanings independent of them.

I have noted that the decomposition of experimental procedures into relatively discrete activities contributes to the localized character of applied knowledge. The qualifier "pertinent" is applied to the status of theories in empirical programs for one principal reason: in these contexts, theories are not routinely used and developed in the far-reaching, systematic manner of presentation encountered during early stages of scientific training or in the discourse of theoreticians. On the contrary, select theories are incorporated in a piecemeal fashion into an account of the operation of the relevant apparatuses and their interaction with a target system. These theories need only be consulted to the extent that they provide the understanding required to make sense of the data gathering practice and assist in the production of data. The terms employed within theories of technique may thus display many features of patchwork concepts (Wilson, 2006; 2018; Haueis, 2020), the meaning of which are best understood when their use is indexed to specific contextual markers: the object domain and scales at which they are applied, the techniques employed to produce a result, and the domain-specific properties they are used to model. Thus, to draw on a favorite example, the concept "hardness" in reference to metals implies the use of indentation tests to measure the material's response to an applied force, whereas in reference to soft polymers it implies the use of durometers to measure the material's elastic modulus. The significance of concepts employed in a theory of technique may be similarly tailored to the target-apparatus interaction in question. This terminology may be highly specific, as it was with X-ray experimentalists of the early 20th century who coincidentally employed their own notion of "hardness" as a measure of the power of these rays to penetrate solid matter.

Theories of technique can vary widely in explanatory depth with respect to the detailed workings of an apparatus and its interactions. Highly developed forms of experimentation, as seen in X-ray crystallography, NMR, or much of high energy physics, produce an extensive literature on experimental methods and techniques, and foster communities that devote themselves to these specialized mathematical frameworks. It is not uncommon to find students or even research groups in high energy physics that spend much of their time refining models of a single detector component or improving statistical estimates of one of many possible collision decay paths. On the other hand, there are cases like Galileo's telescope or Hacking's (1992) account of the scanning tunneling microscope (STM)—devices that lacked a thorough

explanation of their operations or whose inner workings were typically "black boxed" by practitioners. Yet even these novel devices allow for some understanding of their function. Glass-working and basic optics were widespread enough to familiarize Galileo with the capacity of lenses to alter the appearance of things. Prior to his obtaining one, telescopes had already developed a minor reputation for terrestrial uses, being referred to as "spyglasses" (Van Helden, 2010). Galileo even includes a simple test for determining magnifying power of a scope in his *Starry Messenger*. Similarly, despite the theoretical mysteries surrounding quantum phenomena such as tunneling, Hacking overstates the occultation of STM. The basic functioning of the device was already clearly presented in the inaugural papers of Binning et al. (1982), and today a beginning student of materials science will encounter lecture slides that explain the basic principle of signal generation in STM.

The point stands, however, that neither Galileo nor many practicing nanoscientists could be expected to *derive* the results produced by their device, nor understand its workings in full theoretical detail. More important for their purposes is the possession of an account of its functionality—what it operates on, how it does this, and what data are generated in the process—presented in terms of relatively coarse-grained causal generalizations. A large portion of the texts and courses transmitting this empirico-technical knowledge are devoted to familiarizing practitioners with aspects like these. A sample text on fluorescence microscopy thus includes extensive discussion of what the instrument is well-suited to measure (the presence or absence of labeled molecules; relative concentrations of molecules…), its basic limitations (light diffraction's interference with resolution of sub-cellular objects; gradual reductions in fluorescence due to photobleaching…), and other critical issues affecting performance (stability of the light source; chromatic and spherical aberration; dependence of resolution on the wavelength of light emitted…) (Dobrucki, 2013). Similar focus is applied to the preparation stages and data processing stages of experiment. Texts on electron microscopy emphasize that specimens must have a structure that isn't significantly altered by staining techniques and remains stable under electron bombardment (Michler, 2008). Methods that employ CCD detectors cover artifactual and background effects such as dark current, photon noise, and read noise, all of which contribute to recorded data.

When paired with the proper technical training, a theory of technique provides a researcher with knowledge that renders them a competent producer of data. Much of this is achieved through a firm technical grasp of the basic principles and standard procedures of the practice. The rest is due to the control of sources of systematic uncertainty, i.e., biases in results that arise from non-statistical properties of the experimental setup. Philosophers of experiment have routinely commented on the connection between the reduction of systematic uncertainties and researchers' knowledge of those factors within their experimental setup responsible for them. Mayo (2014) writes that the goal for reliable experiments "is not error avoidance but error control, which may be had by deliberately capitalizing on ways we know we can be wrong" (p. 59). On similar grounds, Boumans and Hon claim that "Understanding an error amounts, inter alia, to uncovering the knowledge generating features of the system involved" (2014, p. 11). It is easy to see why this is: an understanding of the dependency relationships between factors contributing to data, as described in terms of a localized theory of technique—an account of the myriad ways factors in the data gathering setup influence one another and affect data outcomes—gives researchers the tools to eliminate as much bias as possible, either through interventions on the setup itself or in the course of data processing.

With the guidance of a theory of technique, then, researchers are better able to shield off or separate out those features of data that are due to the target system from those artifacts that are not. Once data has been shorn of artifacts (as discussed in Section 6), its remaining features provide a basis for inferring target properties. This is consistent with Woodward's (2000) account of reliability, described in terms of a systematic pattern of counterfactual dependence between data outcomes and claims about a phenomenon of interest. For Woodward, this pattern is assessed in light of scientist's understanding of "a large number of highly specific local empirical facts about the causal characteristics of the detection or measurement process" (p. S170). This understanding, which I have called a theory of technique, allows scientists to work "backward" from data outcomes to claims about the phenomenon generating this data, and so characterize the phenomenon in terms of the empirical effects it produces.[49]

[49] The relationship of these data-phenomenon inferences with modeling claims will be explored further in subsequent chapters.

3.2.3.In situ strategies

In situ strategies comprise the other basic element of empirical programs. These refer to frameworks that allow practitioners to evaluate the merits of implementing a component technique in light of local contingencies. In addition to the principles and generalizations bestowed by pertinent theories, researchers have a range of techniques and materials available for carrying out a given task. A sketch of this set of resources in the case of X-ray diffraction experiments are shown in Figure 5.

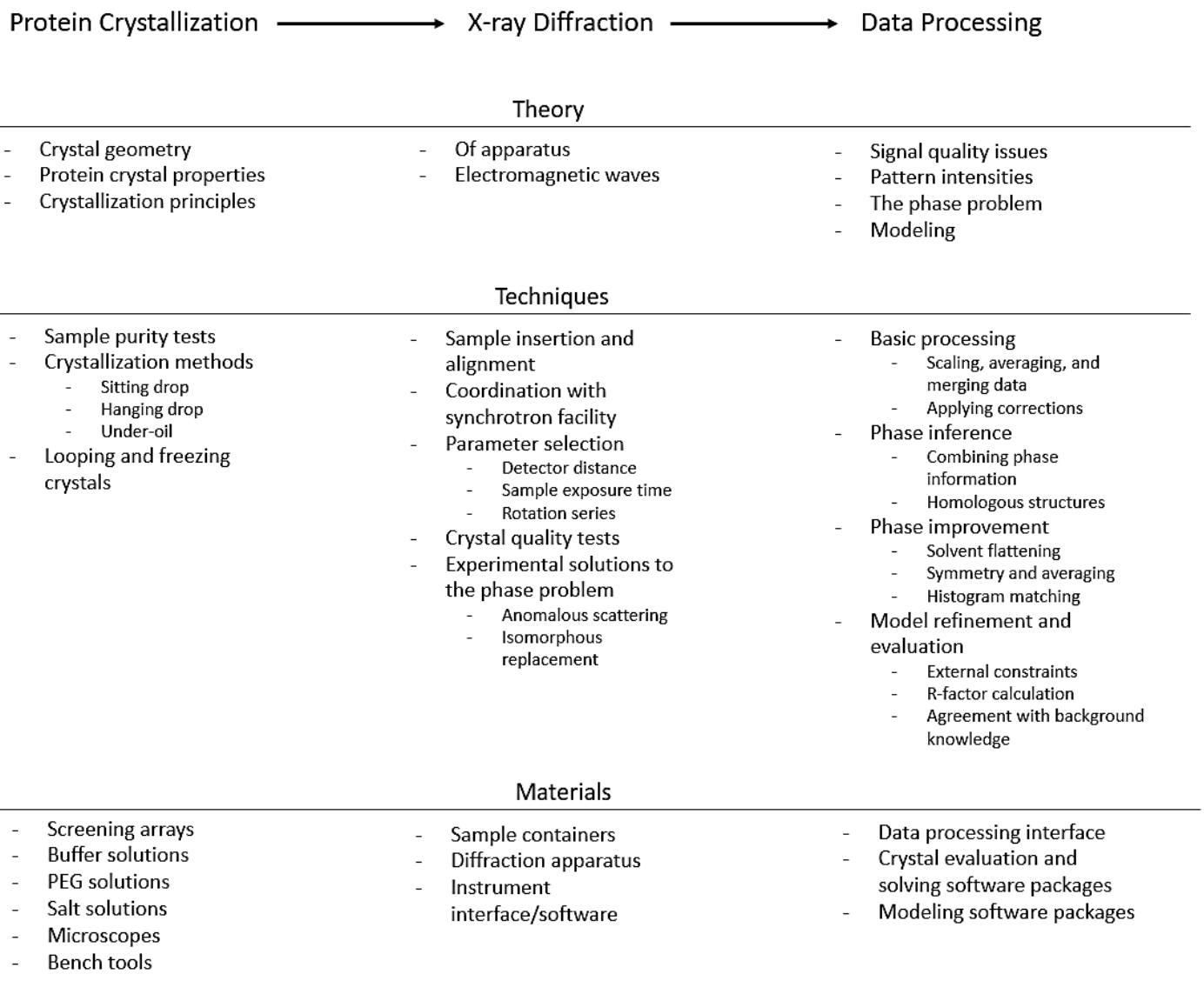


Figure 5: Theory, techniques and materials for each stage.

The use of one technique and set of materials from this table rather than another is rarely a foregone conclusion; such choices depend on researchers' aptitudes and the local contingencies confronting them. The choice of one technique over another typically involves weighing the trade-offs in advantages and disadvantages of each technique. For example, the under-oil crystallization method has the benefit of allowing researchers to vary the mixture of different oils surrounding a drop of protein sample in order to control the speed at which water evaporates from the drop into the air (a process that instigates supersaturation and crystal growth). On the other hand, this evaporation process is continuous and so requires precise timing to stop it once a crystal has emerged. The sitting and hanging drop methods work by equilibration and are not nearly as time-sensitive. A busy researcher with multiple other projects or time spent away from the lab may be safer sacrificing the control of under-oil for the conveniences of the hanging drop approach.

Despite these idiosyncrasies, some overview of the trade-offs associated with different strategies can receive textbook treatment (Fig. 6).

Class	Examples	Brightness	Photo-stability	Bio-compatibility	Environmental sensitivity	Two-photon excitation
AlexaFluor	AlexaFluor 488	××××	××××	××××	××	×××
Atto	Atto 488	××××	×××××	××××	××	×××
BODIPY	BODIPY TMR	×××	××	××××	×××	××
Coumarin	Coumarin 6	×	×	×××	××××	×
Cyanines	Cy5, Cy7	×××	××××	×××	××	×
Fluorescein	FITC	×××	×	×××	××××	××
Rhodamines	Rhodamine 6G	×××	××××	××	×××	×××
Quantum dots	—	×××××	×××××	×	×	×××××
GFP-like proteins	EGFP, EosFP	××	××	×××××	××	××
Phycobiliproteins	R-phycoerythrin	××××	××	××	××	××××

Figure 6: Table of fluorescent markers and properties (Nienhaus & Nienhaus, 2013).

As seen here, quantum dots boast excellent brightness and stability, but are not suitable for unstable environments or biological specimens that do not readily accept foreign markers. Comparable choices are found throughout experimental work in science. For example, estimation of signal background due to Drell-Yan processes in particle collisions can be done with data-driven techniques (the "ABCD method") or other

driven by Monte Carlo simulations (the "Scale-Factor method"). The choice of one rather than another may depend on the quality of a particular region of data under scrutiny. If the quantity of data collected within this region has been insufficient to suppress statistical uncertainties, a simulation-based background estimation may be preferable. Often one choice is not clearly wrong compared to another; it depends on how local contingencies are best navigated in pursuit of a researcher's goals.[50]

3.2.4. Empirical programs in practice: learning in situ strategies and techniques

In order to further illustrate the role of these programs in the execution of a data gathering procedure, I will draw on observations made while embedded in a structural biology laboratory at the University of Pittsburgh. There I was given the opportunity to observe and query Andrew Bellesis throughout the course of sample preparation and an NMR experiment performed in the summer of 2019. At this time Andrew was a graduate student in the Department of Structural Biology, and he had prior training in a biochemistry masters program. Andrew had recently completed the coursework and qualifying examinations in his department and, tasked with developing a PhD project, had begun working to isolate and study a protein complex pertinent to the research interests of the lab.

While shadowing Andrew, I encountered numerous instances where an understanding of the data generating practice was weighed against local contingencies in order to direct a choice of technique. In some cases, Andrew's choice differed from those of more experienced researchers, for example in the choice of a particular pulse program for his NMR experiments. Andrew had decided to use an HSQC (i.e., heteronuclear single quantum coherence) experiment to study his protein sample. This is a very common form of NMR experiment, which meant he could easily borrow parameters from prior experiments saved on the NMR machine's computer and, following an established protocol, get solid data. Unfortunately for

[50] As Wimsatt (1980) has pointed out, this is a basic feature of scientific practice that becomes apparent when scientists are not viewed as embodiments of algorithmic tasks, but as decision-making agents who rely on heuristics for cost-effective solutions to highly specific problems.

Andrew, he was studying Vpr-hHR23a—a complex of two bound proteins—and HSQC data becomes increasingly distorted when used on large macromolecules such as this. As a postdoc in the lab explained to him, HSQC produces an average of four signals (or relaxation peaks) for each nucleus, which can vary significantly in a large molecule due to indirect coupling interactions between the spins of nearby nuclei. An alternative program, TROSY, is better suited to large molecules, since it throws out info from three of these signals and just looks at the sharpest peak. Because the structural biology lab at Pittsburgh is well-funded and possesses the high-strength magnets required for effective use of TROSY, Andrew should have gone with this instead. A second graduate student more familiar with NMR quickly reaffirmed these arguments (Bellesis, 2019).

A similar disagreement arose over settings used for gel electrophoresis. The same postdoc balked at Andrew's intention to run his gel at 200 volts for 30 minutes and insisted he instead use 150V for 70 minutes. Lower voltage gives higher resolution and has less of a chance of "frying the gel"—i.e., overheating it and causing it to lose structural integrity, ruining the data. Andrew later explained to me that he didn't think he was doing anything *wrong* by running it quicker, but would follow this new protocol for the foreseeable future, since the postdoc was more experienced and likely right that it will be more successful more of the time (Bellesis, 2019). For his part, the postdoc claimed that he used to run gels at 230V until a more senior researcher in the lab stopped him for similar reasons.

Where a theory of technique can benefit from systematic presentation in textbooks and seminars, interpersonal encounters in the lab are an ideal medium for the transmission of technical skills and in situ strategies. Learning these strategies often depends on corrective feedback and the communication of past experiences with decisions made in light of local contingencies. This form of knowledge transmission conforms to what Sterelny has called the apprentice learning model. He writes, "most consequential cases of social learning in humans have not depended on pure demonstration or instruction […] Rather, most social learning is hybrid learning: agents acquire skills through socially guided trial and error and socially guided practice" (Sterelny, 2012, p. 28). Andrew's experience learning to use the AKTA purification machine is a case in point. He was trained on the machine by one of the lab's technicians. After several

repeat explanations, he had a grasp of how to use the machine, but it took multiple subsequent runs before he felt confident using it without having others around to help troubleshoot. At first the technician walked him through each step, operating the machine herself. Then he did everything while she watched and corrected him. Finally, he did it himself but made sure a knowledgeable technician was nearby in case he encountered problems.

3.2.5.Empirical programs in practice: warranting actions

Generalizations from a theory of technique may be combined with considerations of local contingencies to inform on site reasoning. These yield chains of inference culminating practical imperatives attuned to a researcher's circumstances. Many instances of this were found in my field work. Here, for example, is the reasoning that motivated Andrew to work quickly between extracting proteins from cells and obtaining a purified sample (Fig. 7):

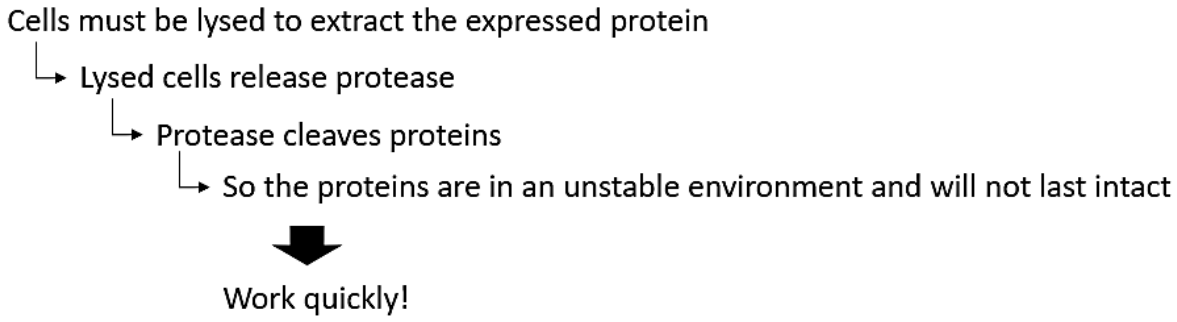


Figure 7: Action-warranting inferential chain, first example

Here is another simple example motivating regular visual inspections of the fluid protein samples, or "fractions," obtained from a purification machine (Fig. 8):

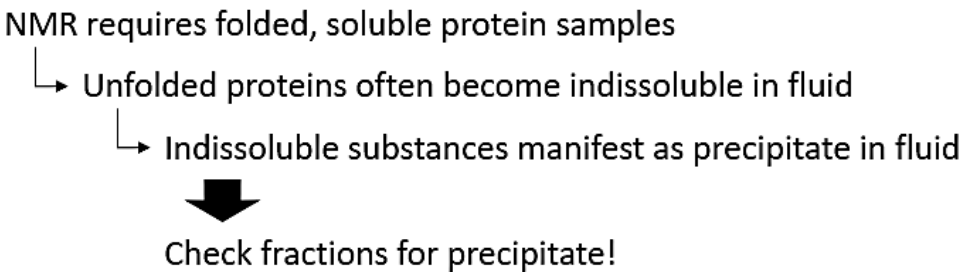


Figure 8: Action-warranting inferential chain, second example

These are not intended as logical reconstructions of researcher behavior. They depict the actual reasoning reported for each act. Nor are they meant to capture the complete set of motives and assumptions at work; the overarching goals of extracting protein and producing clean NMR data are only implicit in the above examples, and the terms and relations are not explicitly defined. Rather than view this reasoning as enthymematic, it can be seen to consist in material inferences. Material inferences are justified on the basis of the meaning of the concepts contained in their premises and conclusion, and are often defeasible, or non-monotonic, in the sense that additional premises can render a practically valid inference invalid.[51]

This framing is more amenable to reasoning of the kind highlighted here than a formal approach, since empirical programs are not learned and applied in ways that honor the strictures of formal logic. They are designed to inform the situated reasoning of practitioners in response to the local conditions and contingencies of empirical work. By articulating how a given technique works, and so establishing the kinds of systems and preparations for which it is an effective means of inquiry, theories of technique delimit the circumstances in which their generalizations hold. Reasoning on their basis thus has a non-monotonic character: change in these background circumstances could render formerly valid inferences invalid. The assumptions involved in such reasoning are geared toward the practical purposes of researchers. Thus, Andrew told me that he takes regular optical density measurements of his growing bacterial cells in order to avoid producing high densities, at which point inter-cellular interactions interfere with protein production, but he was unable to tell me why or how exactly these interactions have this effect. Such

[51] Sellars (1953) originated the theory of material inference, which Brandom (2000) has most notably extended and Brigandt (2010) has previously applied to scientific reasoning.

explanations were seen to extend beyond Andrew's place in the epistemic division of labor. He relegated them to realms of "microbiology" that did not directly impact his work.

This portrayal of practical reasoning in the lab can be expanded to bring out a more complicated ramified structure, in which one imperative entails several procedural sub-imperatives, each with their own line of motivating inferences (Fig. 9).

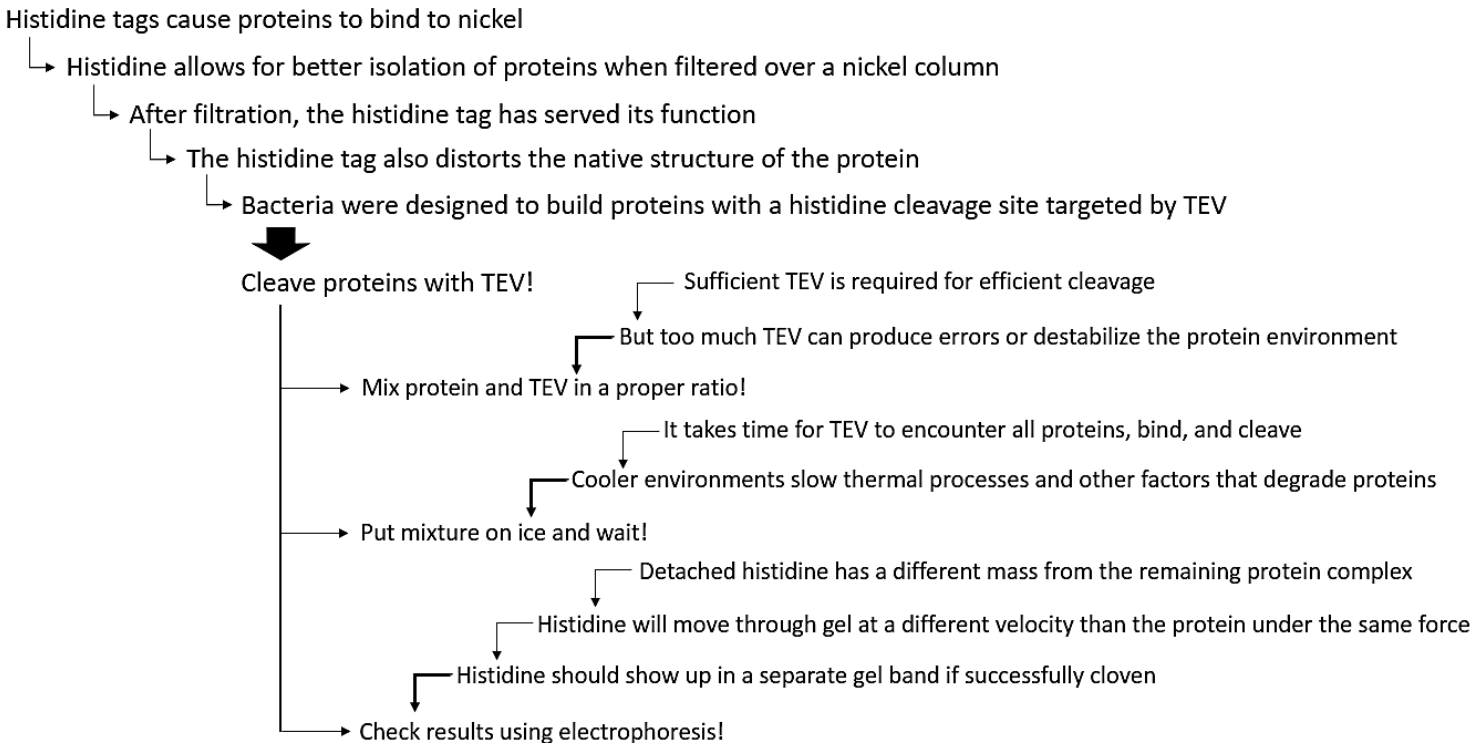


Figure 9: Nested action-warranting inferential chains

In fact, each imperative corresponds to a successive step in a sequence of lab activities. In this way, the inference diagrams put flesh on the procedural skeletons depicted earlier. Each step in a procedure can be reformulated in terms of a practical imperative based on reasoning from a body of empirico-technical knowledge. It is in this sense that the activities constituting a data gathering practice are guided by empirical programs. To the extent that theories of technique account for factors in the data gathering setup responsible for systematic uncertainties, a procedure carried out under their guidance is more likely to produce reliable data.

3.3. Further aspects of practice

Now that I've described the basic features of empirical programs and the way that they demarcate and guide data gathering practices, I want to discuss three more features of these practices that contribute to the consistency and coherence of data: record-keeping, cross-checking, and recursive revision. First, I will recount an episode in Andrew's lab work that will serve to illustrate these features.

The overarching goal of Andrew's lab work was to reproduce NMR results achieved by a former member of Pittsburgh's structural biology lab. The NMR experiments were done on Vpr-hHR23a, an HIV-related protein complex (Jung, et al., 2014). Andrew began work to purify this complex as a project suggested by his advisor. Only later (due, he wagered, to communication issues in the lab and loss of institutional memory) did he learn that similar data had been produced seven years prior. After struggling to get a pure sample of this complex based on the published paper, which lacked a detailed account of the protocol, he went to the lab archives and dug up the technician's notes. There he pinpointed the date and documents where they most likely succeeded with purification. Andrew ran into trouble, however, when the data recorded in his notes at a late stage of his purification process did not match those from the prior experiment (Fig. 10). Rather than two peaks, signifying a separation of free hHR23a from the complex that Andrew sought to isolate, he saw one large peak that overlapped with regions where both smaller peaks were expected.

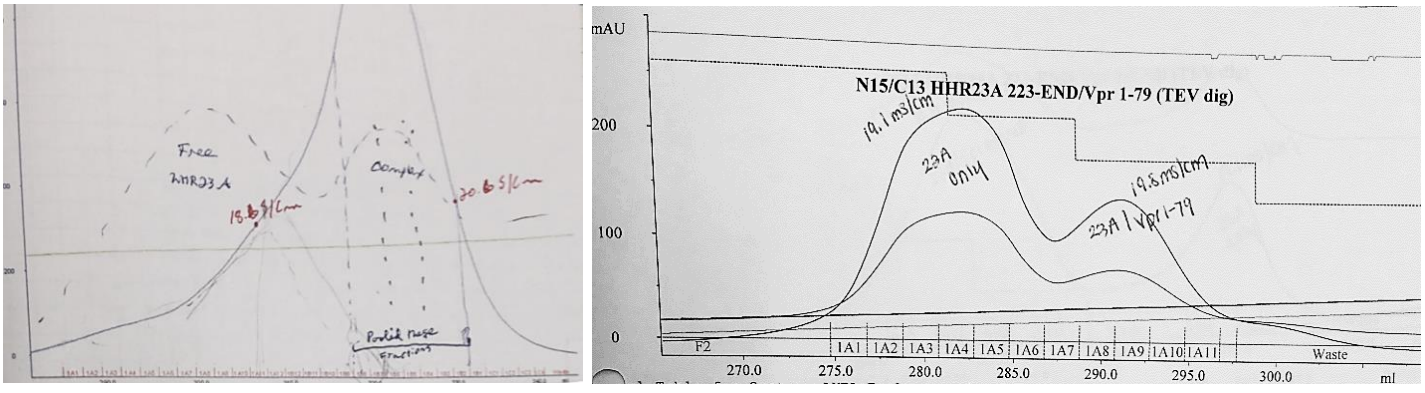


Figure 10: Machine read-outs from Andrew's work (left) did not match those from a prior lab group whose results he sought to replicate (right). For comparison, the prior group's results are sketched in as a dotted line in Andrew's graph.

Bad NMR data ensued, prompting Andrew to consult his lab group. After several proposed explanations for the discrepancy were thrown out, a consensus was reached by review of Andrew's notes that the cause was the substantially higher concentration of protein used in his purification run than in the 2013 experiment. The group reasoned that the filtration machine was overwhelmed by this higher concentration, and that a large quantity of free hHR23a was not successfully separated off from the complex, resulting in a single large peak.

Andrew confirmed this explanation with the help of further record-keeping. He returned to an image in his notes of the gel electrophoresis after this purification run (Fig. 11). The darkness of the lower band relative to the upper was an indication that there was significantly more hHR23a in his sample than Vpr (note: these quantities could be judged independently because proteins are treated with denaturing solution prior to gel electrophoresis, which separates the bonded complex into its components).

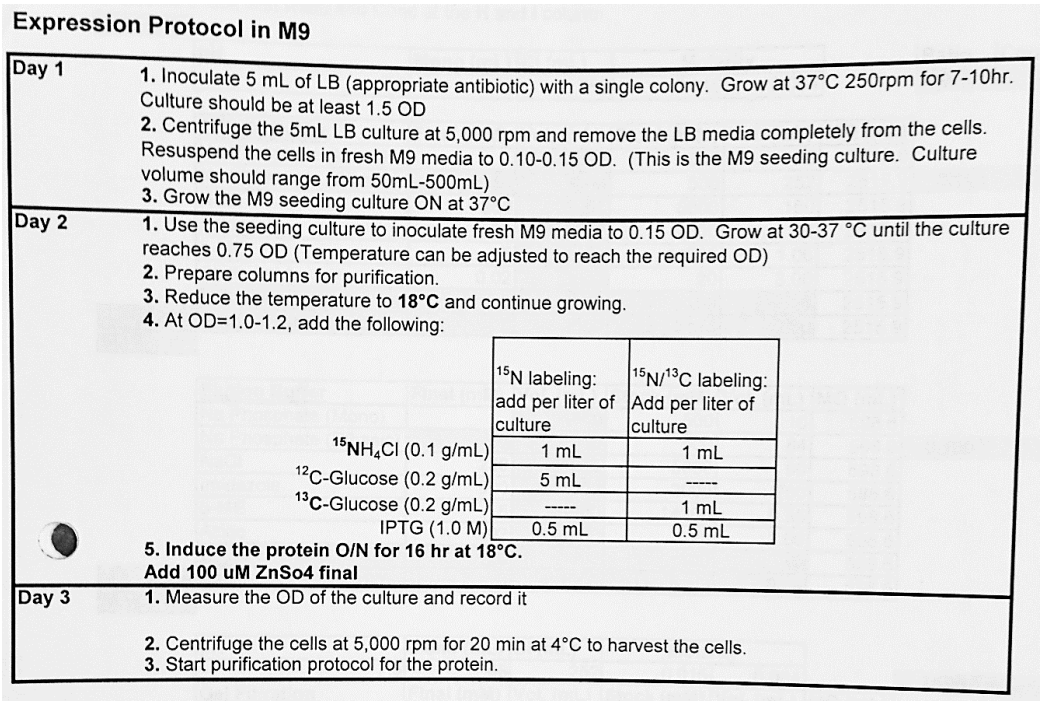
Figure 11: Gel electrophoresis image from Andrew's final purification run.

This was further confirmed by mass spectrometry on left-overs of the purified sample—it detected far more signal at the mass range of hHR23a than Vpr. Finally, and most convincing for Andrew, he took some of an old protein aliquot sample set aside prior to this purification run and ran it through the machine at a lower concentration. The peaks then began to show a slight separation. And so Andrew returned to the first stage of the procedure, intent on running a lower concentration.

3.3.1.Record-keeping

Industrious record-keeping has long been a characteristic of modern science (Daston, 2011; Yeo, 2014). Today it is a methodological rule that researchers save the data recorded from their interventions in some form. This applies not only to the recordings that directly contribute to an empirical result, but also to those auxiliary data generated in the course of a procedure, as seen in the gel images and machine read-outs above. Crystallographers are likewise encouraged to record light microscope images of crystals as they grow in order to determine the time at which they reach an optimum quality. Such recordings provide raw material for downstream decision-making and cross-checking (discussed below) on the way to generating a desired data set.

The production of written records is similarly routine. Science students are trained on maintaining a lab notebook early in their scientific careers, and active laboratories contain storehouses of notes and data from decades-old work. This practice is normalized to such an extent that a failure to produce lab notebooks can cast doubt upon scientific work (Fuyuno, 2005). At their barest, lab notes may state the protocols and recipes followed for a particular procedure (Fig. 12).



Expression Protocol in M9

Day 1

1. Inoculate 5 mL of LB (appropriate antibiotic) with a single colony. Grow at 37°C 250rpm for 7-10hr. Culture should be at least 1.5 OD
2. Centrifuge the 5mL LB culture at 5,000 rpm and remove the LB media completely from the cells. Resuspend the cells in fresh M9 media to 0.10-0.15 OD. (This is the M9 seeding culture. Culture volume should range from 50mL-500mL)
3. Grow the M9 seeding culture ON at 37°C

Day 2

1. Use the seeding culture to inoculate fresh M9 media to 0.15 OD. Grow at 30-37 °C until the culture reaches 0.75 OD (Temperature can be adjusted to reach the required OD)
2. Prepare columns for purification.
3. Reduce the temperature to **18°C** and continue growing.
4. At OD=1.0-1.2, add the following:

	^{15}N labeling: add per liter of culture	^{15}N/^{13}C labeling: Add per liter of culture
^{15}NH$_4$Cl (0.1 g/mL)	1 mL	1 mL
^{12}C-Glucose (0.2 g/mL)	5 mL	-----
^{13}C-Glucose (0.2 g/mL)	-----	1 mL
IPTG (1.0 M)	0.5 mL	0.5 mL

5. Induce the protein O/N for 16 hr at 18°C.
Add 100 uM ZnSo4 final

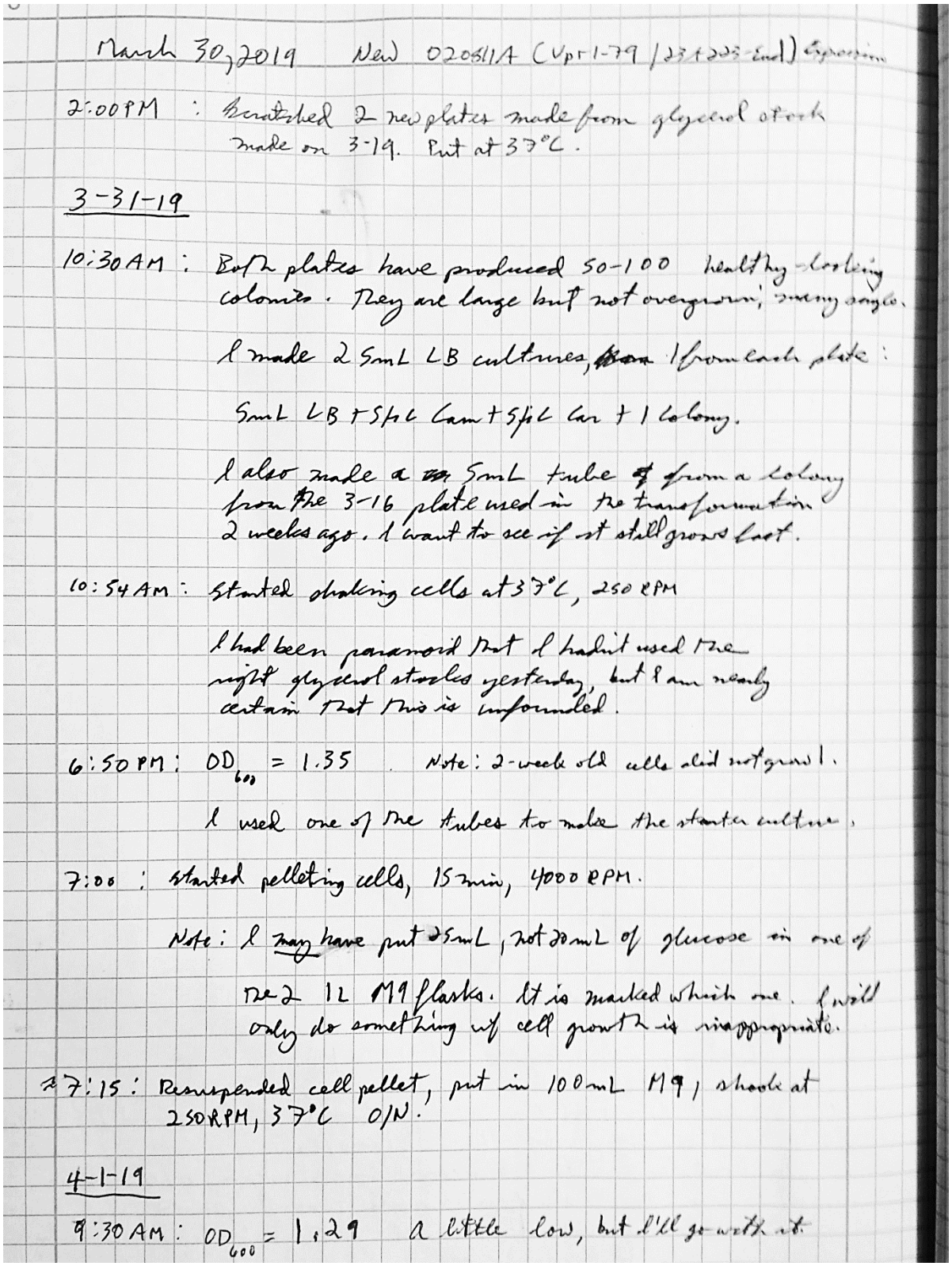
Day 3

1. Measure the OD of the culture and record it

2. Centrifuge the cells at 5,000 rpm for 20 min at 4°C to harvest the cells.
3. Start purification protocol for the protein.

Figure 12: Archived lab notes from the structural biology lab in Pittsburgh.

Consider, by comparison, a page from Andrew's notebook (Fig. 13).



March 30, 2019 New 020611A (Vpr1-79 / 23+223-End) Expression

2:00 PM : Streaked 2 new plates made from glycerol stock made on 3-19. Put at 37°C.

3-31-19

10:30 AM : Both plates have produced 50-100 healthy-looking colonies. They are large but not overgrown; many single.

I made 2 5mL LB cultures, ~~from~~ 1 from each plate:

5mL LB + 5µL Cam + 5µL Car + 1 colony.

I also made a ~~5~~ 5mL tube ~~of~~ from a colony from the 3-16 plate used in the transformation 2 weeks ago. I want to see if it still grows best.

10:54 AM : Started shaking cells at 37°C, 250 RPM

I had been paranoid that I hadn't used the right glycerol stocks yesterday, but I am nearly certain that this is unfounded.

6:50 PM : OD_{600} = 1.35 Note: 2-week old cells did not grow!.

I used one of the tubes to make the starter culture.

7:00 : started pelleting cells, 15 min, 4000 RPM.

Note: I may have put 25mL, not 20mL of glucose in one of the 2 1L M9 flasks. It is marked which one. I will only do something if cell growth is inappropriate.

~7:15 : Resuspended cell pellet, put in 100mL M9, shook at 250RPM, 37°C O/N.

4-1-19

9:30 AM : OD_{600} = 1.29 a little low, but I'll go with it.

Figure 13: A page from Andrew's lab notes.

Each recorded step notes the time it was written, along with relevant quantities (amount of solution; rate, time length, and temperature of automated shaking; rate and time length of centrifugation or "pelleting"). If any ambiguity or uncertainty arises during the process, it is inscribed: at 7:00, "I *may* have put 25mL, not 20mL of glucose in one of the two 1L M9 flasks. It is marked which one. I will only do something if cell growth is inappropriate." The benefits are evident: keeping a concise record of the salient decisions and actions taken during the course of an investigation forces researchers to routinely monitor themselves, providing a safeguard against error.

Alternatively, if an error does take place, written records supply the researcher (or others) with resources to diagnose it and trace its effects on downstream processes. This use of lab notes was manifest in Andrew's case. It allowed his group to identify two important discrepancies between his protocol and the prior experiment's: the purification data and the sample concentration. Drawing on their knowledge of the target-apparatus interaction, they then posited a dependency relation between the two. Andrew gained confidence in this posit by consulting further records and carrying out tests on their basis. He drew on written notes, but also on material records—samples of laboratory materials taken at key junctures in his procedure where the target system underwent alteration (Fig. 14).

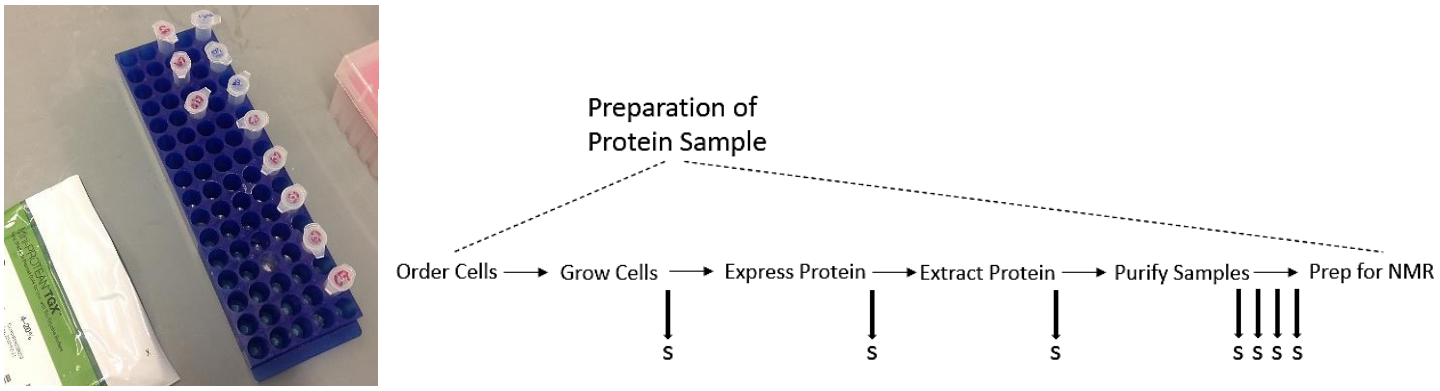


Figure 14: A collection of aliquot samples (left) and the points, marked by an arrow and "s," when a material record was made in Andrew's procedure (right).

The preservation of material records is distinguishable from written note-taking in that it produces items that may serve as a source of new data. Taking regular aliquot samples fine-grains the progressive alterations made to a target in the lab records. It preserves a representative of the target system at separate

stages. This allows researchers to return, in a sense, to preceding stages and investigate factors contributing to a downstream result, as with Andrew's unpurified protein sample. Not all material records serve this specific purpose; researchers involved in fieldwork, such as soil ecologists, will routinely collect samples so that measurements taken in the field may be complemented with laboratory tests. In such a case, the sample functions both as a record and a source for cross-checks.

3.3.2. Cross-checks

Andrew tracked each alteration of his target using various cross-checks. For instance, he applied gel electrophoresis to a set of fractions produced by the purification machine to check whether the molecular contents of an input sample had been successfully separated. This resulted in two forms of data—the machine read-out and the gel image—that could be attributed to a common system, and so checked against one another (Fig. 15).

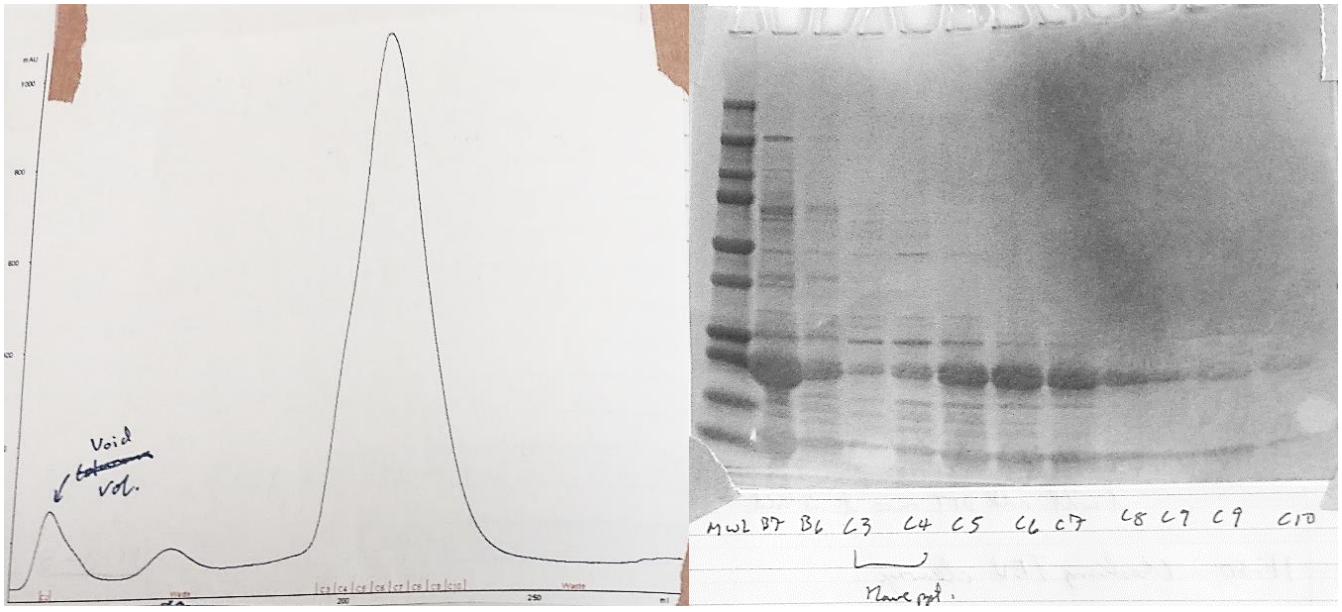


Figure 15: A machine read-out (left) and gel electrophoresis image (right) pertaining to the same fractions.

The signal peaks on the left correspond to an increase in the measured UV absorbance of the material passing through filtration. Organic compounds tend to absorb light from the UV spectrum, especially those

that contain aromatic molecules (as most proteins do). An increased signal thus indicates the presence of more organic material passing through the machine. This material is dispensed into separate vials of equal volume that are then deposited and run side-by-side in gel electrophoresis wells. The gradual darkening and lightening of the gel bands in the vertical columns when viewed from left to right indicates a shift in the concentration of molecules of a given mass (corresponding to each row) during the run time of the filtration process. This is exactly what should be expected from the rise and fall of the UV absorbance peak.

This method of coordination and cross-checking across technical modalities is ubiquitous in empirical practice. It enables evaluation of the quality of an ongoing procedure at multiple stages. When judging the rate of growth of bacterial cells, visual inspection of sample opacity is combined with photospectrometric measurements of optical density. When choosing the best purification fractions to undergo further interventions, electrophoresis gel images are weighed against visual inspection for unwanted precipitate. In evaluating a protein sample for potential crystallization, gel images, Weston blots, and sometimes electron microscope images are consulted. Such methods of cross-checking have been cited by philosophers of science going back at least to Wimsatt (1981) and Hacking (1981) and have been analyzed in terms such as robustness, triangulation, and consilience.[52] Much of the ensuing literature has considered this topic in the context of gaining credence in a hypothesis about some phenomenon on the basis of separate sources of evidence (Fitelson, 2001; Stegenga, 2009; Hey, 2015; Kuorikoski & Marchionni, 2016). Where the tendency among these authors is to regard this method in its application to the *results* of experimentation, here we see the routine occurrence of cross-checks as activities in a data gathering practice. I counted no fewer than seven instances of cross-checking in Andrew's protein preparation. Here robustness was woven into the fabric of experimentation itself (see the following section and Appendix I for a general discussion of the rationale for frequent cross-checking).

Cross-checking is incorporated into a variety of techniques and at multiple scales of scientific inquiry. Some examples include,

[52] They often depend on theories of technique with some independent assumptions and so can be described as instances of "measurement robustness" (Woodward, 2006).

- *Same techniques, same lab, different personnel:* Andrew used the purification machine read-outs and NMR data from a prior experiment to evaluate his success at isolating the Vpr-hRR23a complex.
- *Same techniques, different labs:* the lab at Pittsburgh relied on results from Phil Selenko's lab in Berlin to evaluate their ability to carry out in-cell NMR.
- *Same instrumental techniques, different data processing techniques:* data processing at LIGO was carried out by two different teams who incorporated different assumptions into their techniques (and at times openly fought over them). A result was only accepted when both teams reported an observation of gravitational waves.
- *Different instrumental techniques, different data processing techniques:* the Higgs search was carried out using separate detectors—ATLAS and CMS—which operated according to different theories of technique and so called for different data processing algorithms. For instance, CMS's calorimeter consisted of doped crystals that collected showers of photons emitted by passing particles via Bremmstrahlung, whereas ATLAS sampled the energy of particles using layered columns of excitable argon gas. The former allowed for more accurate energy measurement from the resulting data, while the latter allowed for more accurate momentum measurement. The teams at each detector were (in principle) barred from communicating with each other, and researchers at CERN did not announce an observation of the Higgs boson until both independently reported a positive finding (Hong, 2018).

These disparate cases have a common thread. Each involves the use of separately produced data, subject to an alternate set of biases, as a benchmark for establishing a common result. In the first two replication cases, matching the data published by other groups gave researchers reason to think they had adequate control over the influence of user-specific or lab-specific sources of error in the execution of the same technique. In the latter two cases, this extended to control over biases and variations inherent to different techniques, however they might occur at the level of data processing or production. Each case demonstrates a method

for "calibrating" results produced by one technique against those produced by another technique interacting with the same target under variation of some set of relevant background conditions. This is an indirect method for establishing the reliability of a result, the power of which rests in the fact that it does not depend on a complete accounting of the causal factors contributing to it. That is, convergent cross-checks provide support for the claim that the result cannot be attributed to any of the background conditions that distinguish one technique from the other, and thus the claim that the particular set of causes underlying an individual technique did not introduce significant bias into its result.[53] In this way, cross-checks can compensate for questions about reliability raised by the partiality and gaps within a theory of technique's account of a data-gathering practice.[54]

3.3.3. Recursive revision

What happens, then, when an unexpected discrepancy is found between two sets of results, or between simulated predictions and results from data? Two ATLAS working groups coordinating between Pittsburgh and Marseilles found a substantial gap in detector occupancy measurements (roughly, the number of particle events occurring in the detector) in certain regions. The problem regions were parts of the detector with a less-constrained layout, largely clustered around the end-caps where a large number of cables congregated. Due to potential variability in the detector layout in these regions, it was reasoned that blame for the discrepancy likely fell on the simulation side—a result of inadequate modeling of this region. In particular, the density profile of these regions was not sufficiently sensitive to the material differences therein. Particle events result from the interaction of collision products with materials in the detector, and so a difference in density corresponds to a difference in the number of interactions in a region, and hence a change in occupancy (a measure of detector "hits"). The groups sought to correct these measurements by

[53] The most celebrated version of this form of reasoning is the randomized control trial.
[54] This point is discussed further in Section 5.

more accurately accounting for the broad average differences in densities within these regions. It was further known that, while original design of the inner detector rails of ATLAS consisted of interlocked pieces of metal, the actual design had thicker metal rails and an alternate locking mechanism. The detector model, however, was based on the original design, resulting in an underestimated density profile. With this knowledge on hand, the two groups applied various corrections to the model until the statistical deviance between data and simulation was reduced to levels found within other detector regions (Sapp, 2019).

This is an instance of recursive revision within a data gathering practice: when a result fails to meet researcher expectations, the resources at their disposal—theories of technique, records of prior activity, further cross-checking strategies—are deployed to investigate the local causes responsible. Once researchers gain confidence in an explanation, they then incorporate this into the procedure, reconfiguring their setup to control newly discovered influences and otherwise repeating as before. The process is recursive in the sense that it is normal for data generating procedures to require the repetition of some sub-sequence in the procedure, with some degree of variation.

Recursive revision does not only arise in response to problems; in some cases it is explicitly built into a scientific practice, such as crystal growing. Since there is no strong predictive theory of protein crystallization, researchers resort to a large-scale trial-and-error process, in which protein samples are exposed to a wide range of growth conditions, typically purchased from lab suppliers in the form of "screening arrays" (Fig. 16).

Index™ - HR2-144 Scoring Sheet	Date:	Date:	Date
49. 0.2 M Calcium chloride dihydrate, 0.1 M BIS-TRIS pH 6.5, 45% v/v (+/-)-2-Methyl-2,4-pentanediol			
50. 0.2 M Ammonium acetate, 0.1 M BIS-TRIS pH 5.5, 45% v/v (+/-)-2-Methyl-2,4-pentanediol			
51. 0.2 M Ammonium acetate, 0.1 M BIS-TRIS pH 6.5, 45% v/v (+/-)-2-Methyl-2,4-pentanediol			
52. 0.2 M Ammonium acetate, 0.1 M HEPES pH 7.5, 45% v/v (+/-)-2-Methyl-2,4-pentanediol			
53. 0.2 M Ammonium acetate, 0.1 M Tris pH 8.5, 45% v/v (+/-)-2-Methyl-2,4-pentanediol			
54. 0.05 M Calcium chloride dihydrate, 0.1 M BIS-TRIS pH 6.5, 30% v/v Polyethylene glycol monomethyl ether 550			
55. 0.05 M Magnesium chloride hexahydrate, 0.1 M HEPES pH 7.5, 30% v/v Polyethylene glycol monomethyl ether 550			
56. 0.2 M Potassium chloride, 0.05 M HEPES pH 7.5, 35% v/v Pentaerythritol propoxylate (5/4 PO/OH)			
57. 0.05 M Ammonium sulfate, 0.05 M BIS-TRIS pH 6.5, 30% v/v Pentaerythritol ethoxylate (15/4 EO/OH)			
58. 0.1 M BIS-TRIS pH 6.5, 45% v/v Polypropylene glycol P 400			
59. 0.02 M Magnesium chloride hexahydrate, 0.1 M HEPES pH 7.5, 22% w/v Poly(acrylic acid sodium salt) 5,100			
60. 0.01 M Cobalt(II) chloride hexahydrate, 0.1 M Tris pH 8.5, 20% w/v Polyvinylpyrrolidone K 15			
61. 0.2 M L-Proline, 0.1 M HEPES pH 7.5, 10% w/v Polyethylene glycol 3,350			
62. 0.2 M Trimethylamine N-oxide dihydrate, 0.1 M Tris pH 8.5, 20% w/v Polyethylene glycol monomethyl ether 2,000			
63. 5% v/v Tacsimate™ pH 7.0, 0.1 M HEPES pH 7.0, 10% w/v Polyethylene glycol monomethyl ether 5,000			
64. 0.005 M Cobalt(II) chloride hexahydrate, 0.005 M Nickel(II) chloride hexahydrate, 0.005 M Cadmium chloride hydrate, 0.005 M Magnesium chloride hexahydrate, 0.1 M HEPES pH 7.5, 12% w/v Polyethylene glycol 3,350			
65. 0.1 M Ammonium acetate, 0.1 M BIS-TRIS pH 5.5, 17% w/v Polyethylene glycol 10,000			
66. 0.2 M Ammonium sulfate, 0.1 M BIS-TRIS pH 5.5, 25% w/v Polyethylene glycol 3,350			
67. 0.2 M Ammonium sulfate, 0.1 M BIS-TRIS pH 6.5, 25% w/v Polyethylene glycol 3,350			
68. 0.2 M Ammonium sulfate, 0.1 M HEPES pH 7.5, 25% w/v Polyethylene glycol 3,350			
69. 0.2 M Ammonium sulfate, 0.1 M Tris pH 8.5, 25% w/v Polyethylene glycol 3,350			
70. 0.2 M Sodium chloride, 0.1 M BIS-TRIS pH 5.5, 25% w/v Polyethylene glycol 3,350			
71. 0.2 M Sodium chloride, 0.1 M BIS-TRIS pH 6.5, 25% w/v Polyethylene glycol 3,350			
72. 0.2 M Sodium chloride, 0.1 M HEPES pH 7.5, 25% w/v Polyethylene glycol 3,350			
73. 0.2 M Sodium chloride, 0.1 M Tris pH 8.5, 25% w/v Polyethylene glycol 3,350			
74. 0.2 M Lithium sulfate monohydrate, 0.1 M BIS-TRIS pH 5.5, 25% w/v Polyethylene glycol 3,350			
75. 0.2 M Lithium sulfate monohydrate, 0.1 M BIS-TRIS pH 6.5, 25% w/v Polyethylene glycol 3,350			
76. 0.2 M Lithium sulfate monohydrate, 0.1 M HEPES pH 7.5, 25% w/v Polyethylene glycol 3,350			
77. 0.2 M Lithium sulfate monohydrate, 0.1 M Tris pH 8.5, 25% w/v Polyethylene glycol 3,350			
78. 0.2 M Ammonium acetate, 0.1 M BIS-TRIS pH 5.5, 25% w/v Polyethylene glycol 3,350			
79. 0.2 M Ammonium acetate, 0.1 M BIS-TRIS pH 6.5, 25% w/v Polyethylene glycol 3,350			
80. 0.2 M Ammonium acetate, 0.1 M HEPES pH 7.5, 25% w/v Polyethylene glycol 3,350			
81. 0.2 M Ammonium acetate, 0.1 M Tris pH 8.5, 25% w/v Polyethylene glycol 3,350			
82. 0.2 M Magnesium chloride hexahydrate, 0.1 M BIS-TRIS pH 5.5, 25% w/v Polyethylene glycol 3,350			
83. 0.2 M Magnesium chloride hexahydrate, 0.1 M BIS-TRIS pH 6.5, 25% w/v Polyethylene glycol 3,350			
84. 0.2 M Magnesium chloride hexahydrate, 0.1 M HEPES pH 7.5, 25% w/v Polyethylene glycol 3,350			
85. 0.2 M Magnesium chloride hexahydrate, 0.1 M Tris pH 8.5, 25% w/v Polyethylene glycol 3,350			
86. 0.2 M Potassium sodium tartrate tetrahydrate, 20% w/v Polyethylene glycol 3,350			
87. 0.2 M Sodium malonate pH 7.0, 20% w/v Polyethylene glycol 3,350			
88. 0.2 M Ammonium citrate tribasic pH 7.0, 20% w/v Polyethylene glycol 3,350			
89. 0.1 M Succinic acid pH 7.0, 15% w/v Polyethylene glycol 3,350			

Figure 16: A collection of screening arrays (left). Each well contains different conditions as listed (right).

Each well is examined under a light microscope, and those conditions that show promising growth in initial trials are then selected and run through a second set of screening arrays, where component parameters are systematically varied within a limited range. This process is repeated until the researcher arrives at crystals that appear to be of optimal quality.

Andrew's experience recounted above is also a clear case of recursive revision, and it serves to demonstrate how the different activities highlighted here are integrated in a data gathering practice. In Andrew's case, the problem was identified through cross-checking: his purification read-out did not match the data produced by the prior group. The ability to perform this kind of cross-check relied on records of both his and the prior group's work. So did the identification of the cause: examining his lab notes and reasoning from their technical knowledge, Andrew's lab group posited a dependency between the concentration of protein fed into the purification machine and his results. After further cross-checking this

new causal relation with the help of material records, Andrew incorporated it into the empirical program guiding his lab work, effectively generating a new practical imperative (Fig. 17):

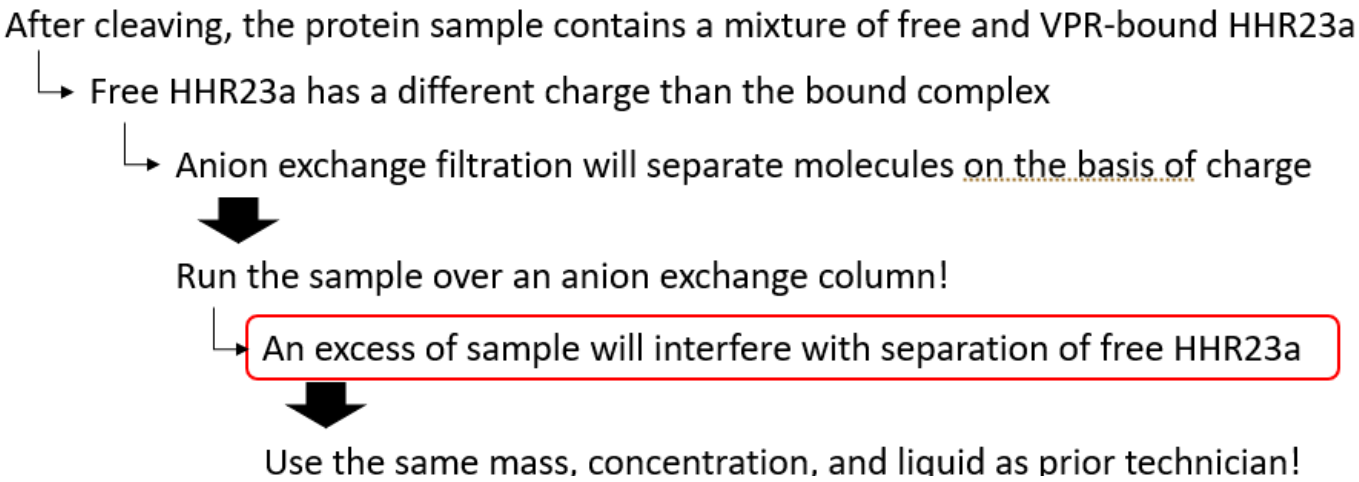


Figure 17: Action-warranting inferential chain, revised

As the PI in Pittsburgh's structural biology lab put it, "95% of what we do is failure. You have to accept that and learn that every time something doesn't work, it's telling you something. The system is telling you something; you have to find out what that is." Indeed, it is through the production of mistakes that scientists are forced to reexamine their understanding of the techniques they employ and uncover previously unacknowledged causes in the data gathering setup. The result of this process is a corrective revision and expansion of the theory of technique in a way that enables researchers to better account for data, and so produce it reliably. The practice of recursive revision shows how error in scientific practice is not merely an obstacle or point of failure, but also plays a positive role in the production of knowledge (Cf. (Karstens, 2014)).

3.4. General remarks on the epistemology and pragmatics of experimentation

Earlier I spoke of the fact that cross-checking, as a form of methodological and metrological triangulation, does not solely occur with respect to the final results of experiments, but is "woven into the

fabric of experimentation itself" (section 4.2 above). I now present a more general account of the epistemological value of these components of intra-experimental practice and the way they figure into researcher decision-making with respect to the goal of producing consistent and coherent data.[55] For the sake of conserving space, I will leave the pragmatic considerations to Appendix I and consider here the epistemic benefit of these practices.

Consider a researcher working, via operations carried out on a target system, from an initial state to an end state. Suppose that there is some error associated with each operation such that there is some uncertainty whether it carries the experimental arrangement from the initial state S to intermediate state I_1 or I_2, and similarly for subsequent steps (Fig. 18).

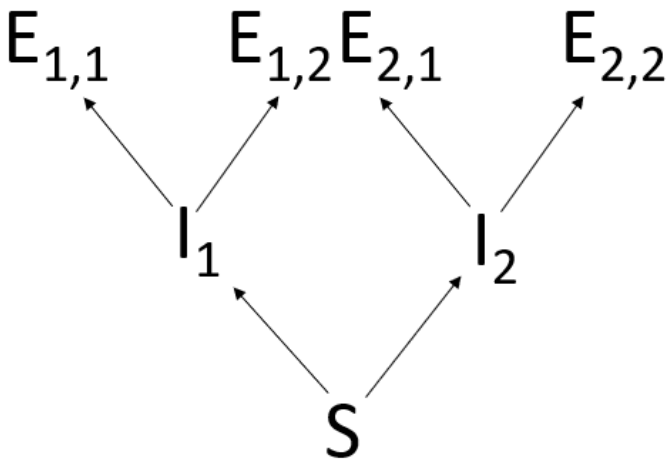


Figure 18: A branching tree structure for operations with some error

I will demonstrate the epistemic value of records and cross-checks in this scenario. We can begin by asking why waiting to perform some measurement at the final stage of this process might not be satisfactory. That is, why bother performing measurements and keeping records prior to this point when the information that one wants is presumably based on the experiment at its end state? One reason for this is to provide ex post facto justification: a colleague who aims to make use of data generated by my phenomenon, which is alleged

[55] Gal Ben Porath deserves partial credit for this section and the appendix, which emerged directly out of work and conversations with him.

to be in some final state (a purified sample, say), may seek some assurance that the system really was in that state. Justifying this claim to the colleague would involve presenting records of the operations carried out. These would provide reasons to believe the experiment was carried along a particular branch of the tree structure.

But this points to a deeper reason for interim checks: recall that a conception of the system interacting with an experimental apparatus is one component of a theory of technique. This entails that a data gathering technique carried out at a later step in a procedure always involves some background assumptions about the current condition of the system, among other factors. The data produced through this technique may not themselves provide a test of all such background assumptions. This is illustrated by cases such as perytons in astronomy (Petroff et al. 2015) or the microtrabecular lattice (Powell 2005)—entities hypothesized to exist due to features that were repeatedly produced in data, which researchers spent years discussing before they were definitively proven to be artifacts. Available knowledge of the techniques involved was insufficient to determine the source of such data patterns. That is, researchers were unable to distinguish between an experimental setup in which the source of the data was an aspect of the target phenomenon and a setup in which it was external to it.

Cross-checks may be beneficial in such circumstances where a particular theory of technique is unreliable. Following Figure 18, suppose that a measurement taken at the final state of an experimental arrangement can distinguish between states $E_{1,1}$ and $E_{2,2}$ but not between states $E_{1,2}$ and $E_{2,1}$. For example, the first pair may include one case $E_{1,1}$ where a target system is fully intact and another $E_{2,2}$ where it has been torn apart by prior operations, while the second pair generate a similar sub-pattern that could be due to causes either internal to the target ($E_{1,2}$, say) or external ($E_{2,1}$) (that is, imagine this difference is only detectable at a scale below the resolution of the data produced by this technique). Suppose, further, that a measurement taken at an intermediate step allows a researcher to determine whether the experiment is in state I_1 or I_2. In this case, having results from the intermediate measurement would allow the researcher to determine the state of their experimental setup, despite generating data at the final step that cannot distinguish between $E_{1,2}$ and $E_{2,1}$. They need only consult whether they were previously at I_1 or I_2 to

eliminate one of these possibilities. Thus, performing checks and keeping records improves the epistemic situation of the researcher by allowing them to locate themselves, as it were, within the branching tree structure of an experiment carried out under conditions of uncertainty. A researcher employing checks is better able to stay "on track" by ensuring their location and determining points at which they must revise their prior operations if the experiment has been carried into an undesired state. Under the assumption that different experimental states cause dissimilarities in data, this improves the consistency of the data generated at the end of the experiment. That is, increased monitoring and revision of their activities increases the probability that a researcher will reduce the tendency to be carried to a state far from the one intended. It also ensures that the data are compatible with the background assumptions accompanying this final stage of data production by giving the researcher a more precise understanding of the state of their experimental arrangement. Thus record-keeping, cross-checking, and revision improve the coherence of data as well.

3.5. Data processing

Empirical programs enable researchers to catalogue factors in the data gathering setup that influence the recording of data, theorize the dependency relationships between these factors and the resulting data, and strategically respond to local contingencies in a manner that is best adapted to the task at hand. Program-guided procedures are further conjoined to activities of record-keeping, cross-checking, and revision, through which their constituent activities are routinely supervised and verified to conform with expectations in a way that tracks and compensates for possible sources of error neglected by theories of technique. If these expectations are violated, records and cross-checks provide the resources to narrow down the stage in a procedure at which an error could have taken place, pinpoint further dependencies between local factors, and revise a theory of technique accordingly. Guided by this knowledge, researchers have grounds for believing that as many sources of uncertainty as are known are being controlled or

suppressed as much as possible. Coherence is a product of this agreement between data and researchers' prior understanding of the target phenomenon, its behavior under preparation, and the target-apparatus interaction. Insofar as data cohere with researchers' knowledge of the data generating context and additional information provided through the variety of available cross-checks, there are good grounds for treating them as produced by the target phenomenon theorized by this empirical program.

Those causal factors contributing to data that cannot be fully controlled in the setup itself can often be accounted for in the data processing stage of a procedure. Researchers develop technique-specific means of intervening on data, again informed by the empirical program, which allow them to extract patterns that can be suitably compared to models. Researchers employ various techniques, often based on a root distinction between a true signal and signals that are due to noise or artifacts. The former are data that are taken to be generated by the target phenomenon. "Noise" and "artifact" refer to data that can be attributed to either signals generated by the operating of the experimental apparatus, the specially prepared state and environment the target is in when the instruments interact with it, or the particular way the signals are recorded after being generated.

This leads to a form of data-reducing intervention in which researchers identify and remove features of data that are produced by the mere operating of instruments, recording of data, or the preparation conditions rather than interaction with the target itself. In X-ray crystallography, the techniques of negative density truncation and solvent flattening provide an example (Drenth, 2007). Often the initial electron density distribution constructed from diffraction pattern data contains regions of negative density. Basic physics, however, denies this possibility; this physical quantity cannot be less than zero, so researchers remove the negative regions of density distribution (Fig. 19).

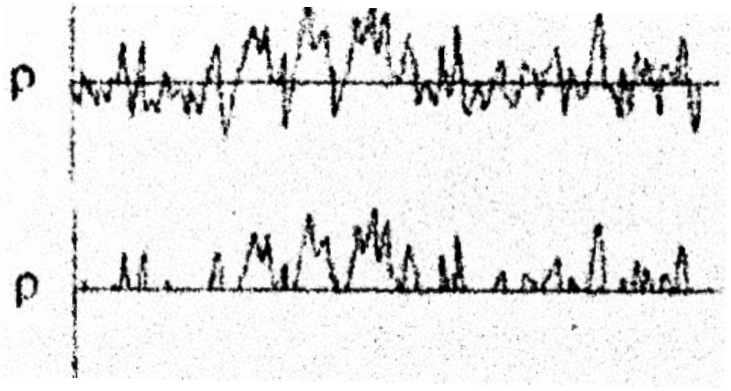


Figure 19: Negative density truncation.

Researchers average the remaining signal and partition the result into regions where the average density falls below or above a certain cut-off. Drawing on their understanding of the chemical properties of the liquid solvent that was used in the crystallization process, they reason that the regions of low-density signal were not generated by the crystallized protein but rather by its surrounding solvent. These regions are therefore eliminated, the averaging effects are reversed, and the negative densities are removed once more (Fig. 20).

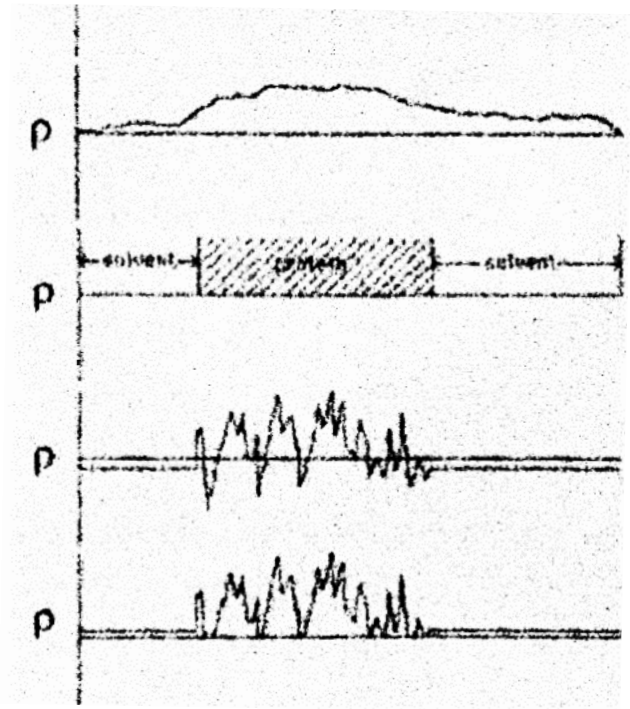


Figure 20: Solvent flattening.

After several iterations, this process can result in dramatic differences between the density map constructed from the initial data and its final form (top and bottom of Fig. 21).

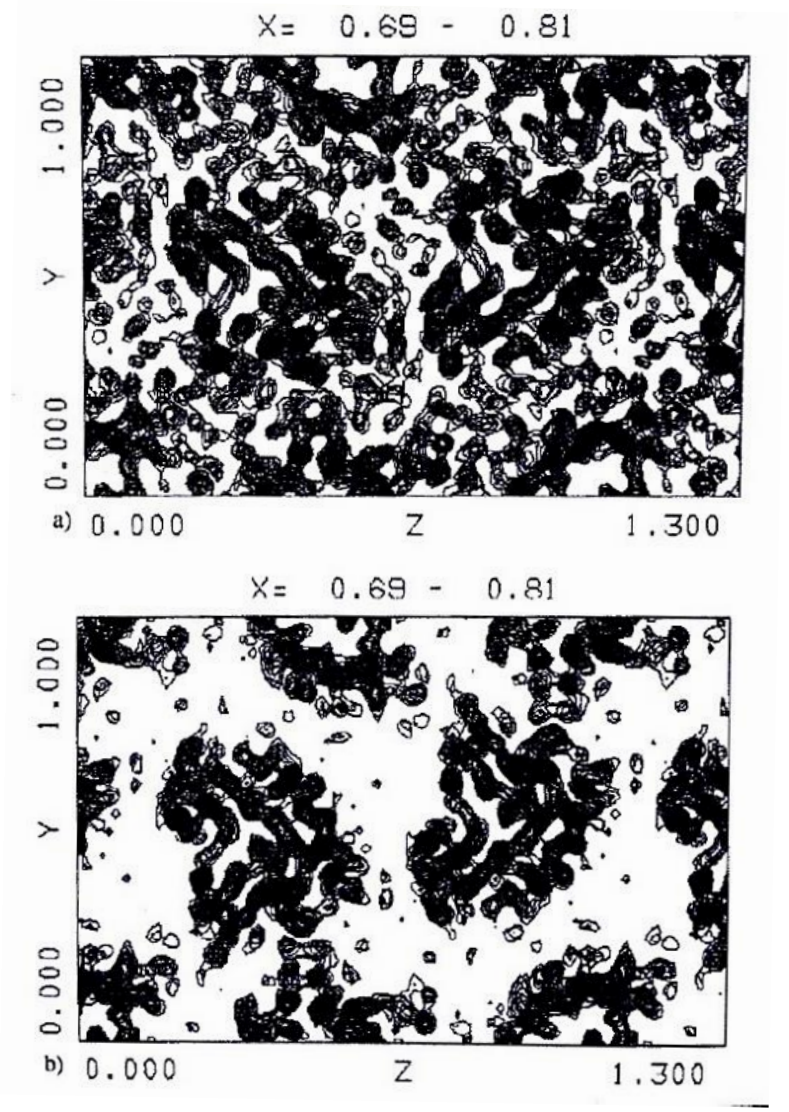


Figure 21: Density maps before (above) and after (below) truncation and flattening.

Each of these interventions on data is motivated by researchers' understanding of data gathering procedures. In the case of negative density truncation, widely accepted and well-confirmed physical principles ruling out the existence of negative densities would have to be abandoned for researchers to accept the negative regions as indicative of an actual feature of the system. Confidence in these principles warrants the removal of such regions. The confidence is buttressed by an alternate explanation of the existence of negative density regions: they are attributed to artifacts resulting from Fourier transforms

applied to inaccurate phase estimates for the recorded signal. As for solvent flattening, an understanding of the specific preparation conditions that apply to the system of interest—knowledge that the proteins are crystallized in solvent—permits researchers to differentiate parts of the data due to the protein itself from those that result from its environment when interacting with the experimental apparatus. Again, this alteration of data is warranted by confidence in a general theory and an alternate causal explanation of the generation of some portion of the recorded data. Here the general theory of protein structure predicts that their electron cloud densities should be concentrated in certain regions and not broadly distributed, as in the top of Figure 21. The recording of X-ray intensities suggesting a broad distribution is then attributed to the interaction of X-rays with the solvent and their scattering off liquid molecules rather than the system of interest.

In this way, we can see the involvement of empirical programs in the processing of data. Features of data are attributed to the specific processing interventions made on them (such as the negative regions resulting from Fourier transforms) or the effects of factors conditioning the system-apparatus interaction (as with the liquid solvent). Subsequent interventions on data are thus warranted by empirical programs in a similar manner to the way they warrant actions resulting from practical imperatives. Because of this warrant, data processed in this way are resolved into patterns that allow researchers to confidently discriminate between inferences drawn from competing models. They thereby satisfy the requirement that data be comparable in order to be reliable.

3.6. Where is the bottom?

This account was geared to show the role of experimental knowledge in producing the kind of data required for the representational use of theoretical models. But it can be charged that I've drastically complicated the representation issue in the process, and for several reasons. First, rather than a simple relationship between one model, one thing in the world, and one data set, I've now presented a framework

that confronts us with multiple, coordinated forms of data production with a collection of theories and techniques corresponding to each sub-step. To the extent that modeling is employed in this process—and work by Edwards (1999), Tal (2012) and Bokulich (2018) among others suggests that this is routine in many areas—this doesn't satisfactorily resolve the general philosophical problem of scientific representation. The relationship between model and target is now ramified into a layered cake of representation problems. It would seem, then, that I haven't explained representation until I can account for these lower-scale instances.

This is an issue that I plan to take up in the remainder of this dissertation. To start, I will approach it by dealing with a smaller, but related issue: In accounting for the reliability of data, I have appealed to techniques and theories trusted by researchers, including causal generalizations about the conditions under which data is generated. I have given reasons to think that this knowledge, combined with routine practices, allow for the control and reduction of uncertainty in outcomes. I have not, however, given any justification for the trust researchers have in these techniques. A bare appeal to "what scientists have been trained to do and believe" is insufficient for this task. We need to understand how these become the things they are trained to do and believe and to understand the justification for this. The following chapter aims to address this by giving a historical account of the way a particular mode of experimental inquiry can take form. I seek to develop an account of the historical dimension of empirical programs, and the justifying role this plays with respect to an ongoing practice. This is intended to provide an additional step towards understanding how models are tied to empirical practice.

References

Anstey, P. (2014). Philosophy of Experiment in Early Modern England: The Case of Bacon, Boyle, and Hooke. *Early Science and Medicine, 19,* 103-132.

Austin, J. L. (1962). *How to Do Things With Words.* London: Oxford University Press.

Baird, D., & Faust, T. (1990). Scientific Instruments, Scientific Progress and the Cyclotron. *The British Journal for the Philosophy of Science, 41*, 147-175.

Bellesis, A. (2019). *Shadowing.* University of Pittsburgh Structural Biology Laboratory.

Binning, G., Rohrer, H., Gerber, C., & Weibel, E. (1982). Surface Studies by Scanning Tunneling Microscopy. *Physical Review Letters, 59*, 57-61.

Bogen, J., & Woodward, J. (1992). Observations, theories, and the evolution of the human spirit. *Philosophy of Science, 59*, 590-611.

Bokulich, A. (2018). Using models to correct data: paleodiversity and the fossil record. *Synthese, Special Issue: Abstraction and Idealization in Scientific Modeling*, Accessed online: <https://link.springer.com/article/10.1007/s11229-018-1820-x>.

Boumans, M., & Hon, G. (2014). Introduction. In M. Boumans, G. Hon, & A. Petersen (Eds.), *Error and Uncertainty in Scientific Practice* (pp. 1-12). London: Pickering & Chatto.

Bourdieu, P. (1977). *Outline of a Theory of Practice.* Cambridge MA: Cambridge University Press.

Brandom, R. (2000). *Articulating Reasons: An Introduction to Inferentialism.* Cambridge MA: Harvard University Press.

Brigandt, I. (2010). Scientific Reasoning is Material Inference: Combining Confirmation, Discovery, and Explanation. *International Studies in the Philosophy of Science, 24*, 31-43.

Chang, H. (1995). Circularity and Reliability in Measurement. *Perspectives on Science, 3*, 153-172.

Chang, H. (2004). *Inventing Temperature.* New York: Oxford University Press.

Chang, H. (2012). *Is Water H2O?* New York: Springer.

Collins, H. (2010). *Tacit and Explicit Knowledge.* Chicago: University of Chicago Press.

Daston, L. (2011). Empires of Observation: 1600-1800. In L. Daston, & E. Lunbeck (Eds.), *Histories of Scientific Observation.* Chicago: University of Chicago Press.

Daston, L., & Galison, P. (2007). *Objectivity.* Cambridge MA: MIT Press.

Dobrucki, J. (2013). Fluorescence Microscopy. In U. Kubitscheck (Ed.), *Fluorescence Microscopy: From Principles to Biological Applications* (pp. 97-142). Singapore: Wiley-Blackwell.

Drenth, J. (2007). *Principles of Protein X-ray Crystallography.* New York: Springer.

Dreyfus, H. (1991). *Being-in-the-World: A Commentary on Heidegger's Being and Time, Division I.* Cambridge MA: MIT Press.

Edwards, P. (1999). Global Climate Science, Uncertainty and Politics: Data-laden Models, Model-filtered Data. *Science as Culture, 8*, 437-472.

Fitelson, B. (2001). A Bayesian Account of Independent Evidence with Applications. *Proceedings of the Philosophy of Science Association*, S123-140.

Franklin, A. (1986). *The Neglect of Experiment.* New York: Cambridge University Press.

Fuyuno, I. (2005). Lack of lab notes cast doubt on RNA researchers' results. *Nature, 437*, 461.

Galison, P. (1987). *How Experiments End.* Chicago: University of Chicago Press.

Galison, P. (1997). *Image and Logic: A Material Culture of Microphysics.* Chicago: University of Chicago Press.

Gangloff, J., & Allamel-Raffin, C. (2014). Commentary: Some Notions of Action. In L. Soler, S. Zwart, M. Lynch, & V. Israel-Jost (Eds.), *Science After the Practice Turn in the Philosophy, History, and Social Studies of Science* (pp. 58-66). New York: Routledge.

Giddens, A. (1984). *The Constitution of Society.* Berkeley: University of California Press.

Gooding, D. (1992). Putting Agency Back into Experiment. In A. Pickering (Ed.), *Science as Practice and Culture* (pp. 65-112). Chicago: University of Chicago Press.

Hacking, I. (1981). Do We See Through a Microscope? *Pacific Philosophy Quarterly, 62*, 305-322.

Hacking, I. (1983). *Representing and Intervening.* New York: Cambridge University Press.

Hacking, I. (1992). The Self-Vindication of the Laboratory Sciences. In A. Pickering (Ed.), *Science as Practice and Culture* (pp. 29-64). Chicago: University of Chicago Press.

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

Harré, R. (2003). The Materiality of Instruments in a Metaphysics for Experiments. In H. Radder (Ed.), *The Philosophy of Scientific Experimentation* (pp. 19-38). Pittsburgh: University of Pittsburgh Press.

Haueis, P. (2020). A Generalized Patchwork Approach to Scientific Concepts. *University of Pittsburgh Center for Philosophy of Science Talk, February 11*.

Heidegger, M. (1963). *Being and Time.* New York: Harper & Row.

Herschel, J. F. (1831). *A Preliminary Discourse on the Study of Natural Philosophy.* London: Longman, Reese, Orme, Brown, and Green.

Hey, S. (2015). Robust and Discordant Evidence: Methodological Lessons from Clinical Research. *Philosophy of Science, 82*, 55-75.

Hong, T. M. (2018). *Topics in LHC Physics.* Seminar: University of Pittsburgh.

Jung, J., Byeon, I., DeLucia, M., Koharudin, L., Ahn, J., & Gronenborn, A. (2014). Binding of HIV-1 Vpr Protein to the Human Homolog of the Yeast DNA Repair Protein RAD23 (hHR23a) Requires Its Xeroderma Pigmentosum Complementation Group C Binding (XPCB) Domain as Well as the Ubiquitin-associated 2 (UBA2) Domain. *The Journal of Biological Chemistry, 289*, 2577-2588.

Karstens, B. (2014). The Lack of a Satisfactory Conceptualization of the Notion of Error in the Historiography of Science: Two Main Approaches and Shortcomings. In M. Boumans, G. Hon, &

A. Petersen (Eds.), *Error and Uncertainty in Scientific Practice* (pp. 13-38). London: Pickering & Chatto.

Kripke, S. (1982). *Wittgenstein on Rules and Private Language.* Cambridge MA: Harvard University Press.

Kuhn, T. (1962). *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

Kuorikoski, J., & Marchionni, C. (2016). Evidential Diversity and the Triangulation of Phenomena. *Philosophy of Science, 83*, 227-247.

Land, W. M., Volchenkov, D., Blasing, E., & Schack, T. (2013). From action representation to action execution: exploring the links between cognitive and biomechanical levels of motor control. *Frontiers in Computational Neuroscience, 7*, 1-14.

Lange, R. (2003). Technology as Basis and Object of Experimental Practices. In H. Radder (Ed.), *The Philosophy of Scientific Experimentation* (pp. 119-137). Pittsburgh: University of Pittsburgh Press.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*, 492-527.

Mayo, D. (2014). Learning from Error: How Experiment Gets a Life (of its Own). In M. Boumans, G. Hon, & A. Petersen (Eds.), *Errors and Uncertainty in Scientific Practice* (pp. 57-78). London: Pickering & Chatto.

Michler, G. (2008). *Electron Microscopy of Polymers.* New York: Springer.

Mikhalevich, I. (2016). Experimenting and Intervening: The Pragmatic Value of Intervention. *USML Colloquium Series, March 11*.

Mitchell, S. (2009). *Unsimple Truths.* Chicago: University of Chicago Press.

Nienhaus, G., & Nienhaus, K. (2013). Fluorescence Labeling. In U. Kubitscheck (Ed.), *Fluorescence Microscopy: From Principles to Biological Applications* (pp. 143-169). Singapore: Wiley-Blackwell.

Polanyi, M. (1966). *The Tacit Dimension.* New York: Doubleday & Company.

Radder, H. (2012). *The Material Realization of Science: From Habermas to Experimentation and Referential Realism.* New York: Springer.

Rheinberger, H. (1997). *Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube.* Palo Alto: Stanford University Press.

Rouse, J. (2006). Practice Theory. In S. Turner, & M. Risjord (Eds.), *Handbook of the Philosophy of Science, Volume 15: Philosophy of Anthropology and Sociology.* New York: Elsevier.

Rule, G., & Hitchens, K. (2006). *Fundamentals of Protein NMR Spectroscopy.* New York: Springer.

Ryle, G. (1949). *The Concept of Mind.* Chicago: University of Chicago Press.

Sapp, K. (2019). Interview. *University of Pittsburgh*.

Seegelke, C., & Schack, T. (2016). Cognitive Representation of Human Action: Theory, Applications, and Perspectives. *Frontiers in Public Health, 4*, 1-6.

Sellars, W. (1953). Inference and Meaning. *Mind, 62*, 313-338.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing. II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127-190.

Stegenga, J. (2009). Robustness, Discordance, and Relevance. *Philosophy of Science, 76*, 650-661.

Sterelny, K. (2012). *The Evolved Apprentice.* New York: Oxford University Press.

Tal, E. (2012). *The Epistemology of Measurement: A Model-based Account.* Dissertation: University of Toronto.

Tal, E. (2017). Calibration: Modelling the Measurement Process. *Studies in the History and Philosophy of Science Part A, 65*, 33-45.

Turner, S. (1994). *The Social Theory of Practices: Tradition, Tacit Knowledge, and Presuppositions.* Chicago: University of Chicago Press.

Van Fraassen, B. (2008). *Scientific Representation: Paradoxes of Perspective.* New York: Oxford University Press.

Van Helden, A. (2010). Galielo and the telescope. In A. V. Helden, S. Dupré, R. v. Gent, & H. Zuidervaart (Eds.), *The origins of the telescope* (pp. 183-202). Amsterdam: KNAW Press.

Waters, K. (2004). What was classical genetics? *Studies in History and Philosophy of Science Part C, 35*, 783-809.

Wilson, M. (2006). *Wandering Significance: An essay on conceptual behavior.* New York: Oxford University Press.

Wilson, M. (2018). *Physics Avoidance.* Oxford: Oxford University Press.

Wimsatt, W. (1980). Reductionistic Research Strategies and Their Biases in the Units of Selection Controversy. In T. Nickles (Ed.), *Scientific Discovery, vol. 2: Case Studies* (pp. 213-259). Dordrecht: Reidel.

Wimsatt, W. (1981). Robustness, Reliability, and Overdetermination. In M. Brewer, & B. Collins (Eds.), *Scientific Inquiry and the Social Sciences* (pp. 124-163). San Francisco: Jossy-Bass.

Wittgenstein, L. (1953). *Philosophical Investigations.* Oxford: Blackwell.

Woodward, J. (2006). Some varieties of robustness. *Journal of Economic Methodology, 13*, 219-240.

Yeo, R. (2014). *Notebooks, English Virtuosi, and Early Modern Science.* Chicago: University of Chicago Press.

4.0. From Blood Crystals to Protein Structure: The Development of an Empirical Program

A central claim of this dissertation is that scientific representation is constituted, in part, by techniques for reliably gathering data from a target phenomenon. The previous chapter presented an account of how reliable data are generated by interactions with a target phenomenon. There, I appealed to the notion of an empirical program, alongside related practices such as cross-checking. An empirical program is an established body of empirico-technical knowledge that guides the actions of researchers in interactions with an experimental target in a way that controls for and corrects sources of error. Reliable data, I claim, is data that is consistent and coheres with the theory of technique belonging to an empirical program.

So far, this approach has passed the buck on some fundamental questions. I have claimed that scientific representation depends on data generating practices whose reliability is underwritten by empirical programs, but I have not explained what epistemically underwrites the empirical programs themselves. So a first question is: what good is a concept of reliability, qua coherence with empirical program, if the program itself fails to accurately characterize the sources of error accompanying some technique? That is, what justifies this appeal to an empirical program as guarantor of reliability? Second, what if a theory of technique implicitly or explicitly employs a model of the target-apparatus interactions? If I am seeking to explain how scientific models represent generally, then falling back on a "lower level" of modeling is unsatisfactory—after all, what makes *these* models sources of informative inferences? And if we cannot appeal to a mapping relation between model and target, then how do we account for the accuracy of such models or understand how they acquire representational content?

These are questions to which the following chapters will respond. My means of answering them are based in an account of the historical process by which an empirical program becomes established, where this involves a reciprocal movement between developments in technical knowledge and the conception of the target system itself. This chapter is primarily historical, covering the details of this process. The following chapter summarizes the major philosophical points that the history is taken to support. The short

answer is that the early stages in the development of an empirical program do not depend on scientific models in the sense employed elsewhere in this dissertation. What is found at these stages is instead a piece-meal accumulation of causal generalizations resulting from an empirical survey of the "parameter space" of a given procedure. At first, different generalizations may come and go, but through a community effort, these can stabilize into a means for consistently producing a phenomenon. This process of stabilization, which establishes a particular characterization of the target system, has a "scaffolding" quality, in which one technique is subsumed into a larger procedural complex, its stable result serving as the input for a subsequent technique. In this way, a conception of the target system becomes entrenched and is gradually articulated in finer detail as a phenomenon initially localized to a single procedure becomes integrated into a range of techniques, leading to more robust theoretical models of the underlying system.

Sophisticated experimentation in science involves the integration of multiple relatively autonomous stages of technical work coordinated around a common system of interest. The cumulative process I describe is often staggered between these stages, each of which may exhibit its own historical trajectory punctuated by points of interaction between them. Stages of experimentation are linked by *points of stability*—relatively stable, portable materials or effects that are the outcome of one process and serve as "raw material" for another. The case study I will use to demonstrate this process unfolds in three acts. It begins with the crystallization of animal blood in the mid-nineteenth century. Once these crystals could be consistently produced, physiological chemists were able to subject them to further scrutiny, and thereby develop a more specific notion of the content of these crystals, eventually dubbed haemoglobin. These chemists came to categorize haemoglobin among the increasingly large class of organic substances known as proteins. Meanwhile, in act two, physicists developed a novel technique for testing their theories of X-rays with the use of crystalline materials; a technique that in turn became a novel means of examining the molecular structure of crystals. Finally, as biochemical debates gradually settled on a conception of proteins as large molecules, it became evident that crystallized proteins could be integrated with X-ray techniques in order to probe their structure. It was this process that established the ongoing empirical program of protein X-ray crystallography, with a three-dimensional model of haemoglobin as one of its first major

successes. This process illustrates how a particular form of scientific representation and its accompanying empirical program get "off the ground," and is thereby meant to address some fundamental questions laid out above.

This way of answering, rooted in historical epistemology, will frustrate those looking for a particular kind of philosophical response. One may object that I have given a genetic account of representation—describing how it comes about—but have failed to explain what representation *is*. That is, this historical approach fails to explain the inferential use of scientific representation in terms that go beyond the process by which scientists come to recognize representations as useful. What this objector wants is to abstract from these epistemological details to arrive at an ontological account of the relationship *itself* between a model and its real-world target. I leave my response to this demand to one of the following chapters.

Two additional notes before embarking on the history: first, I must acknowledge the historiographic influences on this chapter. Galison's (1987) conception of instrumental traditions in the history of physics, Rheinberger's (1997) depiction of researchers engaging with open-ended experimental systems, Kohler's (1994) work on the material culture of laboratory geneticists (with its own influences from environmental history), and Rasmussen's (1999) account of the way biologists standardized the practice of "seeing" with electron microscopes are all especially noteworthy. These authors demonstrate the extent to which scientific work can revolve around the development of empirical techniques and material objects of inquiry, in a manner that is relatively autonomous from high-level theorizing, but which introduces its own conceptual puzzles and localized bodies of knowledge. More recently, authors like Landecker (2007) and Shickore (2007; 2017) have built on these insights in novel ways. Schickore's focus on "second-order concerns," in which researchers investigate the instruments they are using to investigate the world, is of particular interest to me. Like her, I am ultimately less interested in the materiality of scientific technology as such and more interested in the way researchers grapple with their materials to determine "the status, merits, possibilities and limits of their investigative tools" (2007, p. 3) in a way that leads, ultimately, to the kind of knowledge that "finds its expression in instruction manuals" (2007, p. 13). Yet where Shickore is primarily focused on

the evaluation of research instruments, I am also interested in how the development of this second-order understanding interacts with the objects of study. This leads me to adopt a broader historical scope while holding onto the insights of this tradition. In doing so, I take myself to be following a suggestion from Rheinberger (2009), who has noted that many of the case studies on the development of molecular biology have been situated at the level of short-range history, concerning the events around particular experiments, theorems, model organisms, instruments, and the "external" factors surrounding their occurrence. "It is timely," he concludes, "to ask how these micro-histories may become integrated into wider contexts" (p. 11).[56]

A second preliminary note: I must admit that the ambitions of this historical project far outweigh what can be done in a single chapter. There is clearly much more to say on the topics than can be covered here; each sub-section is deserving of its own study. Thus, what I present here is best regarded as an overview with respect to the breadth of historical details that are available. My goal is to get the history right, to tell the story to the extent that I can, but for philosophical ends. The ultimate priority of these ends has determined the amount of detail provided here. Because the chapter forms part of a response to the questions raised at the beginning of this section, I give added attention to the early history on nineteenth century practices of blood crystallization, to which I turn now.

4.1. The study of blood crystals, 1840-1909

The contemporary study of proteins through X-ray diffraction has a basic structure, the stages of which are identified by the Figure 1:

[56] A similar interest has been expressed by Frederic Holmes, who has written, "If we wish to grasp the larger picture we must show how individual investigative pathways are linked into a network of interacting investigations that comprise the moving front of a shared set of research problems, or a scientific subfield" (Holmes, 1993, p. 122), while acknowledging that this may affect the resolution of the individual pathway account.

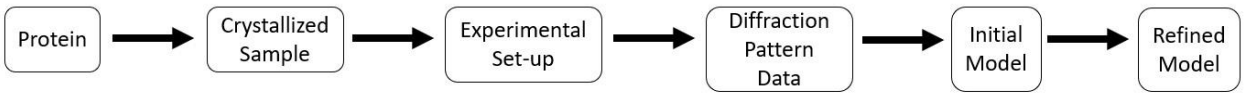


Figure 1: Stages of protein X-ray crystallography

Each of these stages calls on epistemic resources tailored to the particular problems and purposes that dominate at their respective phase of data gathering. The first arrow, denoting the crystallization stage, is just one of these. Given that it occurs upstream of the other stages, the preparation of adequate crystals serves as a basic prerequisite for successful data gathering. To this day its application to a new protein, acid, or virus is an uncertain and potentially laborious process. For this reason, one finds contemporary literature calling crystallization an art lacking universal method (Giegé, 2013). Here I will show how a body of technical knowledge developed around this means of preparation in the nineteenth century, how it took on a consistent form that enabled several diverging lines of research, and how it was situated within a developing understanding of proteins and the components of blood.

The early study of blood crystals and their contents began with their discovery by Friedrich Ludwig Hünefeld in 1840 and culminated in a voluminous study by Edward Reichert and Amos Brown published in 1909. These studies were situated at the intersection of two scientific fields that branched off from chemistry in the beginning of the nineteenth century. The first of these was mineralogy, which typically encompassed or was closely related to studies and theories of crystalline substances. The second was a field known as physiological chemistry in Germany, where much of this research flourished. An overview of each field will help contextualize the work on blood crystals that occurred during this period.

4.1.1.Nineteenth century crystallography

Contemporary crystallography, as employed in X-ray research, is founded in a mathematical theory of the symmetries of crystalline substances. This has motivated some historically-inclined authors[57] to

[57] E.g., (Kahr & Shtukenberg, 2012).

separate mathematical and empirical traditions in the development of the field. However, no such clear distinction is evident in the foundational works. The Danish scientist and bishop Niels Steensen (Latinized as Nicolas Steno) is credited for discovering the first law of geometrical crystallography, but for him this was neither a law nor strictly geometrical. In one of his final scientific papers, a geological essay from 1669 presented as a preliminary tract on "a Solid Naturally Contained Within a Solid," Steensen described various forms of earthen materials and speculated on the mechanical principles underlying their production. Foremost among these was the claim that a solid body "grows by the addition of new particles, secreted from an external fluid" (Steensen, 1669). In a section "Concerning Crystals," Steensen noted a regularity among those growing in this manner: "sometimes the length and number of the sides in the plane of the base are changed in various ways as new crystalline material is added to the plane of the pyramids, without the angles being changed." (p. 659), as illustrated in a diagram (Fig. 2).

Figure 2: Steensen's diagram showing the constancy of angles in a growing crystal.

This observation was repeated throughout the eighteenth century, until the wide-ranging examinations of Romé de l'Isle led him to announce the "law of constancy of interfacial angles" in his *Essai de Cristallographie* of 1772.

The fact that each crystal has characteristic angles which remain constant between its faces provided a straightforward means of classification for different mineral and crystal types. This task was aided by a crucial piece of equipment for crystallographers, the contact goniometer (Fig. 3) invented by de l'Isle's student Arnauld Carangeot in 1782. This device combined two straight-edges with a scaled semi-circle, which allowed the user to read off an angle when each edge was placed at a right angle to adjacent crystal faces.

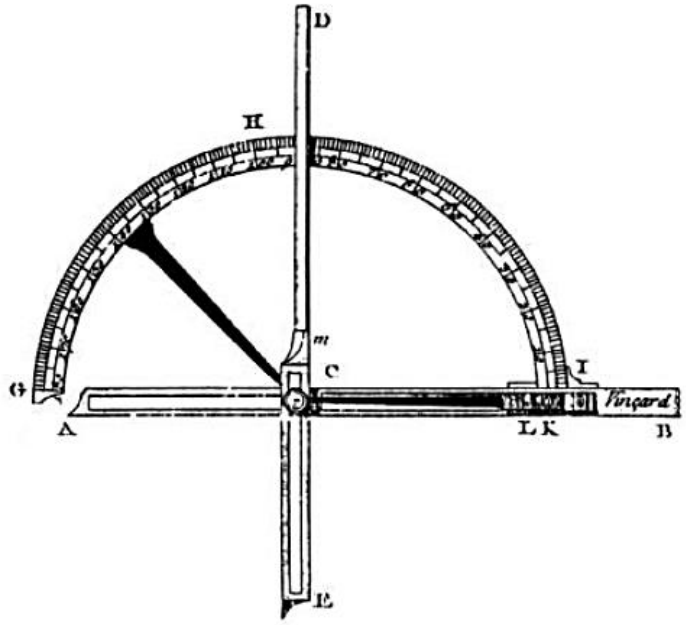


Figure 3: A contact goniometer based on Carangeot's design, from Haüy (1822).

However, it was noted as far back as Steensen that crystal growth was uneven, seemingly affected by a wide range of ambient conditions. In the late eighteenth century, René Haüy experimented with cutting crystalline substances along various axes. He proposed that the characteristic "primitive" form of each substance could be arrived at by repeatedly dividing a crystal along those planes where it cleaves most easily, until no further division was possible. All other forms were deemed "secondary." Haüy announced the aim of crystallography in his first publication: "to determine the form of the constituent molecules of crystals, and the manner in which they are arranged among each other within each crystal. I call this combination *structure*" (1784, p. 9). This use of "constituent molecules" did not refer to a combination of atoms, but only the uniform building blocks assumed to be initially suspended in fluid, which join at their faces through mutual attraction and so build up a crystalline form. Haüy thus established a view of crystals as comprised of regular, repeating units whose nature was systematically related to the macroscopic shape and angles of crystals.

Haüy's stated goal was to establish the laws of growth for each crystal by inferring the structure of its constituent molecules from a primitive form. This effort sometimes took on a speculative geometrical character, largely abandoned by his empirically-minded followers. The Oxford chemist John Kidd, who

drew heavily on Haüy's descriptive work in his *Outline of Mineralogy* (1809), presented an approach where the primitive form of a crystal (of the six fundamental types first identified by Haüy)[58] and its angles are presented alongside a range of other empirical properties. These properties, none of which were consistently listed for each mineral, included fusibility, determined by susceptibility to melting via concentrated heat; hardness, determined by tests such as its ability to scratch glass; the angle of light refraction; and so on. This empirical approach is represented in a more systematic form by the 1822 treatise of Mohs (Fig. 4). Here crystals were primarily classified in terms of their basic geometry according to Haüy's theory, but this information was joined with descriptions of their hardness, color, and other observable characteristics. A further development along these lines resulted from the studies of David Brewster, which laid out correlations between the optical and morphological symmetries of crystals. In his *Optics* (1831), Brewster recounted experiments correlating crystal forms with optical effects. He found, for example, that only crystals with primitive cubic, octahedral, and rhomboidal dodecahedral forms refract light rays without splitting them into two.[59] This treatise also described the phenomena produced by placing different kinds of crystals between a light polarizing apparatus, in which patterns of light of varying hues and intensity are observed depending on the position of the crystal, the polarization of the incident light, and the symmetries of the primitive crystal form (Fig. 5). Further studies at this time by figures such as Franz Neumann established that crystals may vary in other properties (hardness, electrical conductivity, etc.) along their different axes of symmetry.

[58] By his 1822 treatise on mineralogy, Haüy had identified a seventh class, found in later works through the century.
[59] This phenomenon of double refraction was first studied in a treatise on Iceland spar by Bartholin (1669).

GENUS IV. ROCK-SALT.

1. HEXAHEDRAL ROCK-SALT.

Hexahedral Rock-Salt. JAM. Syst. Vol. III. p. 1. Man. p. 9. Common Salt. PHILL. p. 193. Natürlich Kochsalz. WERN. Hoffm. H. B. Th. III. 1. S. 222. Steinsalz. HAUSM. III. S. 843. Steinsalz. LEONH. S. 619. Soude muriatée. HAÜY. Traité, T. II. p. 356. Tabl. comp. p. 20. Traité, 2de. Ed. T. II. p. 191.

Fundamental form. Hexahedron. Vol. I. Fig. 1.
Simple forms. $\dot{H}$ (*P*); $\dot{O}$ (*o*) Vol. I. Fig. 2.; D. Vol. I. Fig. 31.; A_2. Vol. I. Fig. 32.
Character of Combinations. Tessular.
Combinations. 1. H. O. Vol. I. Fig. 3. and 4.
2. H. A_2. Fig. 152.
3. H. D. A_2.
4. H. O. A_2.
Cleavage, hexahedron, perfect. Dodecahedron, often distinct, though generally only a few faces. Fracture conchoidal. Surface generally smooth; the faces of the icositetrahedron sometimes rough.
Lustre vitreous, somewhat inclining to resinous. Colour, generally white, passing into yellow, flesh-red and ash-grey. Sometimes beautifully violet-, berlin-, or azure-blue. Streak white. If scratched with the nail it does not yield any powder, but receives an impression, and becomes a little shining. Transparent ... translucent.
Rather brittle. Hardness = 2·0. Sp. Gr. = 2·257, a yellowish-white transparent variety. Taste saline.

GENUS V. AMMONIAC-SALT.

1. OCTAHEDRAL AMMONIAC-SALT.

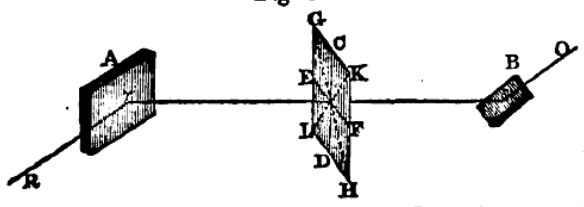
Octahedral Sal Ammoniac. JAM. Syst. Vol. III. p. 11. Man. p. 11. Muriate of Ammonia. PHILL. p. 194. Natürlicher Salmiak. WERN. Hoff. H. B. Th. III. 1. S. 219. Salmiak. HAUSM. III. S. 853. Salmiak. LEONH. S. 631. Ammoniaque muriatée. HAÜY. Traité, T. II. p. 360. Tabl. comp. p. 22. Traité, 2de Ed. T. II. p. 221.

Fundamental form. Hexahedron. Vol. I. Fig. 1.
Simple forms. $\dot{H}$; $\dot{O}$ (*P*), Vol. I. Fig. 2.; D, Vol. I. Fig. 31.; $\dot{C}_1$ (*z*), Vol. I. Fig. 34.
Char. of Comb. Tessular.
Combinations. 1. H. O. Vol. I. Fig. 3. and 4.
Cleavage, octahedron. Fracture conchoidal. Surface smooth.
Lustre vitreous. Colour generally white, often inclining to yellow or grey. Sometimes it is stained green, yellow, or black. Transparent ... translucent.
Very sectile. Hardness = 1·5 ... 2·0. Sp. Gr. = 1·528. Taste acute and pungent.
Compound Varieties. Stalactitic, botryoidal, globular, reniform shapes, also in crusts: composition columnar. Massive: composition impalpable. Fracture conchoidal. Sometimes in a state of mealy efflorescence.

Figure 4: Two pages from the English edition of Mohs's Treatise (1825).

In order to exhibit these phenomena, let a polarising apparatus be prepared, similar in its nature to that in *fig.* 87.; but without the tubes as shown in *fig.* 94., where A is a plate of glass which polarises the ray

Fig. 94.



R *r*, incident upon it at an angle of 56° 45′, and reflects it polarised in the direction *r s*, where it is received by a second plate of glass, B, whose plane of reflexion is at right angles to that of the plate A, and which reflects it

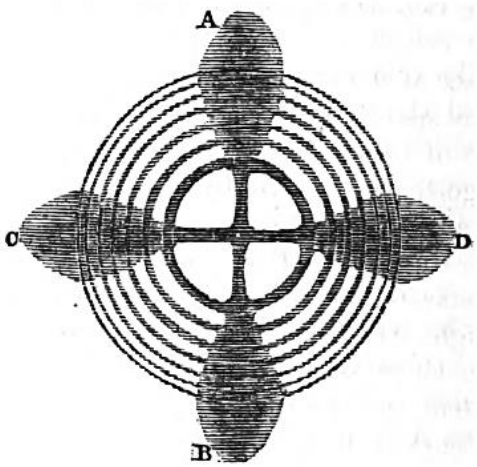


Figure 5: Brewster's 1831 treatise, showing a schematic polariscope and "color rings" produced thereby.

These were the broad contours of crystallographic research within which the study of blood crystals arose: Crystals were grouped into classes based on their primitive polyhedral form, and further specified in terms

of their characteristic interfacial angles, hardness, and other empirical characteristics. The symmetries of the crystal forms served as a key due to their correlation with many other properties and, per Haüy, their supposed link with primitive building blocks. By the mid-nineteenth century, a scientist seeking to identify a given crystal had a range of empirical techniques and classificatory guidebooks to draw on.

Many ensuing developments were technological in nature, pushed by the incessant drive for further precision. Most noteworthy among these were the use of optical goniometers—using reflected light instead of straight-edges to measure interfacial angles—and the pairing of instruments like these with improvements in microscopy (Fig. 6). Similarly, instrument makers fashioned goniometers that could rotate a mounted crystal within two, and then three, planes of motion.[60] These changes obviated the need to remount a crystal in order to measure angles along different zones.[61] Many comparable time-saving and precision-enhancing improvements on earlier designs can be found in the history of mineralogical instruments.[62]

[60] Two- and three-circle goniometers were first developed in 1874 and 1899, respectively (Schuh, 2007).
[61] A crystal zone is a group of faces that intersect in parallel edges. For instance, any two pairs of opposing faces on a cube belong to a single zone.
[62] One example is found in the analytical balance: prior to the advent precision models in the 1820s, common beam-and-pan balances were only sensitive to about 0.5 mg. According to one account, the smallest weight used by the famed chemist Berzelius was 5 mg (Szabadváry, 1986). Within a century, Friedrich Emich had completed a design that could consistently return measurements within 5×10^{-9} g of each other (Schuh, 2007), an improvement of five orders of magnitude.

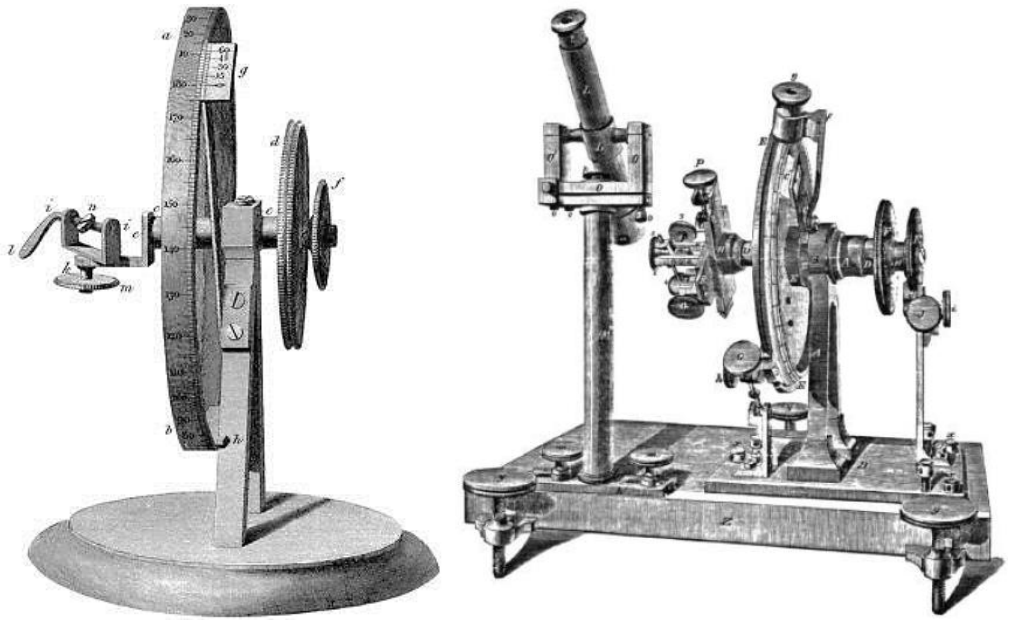


Figure 6: On the left is the first reflecting goniometer design from Wollaston (1809), in which an image is reflected off a face of a crystal fixed to horizontal plane. This plane is rotated by the vertical wheels until the image is seen once more in the crystal, at which point the angle of rotation is noted. On the right is a design from Mitscherlich (1843), which include such improvements as stabilizing legs, more screws for securing the position of the sample, a collimator tube for light on the sample, a microscope for viewing the reflecting crystal, and a second microscope for viewing fine gradations in the angle measurement.

Progress was less forthcoming with respect to the deep structure of crystals. In 1819, the Berlin-based chemist Eilhard Mitscherlich published results showing crystals of the same chemical composition may take on different morphologies and, later, that crystals of the same morphology may have different chemical composition. This compromised any straightforward approach to systematizing Haüy's association of crystal forms and "molecular" constitution. Still, the idea of such a connection remained a lodestar among students of blood crystals into the following century. This might be attributed to the lack of other means of investigation; theories of the structure of crystal constituents remained largely speculative throughout the century. Their main forms are seen in the mathematical lattice studies of continental figures like Bravais (1866) and Sohncke (1879) on the one hand, and in more physicalistic "atom-packing" models of Britons like Wollaston (1813) and Barlow (1883) on the other. Each approach led to its own depiction of crystal structure (Fig. 7) and to accounts of the symmetries within a three-dimensional system of point configurations. However, the connection between these symmetries, ultimately found to comprise 230 "space groups," and those identified in the observable forms of crystals remained obscure.

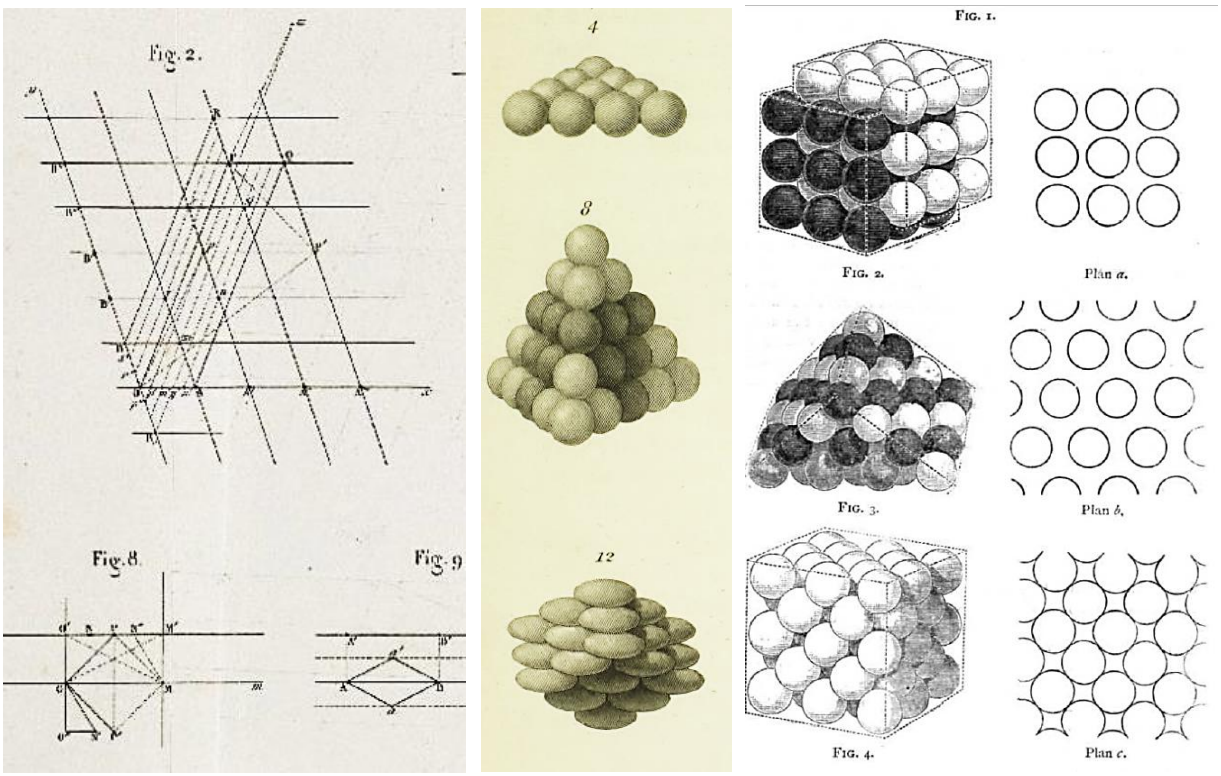


Figure 7: Representations of crystal structure. Left: the lattice approach of Bravais (1866). Center: Wollaston's (1813) atom-packing model. Right: a later atom-packing approach due to Barlow (1883).

4.1.2.Physiological chemistry

Unlike most crystals studied by mineralogists, blood crystals were derived from biological substances. Growing blood crystals required the practical expertise of researchers working in the field of physiological chemistry. This field was dominated by German chemists in the mid-nineteenth century, aided by state investment in the burgeoning research university structure. German state interests in physiological chemistry at this time were exemplified by the development of "experiment stations" in which researchers would study topics such as the impact of different nutrients on livestock health. Major figures in this era of German chemistry—such as Müller, Liebig, and Lehmann—were employed in these stations in the 1850s and 60s (Finlay, 1992).

The development of instrumentation within chemical labs was so gradual that it is only a slight exaggeration to say a biochemist in the early 1900s would easily recognize the lab lay-out of a mid-

nineteenth century chemist, who themselves would be intimately familiar with instruments found in the appendices of Lavoisier (Fig. 8). As with mineralogy, noteworthy advances were made in the material design and precision of equipment. Some significant novelties were introduced, such as Bunsen's spectroscope and gas burner, but the basic operations often remained the same: heating, cooling, mixing, precipitating, filtering, decanting, evaporating, and the like, all performed with familiar glassware. Many shifts in the field instead occurred at the level of procedures and in their accompanying concepts.

Scientists of the early twentieth century credited Justus Liebig's methods with having "laid the foundation of organic chemistry" (Keiser, 1901). Such remarks echoed views from the prior century, such as those of the English translator of Liebig's *Instructions for Chemical Analysis* (1839), for whom the German scientist was "universally acknowledged as the leader in the rapid march of organic chemistry" (p. iii).[63] From 1822 to 1824, a 19-year-old Liebig had spent time working in Paris under the supervision of J. L. Gay-Lussac and Louis Thenard (Rocke, 2000). A decade prior, these authors had developed a method for the quantitative elementary analysis of organic substances. This was based on volume measurements of the gases released upon oxidizing (by combustion—i.e., heating in a tube) a given substance mixed with potassium chlorate. Liebig learned from and built on their methods. After returning to his lab in Giessen he developed novel techniques for separating organic substances into their components. His review of alternative approaches in his *Instructions* exhibits the desiderata for chemical methods of the time. Gay-Lussac and Thenard's method left something to be desired as it "made the accuracy of the results to [sic] depend too much on the dexterity of the operator" (1839, p. 2). He echoed prior complaints from Berzelius and William Henry over the perceived sloppiness of the French chemists, especially with respect to subtle challenges in determining the nitrogen content of novel substances such as alkaloids (Rocke, 2000). These contained a small quantity of nitrogen, for which the use of a eudiometer by Gay-Lussac and Thenard was deemed "not very exact" (1839, p. 2). While praising improvements by Berzelius and Prout, including the

[63] As Rocke (2000) points out, and Liebig's own education makes evident, the historical view of Liebig's legacy expressed by Keiser is overstated.

swapping of slower-burning copper oxide for potassium chlorate, Liebig noted that their methods were only applicable to a limited range of substances.

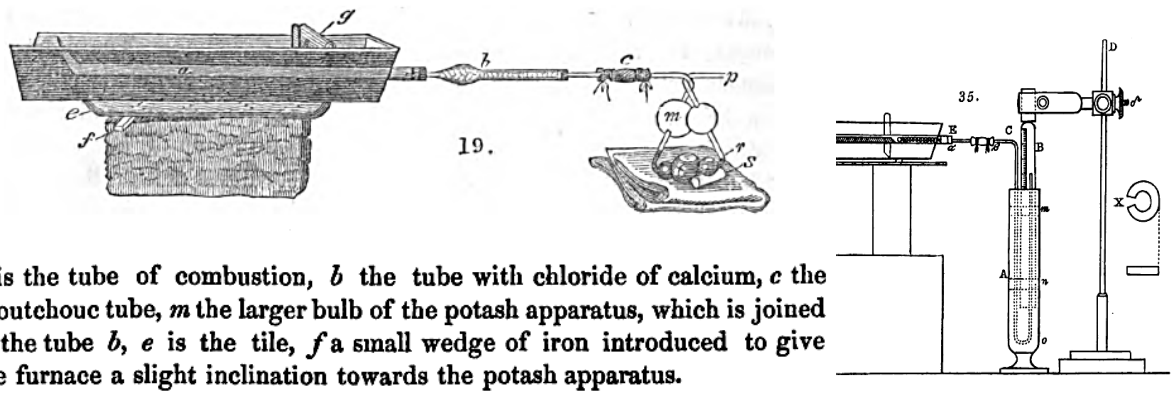
Figure 8: The lab bench in the nineteenth century. Upper left: detail from a widely reproduced engraving depicting Liebig's lab at Giessen in 1840. Upper right: image of a German lab bench in 1875 from Schlöder. Bottom left: a painting of a municipal lab in Paris in 1877 (by Ferdinand Gueldry). Bottom right: a 1901 painting of the first chemistry lab at Victoria College, NZ (by Sybil Johnson).

By contrast, Liebig insisted on the virtues of his own methods:

> It will be observed, that all the parts of the apparatus used for this purpose are extremely simple, and for their employment require no especial dexterity. The essential conditions for performing a good analysis are, the greatest accuracy in weighing, and the strictest conscientiousness in the execution of all the preparatory steps of the process. Let us not

> flatter ourselves that we can obtain an accurate result, if any thing be neglected that can secure it (1839, p. 3).

The basic norms governing Liebig's choice of method thus concerned their range of application, simplicity of instrumentation and execution,[64] and overall accuracy in quantitative results, the latter aided by the scientific ideal of scrupulous self-observation.[65] Thankfully, the *Instructions* give a step-by-step accounting of his general method. This included instructions for blowing a specialized piece of glassware—his "Kaliapparat"—which he combined with the traditional use of a calcium chloride tube and a special technique (sucking air from the far end of the instrument, *p* in Fig. 9) to more thoroughly and effectively separate water vapor and carbonic acid[66] from large quantities of organic substance.



***a* is the tube of combustion, *b* the tube with chloride of calcium, *c* the caoutchouc tube, *m* the larger bulb of the potash apparatus, which is joined to the tube *b*, *e* is the tile, *f* a small wedge of iron introduced to give the furnace a slight inclination towards the potash apparatus.**

Figure 9: Liebig's 1839 diagrams of his heating and separating apparatuses, with heating devices shown on the left-hand side of each. Left: apparatus for separating water vapor and carbonic acid, including the three-bulbed Kaliapparat or "potash apparatus." Right: apparatus for separating out and precisely measuring nitrogen.

Liebig's combustion procedures allowed substances of known composition to be separated from a compound and then measured independently. His book also provided instructions for determining the proportion of an initial substance's weight due to each element.[67] In keeping with the emphasis on accuracy, much of the focus was on sources of error. Determination of carbon, for instance, could be compromised

[64] On this point, Liebig cites a "rule of Berzelius […] of two equally good methods, we prefer the simple to the complicated one" (1839, p. 4).
[65] Nineteenth century emphasis on the reflexive character of scientific subjectivity is a prominent theme in Daston and Galison (2007) and Schickore (2007), among other works.
[66] A substance basically equivalent to carbon dioxide. Dalton (1803), for example, stated that one "atom" of carbonic acid consists of "1 carbone + 2 oxygen." See also the table in Fig. 10.
[67] This was typically presented in terms of "centesimal constitution" where the elements are each given numbers that sum to 100.

by incomplete combustion of the sample, which was corrected by placing it in a longer tube and increasing the amount of the oxidizing agent. Here Liebig again appealed to procedural norms in reference to an extended debate with Berzelius over methods for determining hydrogen: while noting that his method offered no improvement in accuracy, it could be performed much more quickly while claiming "greater simplicity and security, with the same degree of exactness" (1839, p. 30).

Once the proportion of carbon, hydrogen, nitrogen, and oxygen within a substance was determined, their contribution to its original mass could be determined and divided by the known "atomic" or "equivalent" weights of each element to give a chemical formula for the substance (Fig. 10).[68] These formulae had been introduced by Berzelius and were central to the development of organic chemistry in the mid-nineteenth century and beyond. Their main use and advantage, according to Berzelius, was the provision of a scale-independent notational system. That is, the results were independent of the initial amounts of substance analyzed or combined. Importantly, chemical formulae were not thought to convey structural information, beyond showing which basic substance combined with another, and in what proportions (Klein, The Creative Power of Paper Tools in Early Nineteenth-Century Chemistry, 2001);[69] though compatible with atomistic interpretations, some chemists, including Berzelius, purposefully avoided reading them that way.

[68] Equivalent weights are dimensionless units developed from the observations in the late eighteenth century that different masses of different substances would combine with the same amount of a third.

[69] For the notation to carry this information, it had to be assumed that substances forming a compound combined in fixed proportions regardless of the method of preparation—as in Proust's law of definite proportion stated in 1797.

FOUND.		REQUIRED.	
Carbon,	C	Cyanogen,	$N_2 C_2$
Carbonic Acid,	CO_2	Carbon,	C
Carbonate of Baryta,	$Ba O, CO_2$	Carbonic Acid,	CO_2
Carbonate of Baryta,	$Ba O, CO_2$	Baryta	Ba O
Carbonate of Lime,	$Ca O, CO_2$	Carbonic Acid,	CO_2
Carbonate of Lime,	$Ca O, CO_2$	Lime	Ca O
Chloride of Potassium,	KCl_2	Chlorine,	Cl_2

Figure 10: Portion of a table of chemical formulae from Liebig (1839). The use of commas in certain formulae is a means of denoting the perceived "binary constitution" of larger compounds—a subject of dispute until it was abandoned due to its incompatibility with substitution methods (described below).

The basic characteristics of an empirical program for organic chemistry are found in Liebig: a procedural sequence outlining the preparation of an apparatus (how to source instruments, produce copper oxide, dry the substance of interest, mix it with copper oxide, etc.), a target-apparatus interaction (how to initiate combustion and capture reaction products), and data collection and analysis (weight and volume measurements and their conversion to chemical formulae). These steps were paired with an account of possible sources of uncertainty and strategies for coping with different situations (such as an alternate method for measuring results from "volatile substances"). I have called the former a theory of technique; in the case of early chemistry this was largely comprised of generalizations concerning the way particular materials, styles of instruments, or techniques might impact results. In his text, Liebig comments on variants of common instruments, advises on the best materials to use (such as the ceramicist and instrument-maker "Wedgewood's ware"), and compares his methods to others', invoking the above desiderata throughout. In one case, an approach due to Mitscherlich is critiqued, in which the initial weight of a substance of interest is determined by drying it (to remove excess water from a substance—its own source of imprecision), mixing it into a tube of copper oxide, and comparing the weight of the tube before and after this mixture. This method is "less convenient and more troublesome," Liebig complains, since it impacts the potential accuracy of results: "the weight of the tube with oxide of copper, does not permit us to determine the weight of the substance to be analyzed to the 1/500th of a grain. This is a source of uncertainty" (1839, p. 8). While these generalizations employ some theoretical vocabulary, particularly in reference to substances and their

composition, they predominantly concern empirical properties revealed through simple operations. For example, excess water in a substance is not a matter of its having "directly observable" properties corresponding to water, such as being tangibly wet. Rather, excess water is checked by heating a small sample in a test vial and looking for moisture on its sides.

Much of organic chemistry and its cognates in the mid-nineteenth century advanced within this empirical framework. Practitioners were generally devoted to the classification of organic substances through quantitative analyses, like those in Liebig's *Instructions*, and studies of their chemical reactions. Some of these reactions, once properly characterized, came to serve as empirical tests for identifying substances of interest. For example, by the 1850s three "color tests" existed for proteins, all based on the observation of a specific hue when a protein-containing substance was treated with a particular acid solution. A major line of development at this time is seen in the simple accumulation of protocols of this form: new steps for isolating substances, testing for their presence, and otherwise producing observable effects. In this respect, chemists working with organic materials followed the broadly classificatory approach found in concurrent mineralogical and crystallographic texts.[70]

Despite this overall orientation, numerous theories of chemical atomism had sprung up in the first decade of the nineteenth century. These led to distinct sets of formulae and disputes over the proper characterization of compound (Rocke, 2001). This use of formulae was primarily retrodictive, providing compact empirical summaries of proportional combinations (Laszlo, 2001). Though chemists did not shift to more explicitly predictive uses until the advent of structural models in the mid-1850s, the writing of formulae was practically suggestive before then. As Ursula Klein (2003) recounts, the quasi-algebraic character of chemical formulae led to their use as "paper tools," inspiring experiments in which increasingly artificial compounds were synthesized through the successive, step-wise replacement of chemical components. As organic chemists came to focus on applying these manipulations to pure carbon compounds

[70] Classificatory instability was much more the norm for those working directly with plant and animal substances. Early attempts at this were based on their composition out of "primitive substances" that were somewhat arbitrarily chosen based on chemists' abilities to isolate them (Klein, 2003).

like alcohol and ether, the study of plant and animal substances became relegated to the fields of physiological chemistry, medical chemistry, and their ilk.

In the mid-1800s, Germany was rapidly becoming a powerhouse in chemical research. In this context, physiological and medical chemistry were viewed as lesser fields due to their association with applied sciences—both were ultimately concerned with processes in living matter and the functional role of chemical substances within them. Still, leading chemical researchers made their marks on the field, which went on to play a foundational role in training early generations of American and British biochemists.[71] Some examples will illustrate this pedigree: In 1841, Liebig dramatically left the field of pure organic chemistry to study physiology. In Giessen, he trained Adolph Strecker, who then received a position in Tübingen overseeing the work of Felix Hoppe (later Hoppe-Seyler), a Virchow student from Berlin who wrote the foundational textbook *Physiological Chemistry* and started a major journal in the subject. Liebig's longtime collaborator Friedrich Wöhler, supervised Wilhelm Kühne, who went on to succeed Helmholtz in Heidelberg. There Kühne trained several founders of American biochemical research, including R. H. Chittenden (of Yale and Columbia)[72] and J. J. Abel (of Michigan and Johns Hopkins, founder of the *Journal of Biological Chemistry*). Likewise for Britain: Michael Foster, a driving force behind the Cambridge School of Physiology, sent his pupils J. Langely, W. Gaskell, and A. S. Lea to study under Kühne. Ernest Starling at University College London, the "fountainhead of British physiology" (Kohler, 2008, p. 42), had also spent time in Kühne's lab.

Working in Virchow's lab in Berlin, Hoppe trained Ernst Salkowski, who eventually became head of a department of medicinal chemistry and mentored Otto Folin. Folin's colleagues at Harvard included L. J. Henderson, trained under Franz Hofmeister at a leading physiology research institute in Strasbourg. Hoppe-Seyler had left Tübingen in 1872 to run the Strasbourg institute, which served as one of the premiere

[71] Much of the content in the next two paragraphs is derived from Kohler (2008), supplemented by my own proposography.

[72] Yale's chemical training program, the Sheffield Scientific School, was established by Samuel Johnson, a descendant of Liebig. Chittenden's instruction there emulated Kühne's emphasis on pairing general physiology with chemical study.

educational sites for aspiring Anglophone chemists. In addition to Abel and Henderson, students there included Jacques Loeb, founder of the American *Journal of General Physiology* and author of an influential text, *The Mechanistic Conception of Life* (Loeb, 1912). The University of Leipzig had a parallel influence; By the 1850s the school had a separate position for physiological chemistry, then occupied by Karl Lehmann, and subsequently employed figures like Carl Ludwig and Wilhelm Ostwald. Ludwig's physiology-focused model of medical training was a key inspiration for Henry Pickering Bowditch, who established one of the first physiological laboratories in the United States in the early 1870s at Harvard. Loeb and Benjamin Moore (founder of the British *Biochemical Journal* in 1906) were among Ostwald's visiting students.

Upstream links to German education can be found for almost every eminent figure and research institution in American (and much of British) biochemistry. German chemical knowledge and methods were the basis for the emerging physiological and biochemical schools of the English-speaking world, which would become the center of twentieth century protein crystallography.[73] The development of techniques revolving around the analysis of blood crystals are one instance of this general trend; the first dedicated study of these substances was undertaken in 1840s Leipzig, while the last covered here took place in Philadelphia in the early 1900s, during the rise of American biochemistry.

4.1.3. The early study of blood crystals: isolation and analysis of haemoglobin

Blood is the first matter covered in Aristotle's *Parts of Animals*, where it marks the principal division between types—blooded and bloodless. In keeping with their ancient predecessor, Otto Funke's

[73] This is not to say that there was no such work being done in these regions prior to the German influence. Atwater (1978) describes pockets of the United States where pioneering physiological research was carried out before 1870, most prominently in medical studies at the University of Pennsylvania. Americans who made their mark on physiology in the mid-nineteenth century, such as Silas Weir Mitchell and John Call Dalton (by some counts the first full-time professor of physiology in the US), studied under the famed Claude Bernard in Paris, rather than German luminaries. On the other hand, all of this took place against a general background of scarce funding and undeveloped educational institutes, which only began to change in the final decades of the century.

ponderous *Lehrbuch der Physiologie* (1876) and Michael Foster's *Textbook of Physiology* (1880) open with the same topic. It is perhaps no surprise that blood should enjoy this pride of place across millennia; alongside breathing it plays one of the most salient vital functions in larger organisms. By the end of the eighteenth century, French chemists had begun to speculate on a link between blood and breath.[74] This spurred a series of experiments by authors such as Tiedemann, Gmelin, Bertuch, and Magnus on the gaseous contents of blood. In 1837, Magnus overcame experimental obstacles and separated the contents of arterial and venous blood. He noted a larger quantity of oxygen in the former and more carbonic acid in the latter, though his results were not perfectly consistent on this point (Holmes, 1995). It was clear that blood contained gases, yet the manner in which it did so became a point of dispute. In 1803, Dalton published a study of the absorption of gases in liquids via the application of pressure. Though his study was mostly carried out on water, he ventured a theory claiming "All gases that enter into water and other liquids […] are *mechanically* mixed with the liquid, and not *chemically* combined with it" (Dalton, 1803, p. 23). Physiologists argued whether the same could be said for blood.

Various forms of inquiry into the relation between blood, oxygen, and carbonic acid were carried out in the ensuing years.[75] This work pushed researchers to obtain a clearer understanding of the constituents of blood, so to better infer its functional properties. Early on, it was observed that blood contains "globules" among its contents. As microscopy improved, red globules were clearly distinguishable from white (as in (Donné, 1844)). In an early chemical decomposition of blood, Berzelius (1813) focused on three components: fibrin, albumin (so called for its similarity to the substance in egg white), and the "coloring matter" (*matière colorante*, similarly referred to in German texts as *Blutfarbstoff*, later *Blutroth*). Each of these reacted in similar ways to various interventions—coagulable like egg white, soluble in water, but not alcohol or ether; producing a yellow color upon treatment with nitric acid,[76] and so on. Berzelius speculated

[74] Cf. Lavoisier 1790.
[75] For additional lines of research concerning this matter, see Holmes (1995) and Edsall (Edsall, 1972), both of which are drawn on in the following.
[76] An operation due to Fourcory and Vaquelin that would come to be known as the xanthoproteic test for the presence of proteins (Fruton, Proteins, Enzymes, Genes, 1999).

that these "albuminous bodies" might be a modification of the same basic substance, noting that "The largest chemical difference that one finds between fibrin, albumin, and coloring matter consists in a certain quantity of iron oxide contained in the color matter" (p. 45). In 1838, Gerardus Mulder published a method for analyzing the content of fibrin, albumin, and casein (or gelatin). Finding similar proportions of elements and atomic weights for fibrin and albumin (Fig. 11), he too ventured that a common compound may underlie them. He referred to this compound with a term suggested by Berzelius, "protein." At the time, Berzelius's "coloring matter" was distinguished from these proteins. Le Canu had claimed in 1830 that any commonalities were due to a failure to fully separate it from the albumin in blood. Upon boiling the substance in acidified alcohol and removing the residue, he obtained a purified coloring matter that contained iron, but failed to exhibit the same properties observed in 1813. Upon repeating this method, Berzelius chose the name haematin for this substance. He corrected Le Canu on the accompanying residue: it was not albumin, but something else, which he dubbed globulin. Mulder then carried out analyses of each and confirmed that globulin was akin to the other proteins. By 1840 it was thus believed that the coloring matter, which Berzelius had taken to calling *Blutroth*, was a protein-like combination of globulin and the iron-containing haematin.

	Fibrine.	Alb. des oeufs.	Alb. du serum.
Carbone	54,56	54,48	54,84
Hydrogène	6,90	7,01	7,09
Nitrogène	15,72	15,70	15,83
Oxygène	22,13	22,00	21,23
Phosphore	0,33	0,43	0,33
Soufre	0,36	0,38	0,68

Figure 11: A table from Mulder (1838) showing the proportions of elements in fibrin, egg albumin, and albumin from blood serum.

Around this time, the first recorded production of blood crystals took place. Within an extensive treatise on animal chemistry, Friedrich Hünefeld noted a curious result after leaving human and pig blood[77] to dry between two glass plates: "I saw tabular crystalline precipitations appear, which looked sharply

[77] Some sources claim that this was first done with earthworm blood, perhaps because Hünefeld had published a paper on the topic the year prior, though I found no reference to crystallization.

defined and bright red under the microscope" (1840, p. 160). Beyond observing that these crystals would dissolve in water, blood serum, and whole blood, Hünefeld took little interest in the crystals. They were largely forgotten until 1851. In that year a student in Lehmann's lab in Leipzing, Otto Funke, published a study on blood taken from a horse's spleen. When the blood is treated with water, he wrote, "I must describe one of the strangest phenomena […] which to my knowledge has not yet been seen by anyone" (1851, p. 184).[78] Like Hünefeld, Funke had placed the blood between glass and poured a small amount of water on them. As the water evaporated, crystals began to form. In addition, he varied his methods: noting that alcohol and ether could be used as alternate "mother liquors." Funke examined the results with a microscopic goniometer, providing measurements of their angles and concluding that all were rhombic in form. He was otherwise "not yet in a position to publish anything more precise about their nature and chemical constitution" (p. 185).

Such questions would occupy Lehmann and Funke for the following two years. They found that blood from any part of the body could be used, and that different source animals yielded different crystals. The second edition[79] of Lehmann's textbook, *Physiological Chemistry*, summarized their results. Along with varying the mother liquor and source animal, they experimented with ambient conditions, claiming that light and exposure to oxygen, though not strictly necessary, "constitute the most essential conditions toward the rapid formation of these crystals" (1855, p. 349). On the other hand, microscopic inspection showed that the presence of fibrin inhibited growth. The crystals were said to come in three main forms: prisms (human, most mammals, fish); tetrahedra (some rodents); and hexagonal tablets (only squirrels), all rendered in greater detail than prior authors (Fig. 12). Depending on the animal used, certain modifications

[78] Hünefeld's initial finding appears to have fallen into obscurity. Ten years later, Hoppe was using the phrase "Funke's crystals." Funke's 1863 *Textbook of Physiological Chemistry* makes no reference to Hünefeld, though it does acknowledge their discovery by the chemists B. Reichert, Leydig, and Kölliker. These authors, Funke wrote, "had seen blood crystals by chance, but not recognized their true nature" (p. 25). Leydig had in fact wondered in his 1849 note whether these were haematin crystals, but offered no further analysis. The earliest I have seen Hünefeld credited is in Preyer's monograph of 1871.

[79] Published in 1853, first edition 1841. References here are to the 1855 English translation.

in the growing method were required. Lehmann recorded the effects on their form and typically red color upon treatment with various chemicals.

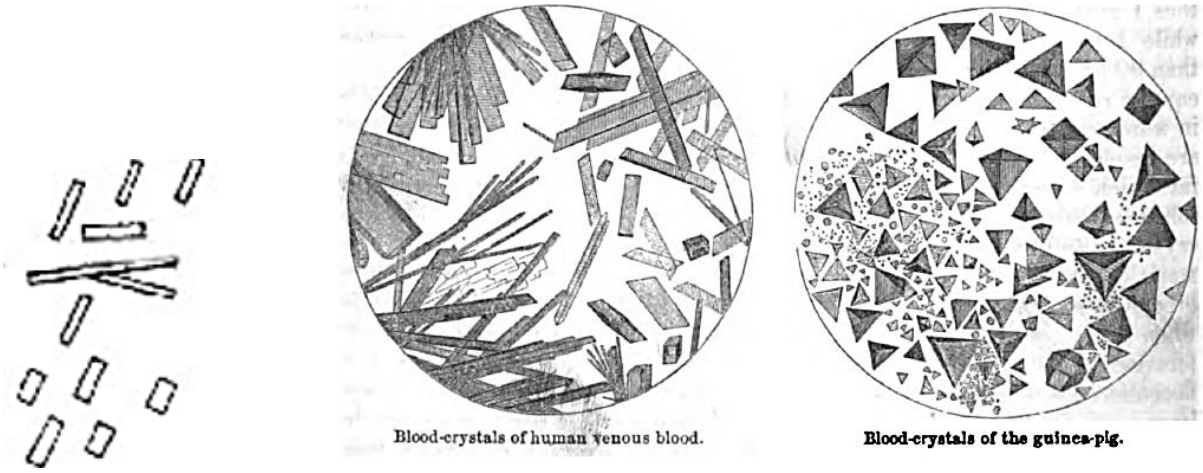


Figure 12: Hünefeld's 1840 depiction of blood crystals (left) compared to Lehmann's in 1855 (right).

Lehmann also wrote of obstacles: "The discovery of a crystallizable protein-substance appeared at once to afford a new means for obtaining more secure points of support for the establishment of its true constitution; but hitherto the elementary analyses of this substance have not furnished the desired information" (p. 347). Part of the reason for this lack of "any satisfactory amount of exactness" (p. 348) was suspected impurities in the crystals; purification of the substance from the cell membranes (or stroma) proved difficult. However, Lehmann was confident enough to assert, "this crystalline substance is not a mixture of a pigment and a protein-body, but a pure chemical compound […] either a salt-like or a conjugated compound," the decomposition of which is "the principal object to be had" (pp. 351-353). Finally, while it "might naturally be supposed that the investigation of this subject would enable us to decide the much-disputed question of the interchange of gases in the circulating blood, […] I have hitherto been unable to obtain any reliable results from my own quantitative determinations" (p. 352).

In a previous section of his textbook titled "Coloring Matters," Lehmann had reviewed the properties of haematin. He noted that it is found in combination with globulin and that experiments on "the action of gases on the color of blood […] certainly indicate that there is a chemical action between the blood-corpuscles and their contents, on the one hand, and the inspired oxygen on the other, in which the

action of haematin doubtless participates" (p. 274). Yet, he opposed the identification of the haematin-globulin complex with his crystals, which he dubbed haematocrystallin. Uncertainty over the contents of the crystals would persist for nearly a decade. Resolution was primarily due to Felix Hoppe, who in the late 1850s began to study the chemical basis of color changes in blood upon exposure to gas. He had begun by investigating *Blutroth*'s (i.e., the coloring matter's) reaction to carbon monoxide. At the same time, Hoppe was harboring doubts about the status of Lehmann's "haematocrystallin," which was said to exist in all vertebrates—how could the same compound be responsible for the variety of crystal forms grown from different animals' blood? To more closely investigate this, Hoppe drew on a novel technique.

Hoppe had taken a position at the University of Tübingen in 1861. The year before, Bunsen and Kirchoff had used new burner technology to show that the light emitted from a pure substance heated to high temperatures had spectra that complemented the substance's absorption lines (known via spectrometry). This provided grounds for thinking these absorption lines carried definite information about the chemical structure of a substance. In his new lab, Hoppe developed methods for observing these lines in liquids (Fig. 13).

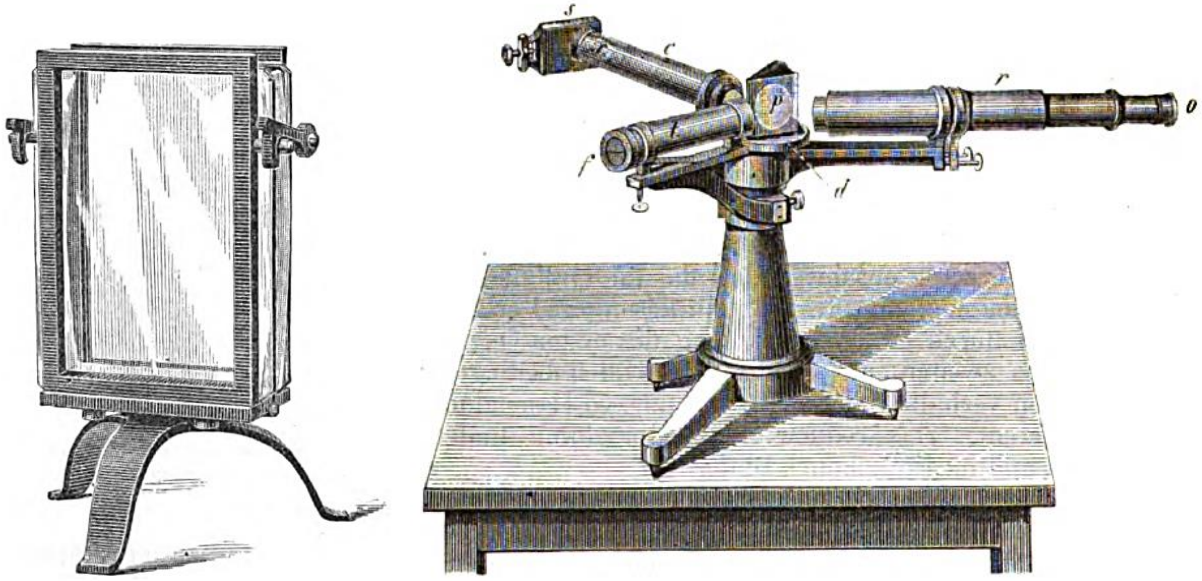


Figure 13: Left: the haematinometer, designed by Hoppe to hold fluid for spectral analysis. Right: A spectroscopic analyzer contemporaneous with Hoppe's work. Both from Funke (1876). Light enters through an adjustable slit (s or p on the left side of the device), passes through a layer of the substance of interest, and is then prismatically refracted (at P) and viewed through an eyepiece (o on the right side of the device). The center region of the device is covered during use to prevent further light entering.

He grew blood crystals, dissolved them in water, and recorded the absorption lines, which he observed to be the same as those of dissolved *Blutroth*. He then treated these preparations with various reagents, finding only those that broke down proteins caused the characteristic lines to disappear. Significantly, individual examination of the decomposition products of *Blutroth*, globulin and haematin, did not produce the same lines as the crystals. For Hoppe, this suggested the crystallization process was in fact a means of obtaining purified samples of the protein *Blutroth*. He claimed "this substance is the same one that forms Funke's crystals" (Hoppe, 1862, p. 449), which he dubbed 'haemoglobin' in 1864. These results were rapidly accepted. Funke summarized the state of knowledge in the following years:

> The crystalline haemoglobin is itself the preformed coloring matter of the blood; It is only through chemical decomposition that it is broken down into a protein substance [globulin], which, depending on the nature of the decomposing agent, either remains in solution or coagulates, and an iron-containing dye: haematin. This decomposition occurs in the aqueous solution of haemoglobin when it is heated, or when alcohol, acids, alkalis, or metal salts are added; it is manifested by the the garnet-red color of the solution changing to a dirty brownish-red, and by a characteristic change in the absorption of light which is revealed when the solution is examined with the spectral apparatus (Funke, 1876, p. 26).

The identification of blood crystals with *Blutroth* motivated Hoppe to refine his crystallization technique. He developed a method based on defibrinating blood (removing the fibrin contents), separating out the blood cells from the remaining serum with salt solution, adding an ether solution to dissolve the cell lining, filtering out the *Blutroth*, and then putting it through multiple cycles of cooling in alcohol (to induce crystallization), filtration, dissolution, and recrystallization. In doing so, he gained confidence in the purity of the crystals. The dissolved crystals showed sharp spectral lines, consistent with his prior results, but now capable of revealing a difference between oxygenated arterial blood and venous blood (Holmes, 1995).

Hoppe's observations inspired G. G. Stokes to further study. Rather than drawing arterial and venous blood for crystallization, Stokes devised a salt solution that, when applied to "colouring matter" extracted from sheep and oxen blood, allowed him to directly change its color from a brighter to a darker more purple red. Recording the spectra of each, he observed the appearance of a dark absorption band between the Fraunhofer lines D and E, where none previously existed (Fig. 14).

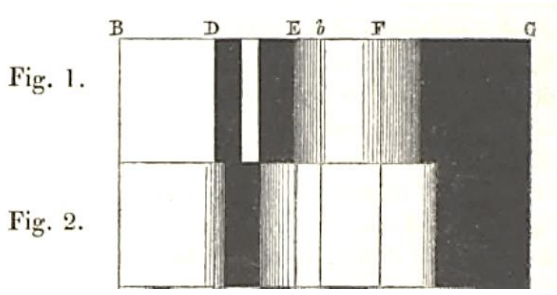


Figure 14: Stokes's (1864) spectral data, showing the spectrum of coloring matter before (top) and after (bottom) treatment with his reducing agent.

Stokes found he could convert these substances into one another by exposing the coloring matter solution to air, then treating it again with his reagent. He concluded,

> We may infer from the facts above mentioned that the colouring matter of blood, like indigo, is capable of existing in two states, distinguishable by a difference of colour and a fundamental difference in the action on the spectrum. It may be made to pass from the more to less oxidized state by the action of suitable reducing agents, and renters its oxygen by absorption from the air (Stokes, On the reduction and oxidation of the colouring matter of the blood, 1864, p. 357).

He proceeded to tie these results to physiological debates over the incorporation of gas into blood. For decades, the authority of Dalton and experiments by Magnus had encouraged the belief that oxygen was

mechanically absorbed into blood. Now, however, Stokes had unambiguous evidence for the presence of a substance in the blood capable of oxidation. Within two years, Hoppe was referring to oxyhaemoglobin and reduced haemoglobin to refer to the change in this oxygen-carrying molecule.

4.1.4. Core features and limitations of methods drawing on crystallization

Hoppe's intervention resolved questions concerning the relation of Lehmann's haematocrystallin to the contents of *Blutroth*, identifying the two. Together with Stokes, he had shown that this substance could chemically incorporate oxygen. Other questions were not so easily resolved. Silas Weir Mitchell's "Observations on the Blood Crystals of the Sturgeon" (1858) provides an example of the characteristic instabilities accompanying the understanding of blood crystals. Alongside Funke's crystallization method, he offered another: leave the fish's blood in an open container, exposed to light and temperatures of 60-70 degrees Fahrenheit for at least two days, then take a drop of the decomposed substance and let it slightly evaporate without any added water to yield "the most beautiful crystals" (p. 2). As with Lehmann's text, Mitchell subjected his crystals to various chemical treatments and described their effects, primarily in terms of changes in color. But in describing his crystals, Mitchell also diverged from Lehmann at several points. For one, those he identifies as "the true albuminoid crystals" reported by others are hexagonal in shape, where Lehmann only found prisms from fish blood. Second, where Lehmann thought the coloring matter was an essential constituent of blood crystals, Mitchell claimed he could remove their redness by treatment with alcohol or dissolving and recrystallizing them. He also claimed to see the same hexagonal form in crystals grown from a wide range of human bloods (varying over sex, various diseases, and even drawing from a fetus and placenta). This appeared to show that the form is constant for a given species, yet "Professor Johnston, of Baltimore, informs me as a contrast to this statement that the splenic vein blood of the opossum afford tetrahedral forms, whilst all the other blood of this animal yields rhombic crystals" (p. 3).

This speaks to two basic features of blood crystal research that persisted into the following century. First, a proliferation of different techniques for obtaining crystals, which initially outpaced the ability to discern one technique's superiority over another. William Preyer's *Die Blutkrystalle* (1871), the first substantial monograph dedicated to the subject, reviews six different methods for obtaining crystals. A review by Arthur Gamgee (1898) mentions nine distinct procedures and several modifications. Reichert and Brown (1909) refer to as many as nineteen methods before introducing four variations on their own. These present a menagerie of seemingly contradictory techniques: scientists were freezing blood, cooling blood, heating blood, boiling blood, drying blood, coagulating blood, electrocuting blood, adding alcohol, adding ether, adding salts, adding bile, extracting gases, adding gases, drawing blood from living animals, drawing blood from asphyxiated animals, injecting water into animals, crystallizing in a vacuum, under sunlight, under artificial light, in the dark…

It appears as though authors reviewing these methods were more interested in providing a historical catalog than methodological guidance. Yet, on closer reading, a strategy can be inferred from this accumulation: as a whole, the variety of nineteenth century techniques present a dogged survey of blood crystallization "parameter space." In running through these procedures, researchers varied a wide range of conditions that could have impacted crystal growth—lighting, temperature, source animal, gaseous atmosphere, and so on. As I noted, even today crystallization is not understood in such theoretical detail that a predictive method is available for any protein; instead, researchers often make use of commercial arrays designed to expose samples to a systematically varying range of growth conditions (see Ch. 3, Fig. 14). And despite the reticence of nineteenth century reviewers to declare one procedure a victor, some clear findings emerged: for example, early suggestions that light, oxygen, or the addition of carbonic acid aided crystal growth[80] were discarded. Other steps suggested in these early texts, such as the removal of fibrin, became widely adopted. In a comparison of 11 procedures introduced after 1858, 7 explicitly refer to the use of defibrinated blood. Solutions to Lehmann's complaints about residual stroma contaminating crystals

[80] Variants of these claims are found in Lehmann and Mitchell's writing.

had also emerged: by 1876, Funke had five ways to fully "liberate" the compound from cell membranes,[81] combinations of which one became standardized as a process called "laking" the blood. 10 of the 11 reviewed procedures involve some way to separate crystal contents from cells. 9 of these involve some combination of freezing blood or mixing it with water, alcohol or ether solution—all methods suggested by Funke—and in all cases this follows shortly after the defibrination step. Cooling, typically to a specified temperature of 0 degrees centigrade, likewise occurs in 7 of these procedures. Beneath the noise of diverse and novel techniques, a common approach was emerging.

As more scrutiny was applied to these methods, physiologists would check the results of different steps under the microscope, developing a rationale for their actions that was lacking in earlier works. Actions were justified by reference to the chemical properties of blood contents. To take one example, it had been observed that less soluble blood (e.g., rodent blood, compared to human or cow) was generally easier to crystallize. As a result, authors began to recommend methods to reduce the solubility of the mother liquor in order to initiate crystallization, and prior procedures came to be understood in these terms. Thus, the "drying" or gradual evaporation of a droplet of blood in solution, employed from the start of crystallization, came to be viewed through the lens of solubility (as in (Funke, 1876)).[82] Following significant results from Hoppe in the early 1860s, his technique became notably popular on the continent. It was imported, with adjustments, in another significant result due to Zinoffsky (1886) and presented as the main way to grow purified crystals in the 1893 lab manual *Practicum for Physiological and Pathological Chemistry* by Salkowski.[83]

All of this aided in meeting something like Lehmann's desire for greater certainty, or "exactness." Liebig (1839), among other early chemists, insisted on analyses using a substance of the "highest degree of purity possible." The reason is straightforward: with questionable purity, no amount of diligence and

[81] By this point multiple authors, including Funke and Mitchell, had observed that the crystals could form inside red blood cells. This further supported the belief that the compound originated from within the cells, along with the straightforward connection with their red color.
[82] This period hereby gave rise to techniques akin to present day "sitting drop" and "salting out" methods for crystallization.
[83] English translation printed in New York (1903).

dexterity on the part of the chemist can prevent inconsistencies in measurement results, hence "No means should be neglected to satisfy ourselves of the absence of foreign matters" (1839, p. 4). By exploring procedural possibilities for crystallization in this way, physiological chemists were seeking to achieve reliability in two ways: consistency in the production of satisfactory crystals, and in the outputs of procedures performed on these crystals.[84] Still, having a purified substance did not always guarantee insights, as slowly became clear to researchers. In fact, the second prominent feature of blood crystal research during this period is a marked confusion over how to interpret the variety of crystallographic features observed under the microscope.

Aside from reactions to chemical tests and treatments, it was not clear how physiologists could determine the nature of this purified substance. Limits to insights afford by elementary analysis were also becoming clear. In the 1830s, Liebig and Wöhler had established the phenomenon of isomerism: substances with the same formula could have different chemical properties. This meant that chemical properties were not determined by atomic make-up alone.

The main alternative to chemical interventions was crystallographic description. The compound microscope allowed researchers to describe minscule features of the crystals and depict these for readers (Fig. 15). When coupled with the goniometer, this allowed for a potentially deeper characterization, exploiting Haüy's hypothesized link between crystal form and "molecular" structure.

[84] A desire for reliability in the sense of Liebig and Lehmann is found in later authors. For example, Zinoffsky (1886) justified his refinement of prior crystallization methods with reference to inconsistencies in prior determinations of crystal iron and sulfur content. It is a testament to the claimed superiority of Zinoffsky's techniques that they yielded analyses of haemoglobin much closer to contemporary values. Where Preyer cites results finding 600 carbon atoms, 960 hydrogen, 154 oxygen, and 179 nitrogen in what we would today call a single heme group (ignoring iron and sulfur), Zinoffsky measured 712C 1130H, 214O, and 245N. The contemporary values are 738C, 1166H, 203O, and 208N.

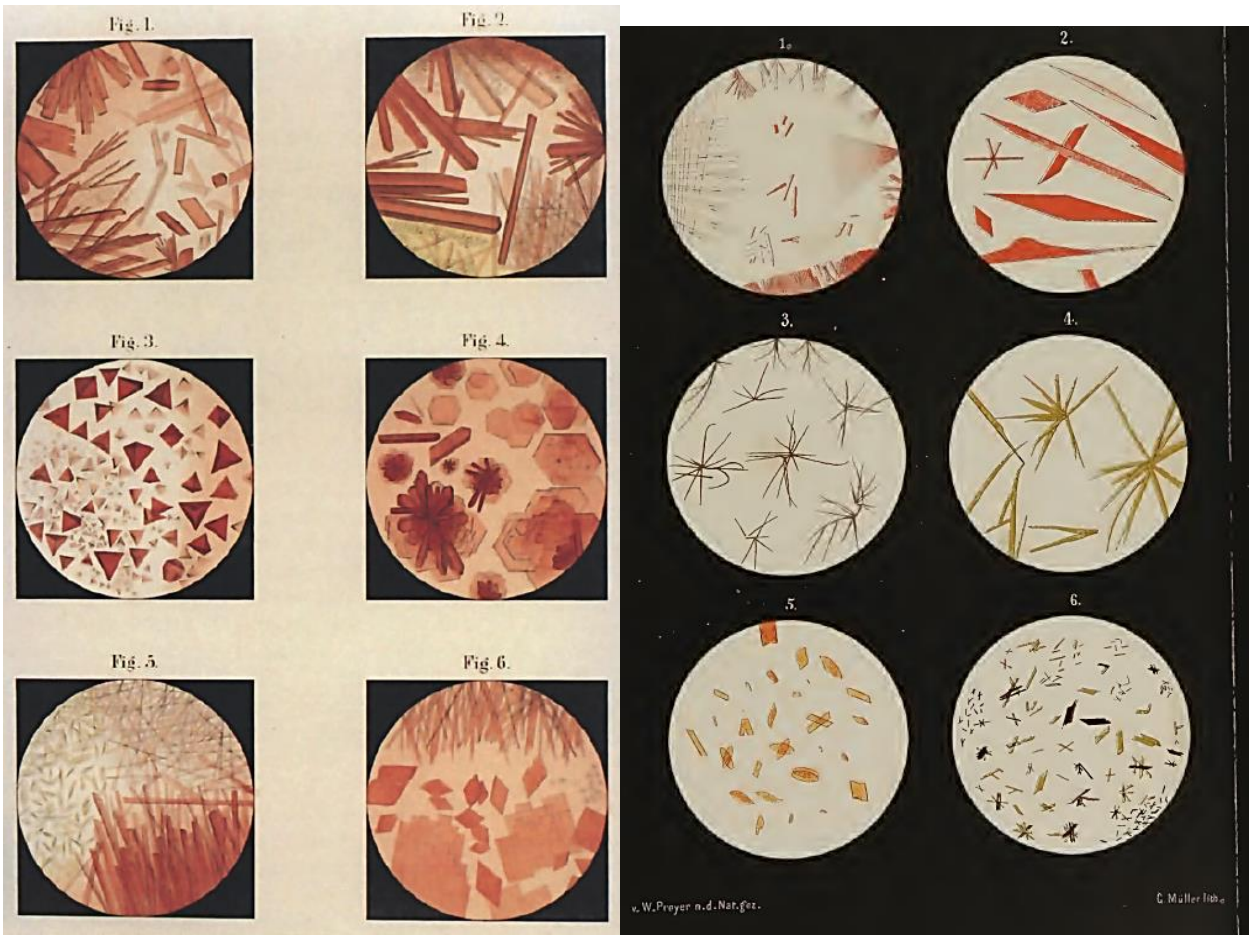


Figure 15: Left: drawing from Funke's Atlas of Phys. Chem. (1853). Right: Preyer's The Blood Crystal (1871).

Mitscherlich may have shown that substances of the same chemical make-up could crystallize differently (a feature referred to as isomorphism), but this did not entail the absence of any link between the form of a crystal and the nature of its building blocks. The notion that such chemical differences could be identified through crystal forms was a major inspiration to blood crystal researchers. Lehmann, after remarking that proteins may have the form of a complex with "several proximate constituents conjugated together," added,

> This view seems to gain support from the remarkable circumstance that there is an albuminous substance in the blood of carnivorous animals which crystallizes in prisms, while the corresponding substance in the blood of guinea-pigs and rats crystallizes in tetrahedra. This obviously points at combinations of an analogous kind, in which only one different constituent has entered, which, however, is the cause of the difference in the crystalline form of the otherwise perfectly analogous body (1855, p. 293).

If there was a link, as crystallographers asserted, between chemical structure and crystal form, then the study of the different forms of crystals found in animal blood could provide insight into the underlying structural differences in the blood of diverse species.

Of course, this potential insight immediately raised questions: for instance, what to make of the results from Silas Mitchell's friend, Johnston of Baltimore, for whom blood from two different places in the same animal formed into different crystals? Many approached these questions with enthusiasm. Preyer wrote that studying blood crystals opened the way to a new form of comparative evolutionary biology, which "sets variability alongside the constancy of proportions" within the temporal progression of animal forms:

> For, since *Blutroth* crystallizes in different systems in very closely related animals, and since their common ancestor had only one type of blood pigment, a material metamorphosis must have taken place at the same time as the change in form. And this means that, even if many simple connections between animal bodies are everywhere identical, the more relevant chemical constituents, especially the albumins and the coloring matter, are not merely different in kind, but are variable in time, variable in the same degree as the zoological species (1871, p. iv).

Over thirty years later, this assumption remained operative in the thinking of American chemists. Thomas Osborne wrote in 1902 of edestin that "the form of the crystal is determined by the protein molecule" (1902, p. 55). The same idea motivated Reichert and Brown's sprawling study of blood crystals in 1909: "Differences of chemical constitution are accompanied by differences of physical structure, and the crystallographic test of differences of chemical constitution is recognized as the most delicate test of such differences" (1909, pp. 144-145). Why is this? Because "*Substances that show differences in crystal structure are different chemical substances*" (p. 145).

These authors may have been emboldened by the rise of structural models in chemistry in the 1850s. Such models initially took many forms but began to stabilize into a consistent graphical presentation following Edward Frankland's popularization of Brown's structural notation in 1866 ((Ritter, 2001), Fig. 16). With these models, the notion of isomerism was easily comprehensible: different arrangements of the same basic units could intuitively be seen to yield different kinds of compounds.

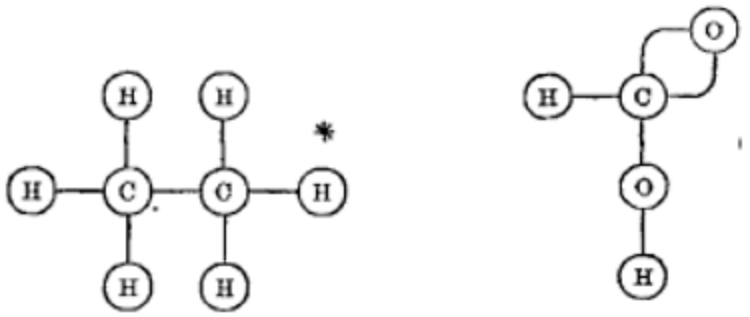
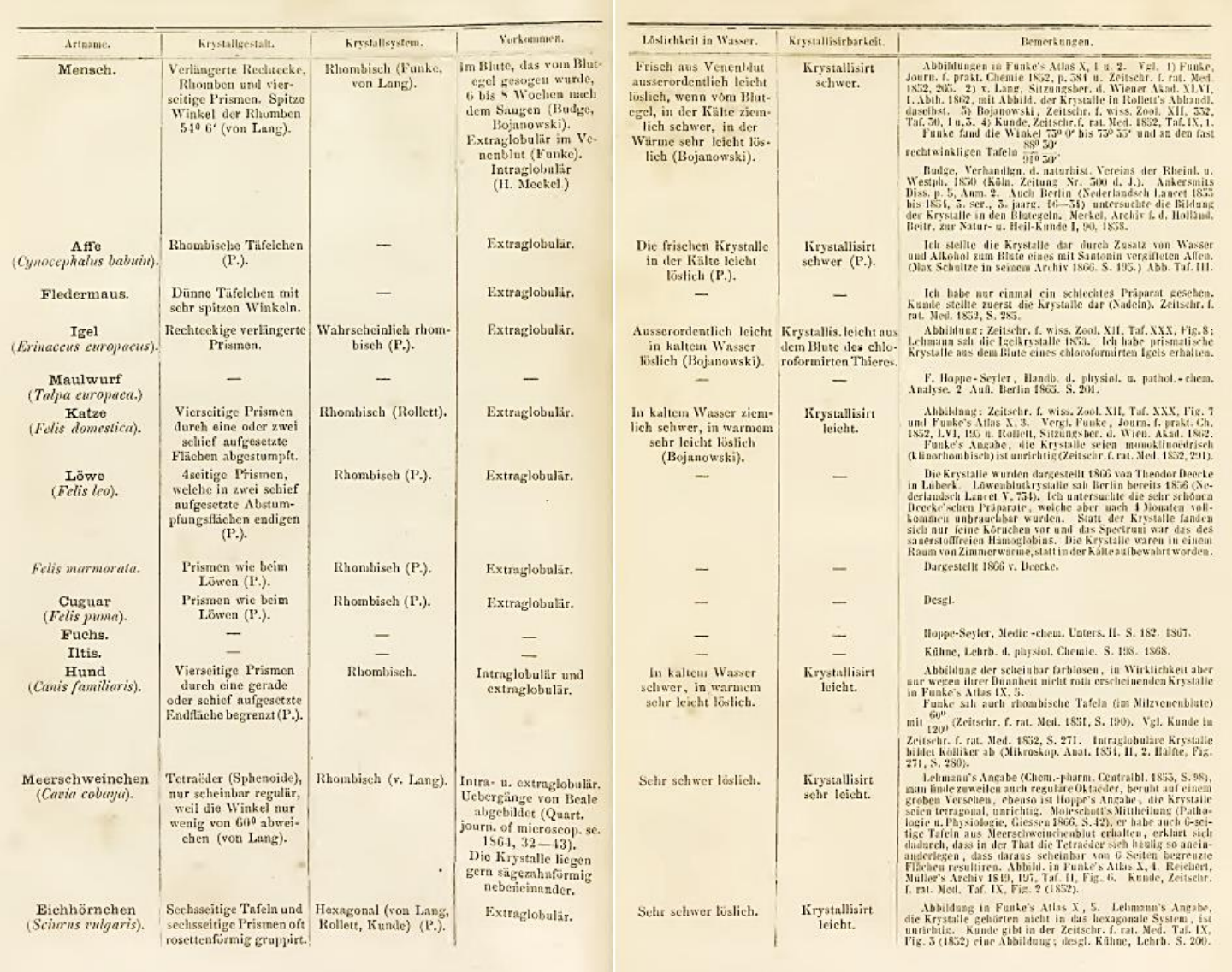


Figure 16: Brown's graphical notation (1864), created in 1861 and later popularized by Frankland.

But physiological chemists never arrived at a detailed presentation of the structure of blood crystal molecules. Instead, studies carried on in the familiar classificatory format, supplemented with an account of the different crystal shapes found in different species. Preyer's monograph, for instance, contains a long table listing the crystals found in over forty animal species: their appearance, the fundamental crystal system they belong to, where they form, their solubility in water, and how easily they crystallize (Fig. 17). On the other hand, he provides no well-supported conclusions about the deeper chemical differences between the blood of different animals.



Artname.	Krystallgestalt.	Krystallsystem.	Vorkommen.	Löslichkeit in Wasser.	Krystallisirbarkeit.	Bemerkungen.
Mensch.	Verlängerte Rechtecke, Rhomben und vierseitige Prismen. Spitze Winkel der Rhomben 54° 6′ (von Lang).	Rhombisch (Funke, von Lang).	Im Blute, das vom Blutegel gesogen wurde, 6 bis 8 Wochen nach dem Saugen (Budge, Bojanowski). Extraglobulär im Venenblut (Funke). Intraglobulär (H. Meckel.)	Frisch aus Venenblut ausserordentlich leicht löslich, wenn vom Blutegel, in der Kälte ziemlich schwer, in der Wärme sehr leicht löslich (Bojanowski).	Krystallisirt schwer.	Abbildungen in Funke's Atlas X, 1 u. 2. Vgl. 1) Funke, Journ. f. prakt. Chemie 1852, p. 584 u. Zeitschr. f. rat. Med. 1852, 265. 2) v. Lang, Sitzungsber. d. Wiener Akad. XLVI, I. Abth. 1862, mit Abbild. der Krystalle in Rollett's Abhandl. daselbst. 3) Bojanowski, Zeitschr. f. wiss. Zool. XII, 352, Taf. 30, 1 u. 3. 4) Kunde, Zeitschr. f. rat. Med. 1852, Taf. IX, 1. Funke fand die Winkel 73° 0′ bis 73° 35′ und an den fast rechtwinkligen Tafeln $\frac{88° 30'}{91° 30'}$. Budge, Verhandlgn. d. naturhist. Vereins der Rheinl. u. Westph. 1850 (Köln. Zeitung Nr. 500 d. J.). Ankersmits Diss. p. 5, Anm. 2. Auch Berlin (Nederlandsch Lancet 1853 bis 1854, 3. ser., 3. jaarg. 16—34) untersuchte die Bildung der Krystalle in den Blutegeln. Merkel, Archiv f. d. Holländ. Beitr. zur Natur- u. Heil-Kunde I, 90. 1858.
Affe (*Cynocephalus babuin*).	Rhombische Täfelchen (P.).	—	Extraglobulär.	Die frischen Krystalle in der Kälte leicht löslich (P.).	Krystallisirt schwer (P.).	Ich stellte die Krystalle dar durch Zusatz von Wasser und Alkohol zum Blute eines mit Santonin vergifteten Affen. (Max Schultze in seinem Archiv 1866. S. 195.) Abb. Taf. III.
Fledermaus.	Dünne Täfelchen mit sehr spitzen Winkeln.	—	Extraglobulär.	—	—	Ich habe nur einmal ein schlechtes Präparat gesehen. Kunde stellte zuerst die Krystalle dar (Nadeln). Zeitschr. f. rat. Med. 1852, S. 285.
Igel (*Erinaceus europaeus*).	Rechteckige verlängerte Prismen.	Wahrscheinlich rhombisch (P.).	Extraglobulär.	Ausserordentlich leicht in kaltem Wasser löslich (Bojanowski).	Krystallis. leicht aus dem Blute des chloroformirten Thieres.	Abbildung: Zeitschr. f. wiss. Zool. XII, Taf. XXX, Fig. 8; Lehmann sah die Igelkrystalle 1853. Ich habe prismatische Krystalle aus dem Blute eines chloroformirten Igels erhalten.
Maulwurf (*Talpa europaea.*)	—	—	—	—	—	F. Hoppe-Seyler, Handb. d. physiol. u. pathol.-chem. Analyse. 2 Aufl. Berlin 1865. S. 201.
Katze (*Felis domestica*).	Viereitige Prismen durch eine oder zwei schief aufgesetzte Flächen abgestumpft.	Rhombisch (Rollett).	Extraglobulär.	In kaltem Wasser ziemlich schwer, in warmem sehr leicht löslich (Bojanowski).	Krystallisirt leicht.	Abbildung: Zeitschr. f. wiss. Zool. XII, Taf. XXX, Fig. 7 und Funke's Atlas X, 3. Vergl. Funke, Journ. f. prakt. Ch. 1852, LVI, 195 u. Rollett, Sitzungsber. d. Wien. Akad. 1862. Funke's Angabe, die Krystalle seien monoklinoedrisch (klinorhombisch) ist unrichtig (Zeitschr. f. rat. Med. 1852, 291).
Löwe (*Felis leo*).	4seitige Prismen, welche in zwei schief aufgesetzte Abstumpfungsflächen endigen (P.).	Rhombisch (P.).	Extraglobulär.	—	—	Die Krystalle wurden dargestellt 1866 von Theodor Deecke in Lübeck. Löwenblutkrystalle sah Berlin bereits 1856 (Nederlandsch Lancet V, 754). Ich untersuchte die sehr schönen Deecke'schen Präparate, welche aber nach 4 Monaten vollkommen unbrauchbar wurden. Statt der Krystalle fanden sich nur feine Körnchen vor und das Spectrum war das des sauerstofffreien Hämoglobins. Die Krystalle waren in einem Raum von Zimmerwärme, statt in der Kälte aufbewahrt worden.
Felis marmorata.	Prismen wie beim Löwen (P.).	Rhombisch (P.).	Extraglobulär.	—	—	Dargestellt 1866 v. Deecke.
Cuguar (*Felis puma*).	Prismen wie beim Löwen (P.).	Rhombisch (P.).	Extraglobulär.	—	—	Desgl.
Fuchs.	—	—	—	—	—	Hoppe-Seyler, Medic.-chem. Unters. II. S. 182. 1867.
Iltis.	—	—	—	—	—	Kühne, Lehrb. d. physiol. Chemie. S. 198. 1868.
Hund (*Canis familiaris*).	Vierseitige Prismen durch eine gerade oder schief aufgesetzte Endfläche begrenzt (P.).	Rhombisch.	Intraglobulär und extraglobulär.	In kaltem Wasser schwer, in warmem sehr leicht löslich.	Krystallisirt leicht.	Abbildung der scheinbar farblosen, in Wirklichkeit aber nur wegen ihrer Dünnheit nicht roth erscheinenden Krystalle in Funke's Atlas IX, 5. Funke sah auch rhombische Tafeln (im Milzvenenblute) mit $\frac{60°}{120°}$ (Zeitschr. f. rat. Med. 1851, S. 190). Vgl. Kunde in Zeitschr. f. rat. Med. 1852, S. 271. Intraglobuläre Krystalle bildet Kölliker ab (Mikroskop. Anat. 1854, II, 2. Hälfte, Fig. 271, S. 280).
Meerschweinchen (*Cavia cobaya*).	Tetraëder (Sphenoide), nur scheinbar regulär, weil die Winkel nur wenig von 60° abweichen (von Lang).	Rhombisch (v. Lang).	Intra- u. extraglobulär. Uebergänge von Beale abgebildet (Quart. journ. of microscop. sc. 1864, 32—43). Die Krystalle liegen gern sägezahnförmig nebeneinander.	Sehr schwer löslich.	Krystallisirt sehr leicht.	Lehmann's Angabe (Chem.-pharm. Centralbl. 1853, S. 98), man finde zuweilen auch reguläre Oktaëder, beruht auf einem groben Versehen, ebenso ist Hoppe's Angabe, die Krystalle seien tetragonal, unrichtig. Moleschott's Mittheilung (Pathologie u. Physiologie, Giessen 1866, S. 42), er habe auch 6-seitige Tafeln aus Meerschweinchenblut erhalten, erklärt sich dadurch, dass in der That die Tetraëder sich häufig so aneinanderlegen, dass daraus scheinbar von 6 Seiten begrenzte Flächen resultiren. Abbild. in Funke's Atlas X, 4. Reichert, Muller's Archiv 1849, 197, Taf. II, Fig. 6. Kunde, Zeitschr. f. rat. Med. Taf. IX, Fig. 2 (1852).
Eichhörnchen (*Sciurus vulgaris*).	Sechsseitige Tafeln und sechsseitige Prismen oft rosettenförmig gruppirt.	Hexagonal (von Lang, Rollett, Kunde) (P.).	Extraglobulär.	Sehr schwer löslich.	Krystallisirt leicht.	Abbildung in Funke's Atlas X, 5. Lehmann's Angabe, die Krystalle gehörten nicht in das hexagonale System, ist unrichtig. Kunde gibt in der Zeitschr. f. rat. Med. Taf. IX, Fig. 3 (1852) eine Abbildung; desgl. Kühne, Lehrb. S. 200.

Figure 17: A table accounting for blood crystals from different animals by Preyer (1871). The species list shown here begins with human, then includes donkey, cat, fox, dog, and squirrel blood. Most species are seen to belong to the rhombic system, though some of these present as prisms while others as tetrahedra (this distinction between appearance and system was not always appreciated, e.g., by Lehmann).

Preyer's book also includes one of the first quantitative elementary analyses of blood crystals in terms of a precise atomic weight. Where prior authors only gave proportional analyses, Preyer's results—given by the formula $C_{600}\ H_{960}\ N_{154}\ O_{179}\ F\ S_3$—suggested that the *Blutroth* was much larger than anything dealt with in standard chemistry. Later authors carried out similar analyses on crystals derived from a range of species, encountering notable variation. By the end of the century, these methods were also plagued with uncertainty. As Gamgee wrote, analyses by different authors show "such extraordinary discrepancies in the

results of ultimate organic analysis as to preclude a precise answer being given to such simple questions as the following: Is haemoglobin a body, having a constant composition in animals of the *same species*? Does the haemoglobin of *different animals* vary in chemical composition, and if so, within what limits?" (1898, p. 199). If available results were taken at face value, the answer to each of these questions would be yes, but for Gamgee this variation was "doubtless due to differences in the purity of the substance analysed, and to errors of analysis" (p. 200). He argued that prior attempts at purification likely resulted in the analysis of a mixture of haemoglobin and products of decomposition. Inconsistencies in measurement results caused a recoil; crystallization methods that Hoppe had regarded as a sure path to purity were no longer trusted. Definitive insights into chemical structure were not, it seemed, readily obtainable from microscopic inspection nor from analyses of the crystals. For all its advances, blood crystallization remained imprecise and not fully trustworthy as a purification method.

4.1.5. Spectroscopic improvements and culmination of the nineteenth century approach

Spectroscopy, on the other hand, had enabled physiologists to infer differences in the underlying chemical structure of blood crystals with greater confidence than other techniques. In addition to his crystal tables, Preyer devoted large sections of his 1871 book to an account of the spectral lines of haemoglobin, of its products of decomposition, and of modified forms resulting from treatments with salts, acids, and other chemicals. Hoppe's successor at Tübingen, Gustav von Hüfner, took this line of work even further, commissioning new precision instruments, synthesizing novel forms of haemoglobin, and recording their distinct spectral properties. By the end of the nineteenth century, oxyhaemoglobin and its reduced form had been joined by methaemoglobin,[85] CO-haemoglobin,[86] NO-haemoglobin, CO-methaemoglobin, and NO-methaemoglobin.

[85] Studied by Hoppe in 1866.
[86] Studied, but without being aware of it as such, by Hoppe in the late 1850s.

As with microscopy, spectroscopy enabled second-order inquiries into crystallization techniques. A simple example is found in Preyer's spectral line illustrations (Fig. 18). Lines 2-8 correspond to gradual increases in the concentration of dissolved oxyhaemoglobin crystals in clear solution. These show a gradual widening of absorption bands as an effect, up to the point where the outcome resembles the large gap seen in the *reduced* haemoglobin spectrum in Line 9. This showed how an overly dilute or overly concentrated sample could yield spectroscopic errors. Spectroscopists gradually became aware that the intensity of spectral lines could be misperceived by unaided observation, leading scientists to infer more or less absorption at a given frequency than had actually taken place. This was corrected by changes in instrumentation—such as a modified eyepiece with a moveable slit that allowed users to view an adjustable range of the spectrum—and further development of the theory of technique under the heading of "spectrophotometry." Among other factors, spectroscopists sought to understand the general quantitative relationship between light absorption and concentration of a fluid in order to correct for density effects and distinguish a molecule's independent contribution to absorption results.

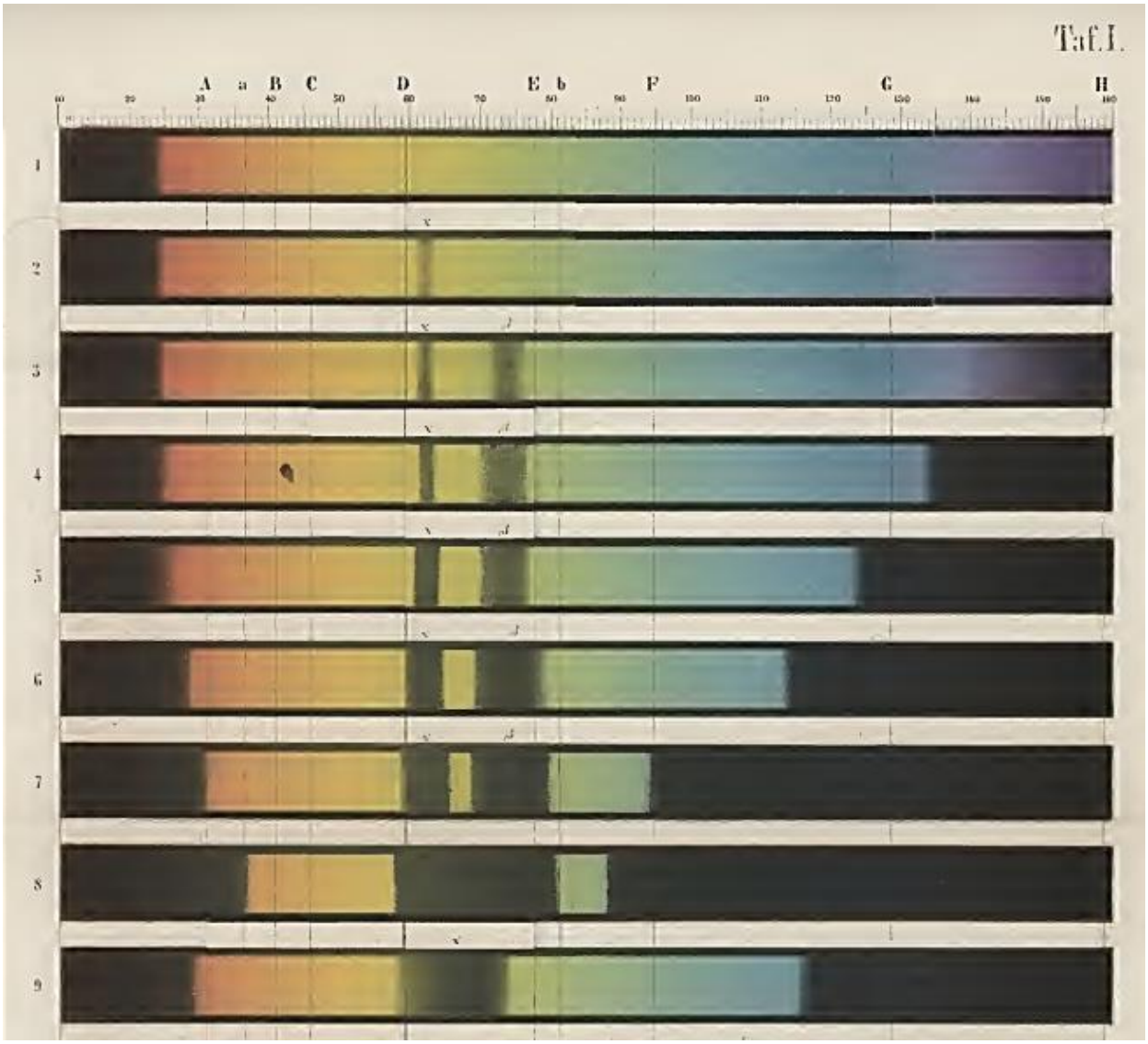


Figure 18: Spectral lines from Preyer (1871). Line 1 is the full light spectrum for comparison. Lines 2-8 show gradual increase in the concentration of oxyhaemoglobin in solution. Line 9 is reduced haemoglobin.

As with microscopy, spectrophotometry was employed to better understand and correct the effects of chemical interventions on blood crystals. For physiological chemists, spectrophotometry introduced a new way to check the purity of a crystallization result:

> Since spectrophotometry has supplied us with a method of determining, with an accuracy previously unattainable, the purity of a colouring matter, it has been found that although oxyhaemoglobin which has been recrystallised, when examined in the ordinary manner, exhibits a spectrum which appears identical with that of the colouring matter which has been only once crystallised, its spectrophotometric constants have changed (Gamgee, 1898, p. 197).

Following Hoppe, many had recommended cycles of recrystallization to remove all traces of impurities from haemoglobin. Repeated recrystallization was central to Zinoffsky's (1886) preparation and chemical analyses, itself motivated by perceived impurities in prior methods. The result cited by Gamgee caused authors to cease recrystallizing for fear of disturbing their sample. Such results likely hastened the broader adoption of the centrifuge as an alternative method of separation.

Two researchers at the University of Pennsylvania, Edward Reichert and Amos Brown (R&B), were sensitive to these developments. As a post-graduate, Reichert had studied in Leipzig and Berlin (Frank, 1987), and subsequently received training from Henry Newell Martin, the first professor of physiology at Hopkins, himself a student of Michael Foster of Cambridge. Reichert's physiological credential were in place and like his German forebears he had a taste for exhaustive applications of method. R&B's *Crystallography of Hemoglobins* (1909) included a thorough review of prior works on the subject before detailing the properties of blood crystals from several hundred species, many obtained from the Philadelphia zoo. This was supplemented with over 100 plates of photomicrographs.[87]

R&B took great care to ensure that they were working with unaltered specimens. By adding and subtracting steps to their procedures and comparing the results, they were able to winnow down the crystal preparation process further than prior authors. Where nineteenth century methods regularly advised for the use of alcohol or alkali salts to initiate crystallization, R&B avoided them due to deleterious effects. Alcohol, for instance, lowered the extinction coefficient (for spectral study), decolorized the crystals over time, and affected the water that sometimes crystallizes with haemoglobin. When they couldn't defibrinate blood, they made use of anti-coagulants like oxalate—but not before inspecting its effects on crystals and determining that its only impact was on the proportion of oxyhemoblin in horse and mule blood. They worked quickly to circumvent the deterioration of crystals over time, and avoided recrystallization at all costs, opting for the centrifuge instead. In short, they built on and scrutinized prior procedures, in the process producing the most extensive survey of blood crystals to date.

[87] This was the first of two volumes that Reichert, like Preyer before him, considered chemical contributions to the understanding of biological taxonomy and evolution. The second was a study of plant starches published in 1919.

Working with so many animals gave R&B the ability to make certain methodological generalizations:

> We found, as we gained experience with the bloods of different species, that, while the blood of each species must be treated as an individual, we could nevertheless depend with some confidence upon the guidance of certain generalizations in the selection of the best method to be pursued. Thus, we found that usually the haemoglobins of Rodentia and Canidae crystallize with great readiness, those of Marsupialia very readily, those of Felidce readily, those of Ungulata not readily, those of Aves with difficulty, etc. (1909, p. 141).

> In general the best results were obtained with crystals that formed at room temperature and did not dissolve on slight increase of temperature, but in many cases all the observations had to be made near freezing temperatures. This was done by working in a cold room, at a temperature near 0° C. Even at such a temperature the heat of the body or breath often produced partial solution of some of the crystals, so that the measurements had to be made rapidly (p. 147).

On the whole, however, the work is a culmination of nineteenth century research on blood crystals, including its characteristic limitations.

R&B used a specialized crystallographic device that combined numerous functions. They describe how the microscopic study of crystals "has resulted in the evolution of a form of microscope which is at once a goniometer, a polariscope, and an instrument for measuring optic axial angles—in short for determining the physical crystallographic constants of small crystals" (1909, p. 145). After haemoglobin crystals from a given sample were prepared, these were put under the microscope to determine their plane angles, the relations of the parts of composite crystals to one another, their pleochroism (i.e., change in color at different angles), the positions and relative values of their light elasticity axes (a refractive property), and the position and angle of inclination of their optical axes.

Two sample pages provide an adequate representation of the content of their study (Fig. 19; Fig. 20). For each species, the above crystallographic features are described in great detail, accompanied by illustrations and photomicrographs in the appendix. Within each section, summary tables allowed for comparisons, though any remarks to this extent are left to the book's conclusion. There, R&B provide answers, of a sort, to problems raised as far back as Lehmann (1855). But even these remain unsatisfactory.

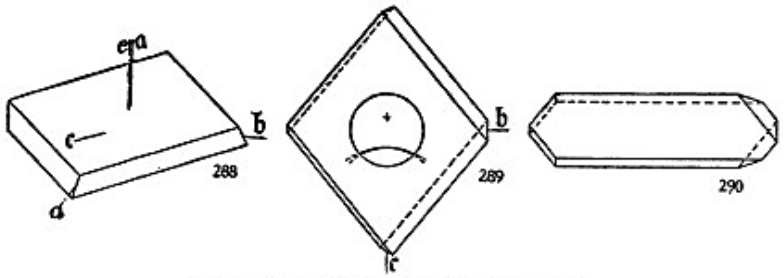
case of the black bear, the crystals of the first crop were mostly trillings, and those of the second crop were mainly single, untwinned crystals. The color of the blood and of the crystals was decidedly brownish, but the spectrum was that of oxyhemoglobin. The solution probably contained met-oxyhemoglobin, which seems to be common in bear blood, at least during the winter.

Oxyhemoglobin of Ursus maritimus.

Monoclinic hemimorphic (monoclinic sphenoidal): Axial ratio $a : b : \dot{c} = 1.2088 : 1 : \dot{c}$; $\beta = 73° 2'$ (calculated).

Forms observed: Unit pyramid ($\bar{1}11$), imperfectly developed; unit prism ($1\bar{1}0$), prism ($2\bar{3}0$), in twins only; basal pinacoid (001); also rarely (100) ? or (101).

Angles: Traces of the unit prism on the base, edges 110–$001 \wedge \bar{1}10$–$001 = 79° 12\frac{1}{2}'$, average of a number of measurements; $110 \wedge 001 = 77°$ (about); $\bar{1}11 \wedge 001 = 48° 30'$. The angle of the dome (101), or orthopinacoid (100), on the base was not determined.



FIGS. 288, 289, 290. *Ursus maritimus* Oxyhemoglobin.

Habit tabular on the base, the crystal consisting of the basal pinacoid (001) cut by the unit prism at one end of the ortho-axis and by the unit pyramid at the other end (text figures 288 and 289). The first crystals to form are of this habit, porous-looking tabular crystals that are proportionately rather thick, but they do not attain very large size. They are usually in trillings as already described for the black bear oxyhemoglobin (text figures 286 and 287) and this kind of twinning is normal in the bears. As these crystals of the first crop disappear, they are succeeded by crystals of the second crop, in which the base (001) and the unit prism (110) are well formed, and usually have sharp angles and smooth faces; but the unit-pyramid planes are generally imperfect in these second-crop crystals. Nevertheless, the angle of the unit pyramid on the base was obtained from crystals of the second crop. The tabular crystals of the first crop are proportionately much thicker than those of the second crop, the ratio of the long diagonal of the plate to its thickness being about 5 : 1; while in the second-crop crystals this ratio will average nearer to 25 : 1. The crystals of the second crop produce parallel growths, but do not twin, as do those of the first crop. A few crystals were observed in the second crop that were prismatic on the ortho-axis, by development of planes in the zone 100–001, which seemed to be (100) or (101), and (001) (text figure 290); but the angle on the base was not measurable, owing to the position in which the crystals were lying.

Pleochroism was rather marked in positions where the axis of greatest elasticity appeared in the section; the colors were: a nearly colorless, b and c about equal and deep red, somewhat brownish-red. The crystals of the first crop, which look soft and porous, do not show as strong double refraction as those of the second crop; the extinction on the flat is symmetrical with the outline of the plate; on edge views of the plate the extinction runs up to 20° with the axis a, or the long dimension of the plate. The orientation of the elasticity axes is $c \wedge a = 20°$, in the obtuse angle; $b = b$, $a \wedge \dot{c} = 0° 6'$ (calculated); or a is parallel to $\dot{c}$ as nearly as can be measured. On the flat aspect, in convergent light,

Figure 19: A summary of properties of crystals from black bear blood (Reichert & Brown, 1909).

TABLE 42.—*Crystallographic characters of the hemoglobins of the Rodentia.*

Name of species.	Axial ratio $a:b:\dot{c}$, etc.	Prism angle or traces of prism on base (normals).	Angle β.	Extinction angle.	Optical character.	System and class.	Substance
		° ′	°				
Sciurus vulgaris	.....	60 0	90	0°	Weakly positive	Hexagonal or pseudohexagonal	α-OHb.
Do	0.577 : 1 : $\dot{c}$	60 0	90	0°	Negative	Orthorhombic	β-OHb.
Sciurus rufiventer neglectus	.....	60 0	90	0°	Weakly positive	Hexagonal	OHb.
Sciurus carolinensis	0.577 : 1 : $\dot{c}$	60 0	90	0°	Negative	Orthorhombic	OHb.
Sciuropterus volans	.....	60 0	90	0°	Do.	Hexagonal	OHb.
Tamias striatus	0.9246 : 1 : 0.589	85 30	90	0°	Positive?	Orthorhombic	OHb.
Cynomys ludovicianus	.....	... 0	90	0°	.....	Do.	OHb.
Marmota monax	.....	60 0	90	0°	Negative	Hexagonal or pseudohexagonal	α-OHb.
Do	.....	... 0	..	0°	Positive	Orthorhombic	β-OHb.
Do	1.804 : 1 : $\dot{c}$	122 0 (58* 0)	near 90	$c \wedge a = 11°$	Negative	Monoclinic	γ-OHb.
Castor canadensis	1.732 : 1 : $\dot{c}$	120 0 (60* 0)	78	$a \wedge a = 8°$	Positive	Do.	OHb.
Fiber zibethicus	1.6318 : 1 : $\dot{c}$	117 0 (63* 0)	68	$c \wedge a = 15°$	Negative	Do.	OHb.
Mus norvegicus albino	0.7829 : 1 : 0.7332	76 7	90	0°	Positive	Orthorhombic	OHb.
Mus norvegicus	0.7829 : 1 : 0.7332	76 7	90	0°	Do.	Do.	α-OHb.
Do	1 : 1 : 1	90 0	90	..	Isotropic	Isometric	β-OHb.
Mus rattus	0.7829 : 1 : 0.5864	76 7	90	0°	Positive	Orthorhombic	OHb.
Mus alexandrinus	0.7829 : 1 : 0.5880	76 7	90	0°	Do.	Do.	OHb.
Erethizon dorsatus	0.5543 : 1 : $\dot{c}$	58 0	56	$b \wedge a = 20°$	Do.	Monoclinic	α-OHb.
Do	0.8170 : 1 : $\dot{c}$	78 30	90	0°	Do.	Orthorhombic	β-OHb.
Cavia cutleri, domesticated variety	.....	.....	90	0°	Do.	Do.	OHb.
Hydrochœrus capyvara	1 : 1.8184	90 0	90	0°	Negative	Tetragonal	α-OHb.
Do	.....	.....	90	0°	Do.	Orthorhombic	β-OHb.
Lepus cuniculus	0.643 : 1 : 0.797	65 30	85	$a \wedge \dot{c} = 15°$	Positive	Monoclinic	α-OHb.
Do	0.5317 : 1 : $\dot{c}$	56 0	90	0°	Do.	Orthorhombic	β-OHb.
Lepus europæus	0.6588 : 1 : 0.8069	66 0	85	$a \wedge \dot{c} = 15°$	Negative?	Monoclinic	α-OHb.

*True angle of traces of prism on base.

Figure 20: A table summarizing the crystallographic characters of rodent blood (Reichert & Brown, 1909).

R&B's strategy for inferring structural properties of haemoglobin from crystal form were based entirely on the principle cited earlier, that "*Substances that show differences in crystal structure are different chemical substances*" (1909, p. 145). The major claims asserted in their study on this basis are given here:

> It was found in the case of many species that the fresh blood would first crystallize in one form of oxyhaemoglobin; that later a second crop of crystals would appear having a totally different habit and even crystal system, or, in other words, different constitution; and that sometimes this would be succeeded by a third crop having a still different form (p. 323).
>
> The crystals obtained from different species of a genus are characteristic of that species, but differ from those of other species of the genus in angles or axial ratio, in optical characters, and especially in those characters comprised under the general term of crystal habit, so that one species can usually be distinguished from another by its haemoglobin crystals. But these differences are not such as to preclude the crystals from all species of a genus being placed in an isomorphous series (p. 327).
>
> In large molecules like those of the haemoglobins, *plasticity of the molecule* is very likely; moreover, there is no doubt from the recorded observations of the practical *plasticity of the*

> *crystal structure* […] Their elementary composition may be various or they may be stereoisomers of the same centesimal composition, but all are connected by the common nucleus hemin, whose crystals show angles that belong in the same isomorphous series (p. 332).

At a foundational level, R&B did not significantly advance chemical knowledge of haemoglobin. Their findings, admirably thorough at the level of each animal species, primarily served to better secure familiar chemical results. Their work gave firm crystallographic grounds for thinking, with Preyer, that interspecies variation in crystal form corresponded to some form of chemical differentiation with phylogenetic implications. However, their inferences from crystallographic to constitutional differences were compromised by a fundamental vagueness around the notion of "constitution." While their claims are broadly consistent with present-day observations of intra- and interspecies variation in the amino acid sequence of haemoglobin, R&B do not—nor could they—explain how features corresponding to "constitution" or being "different chemical substances" would account for the observed differences between haemoglobin crystals. The assumed link between crystal properties and chemical structure offered little insight on its own, and by the middle of the twentieth century leading haemoglobin researchers had abandoned it (Cf. footnote 76).

The chief developments over this period of study can be summarized in the branching figure below (Fig. 21). Crystal purification, joined with chemical, microscopic, and spectroscopic techniques, allowed for the sequential parsing of the components of blood. Chemical distinctions between types of haemoglobin were also achieved. But at this point, questions came to rapidly outpace the answers offered by available techniques. Nineteenth century crystallographic methods provided a platform for speculations about the chemical nature of haemoglobin, but they could offer no real insight into the structural features of molecules. Instead, generalizations were restricted to the features of the crystals themselves. R&B's talk of the "the haemoglobins" used updated chemical nomenclature to retread familiar uncertainty over the chemical significance of the different crystal forms.

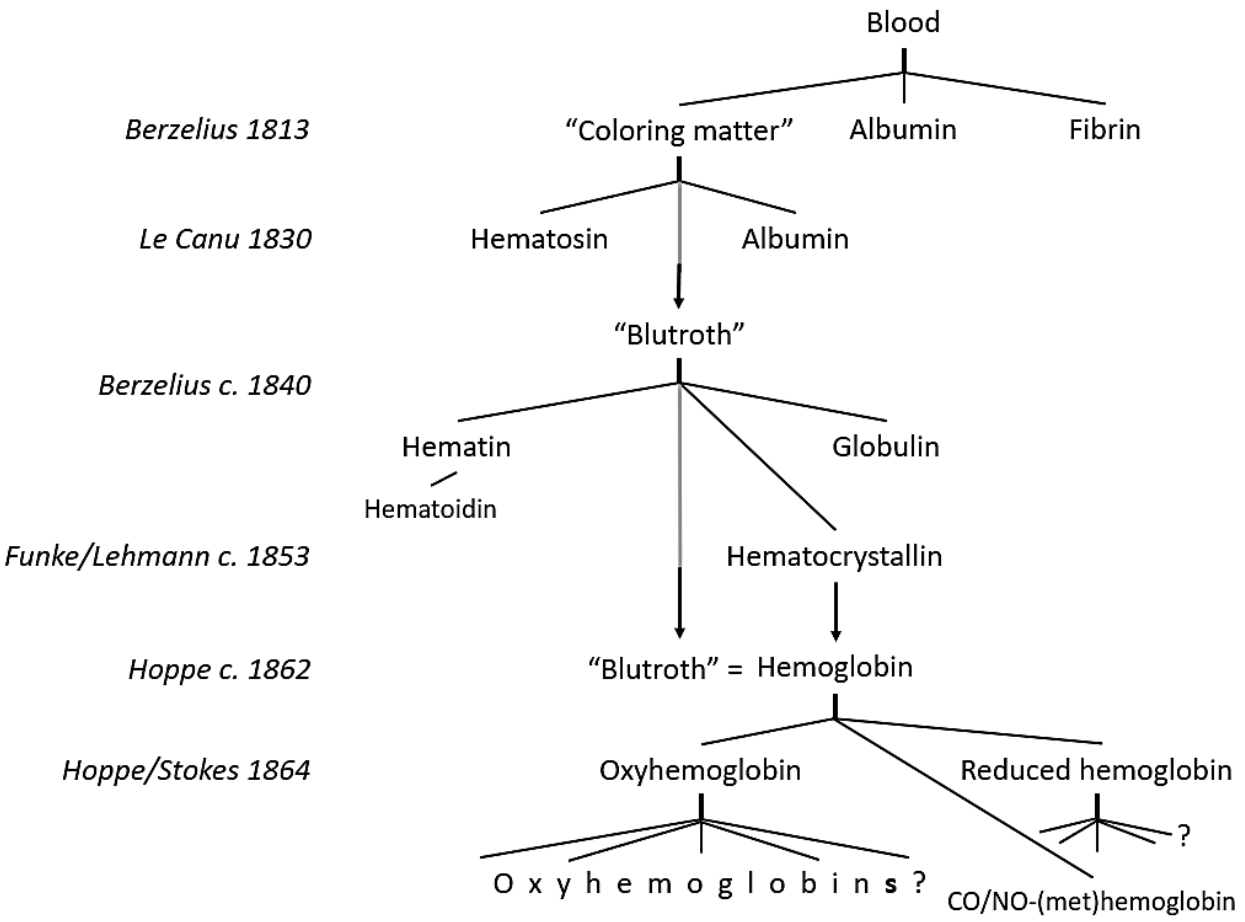


Figure 21: Sequential parsing of the components of blood.

4.1.6. Crystallization after Reichert and Brown

Reichert was a physiologist by training and Brown a mineralogist. Their pairing reflected the dominant mode of inquiry into blood crystals in the prior century. Much like this mode of inquiry, they were also aging in 1909; by 1920, both had exited academia. In the meantime, biochemistry emerged as an independent field in the United States. Crystallographic work was less celebrated here than in the physiological tradition. This may have been due to the extensive labor required from crystallization. When researchers turned to novel substances, they were often confronted anew with the uncertainties affecting the early days of haemoglobin crystallization. Despite a growing variety of animal and plant proteins

identified after 1850, only three noteworthy cases—egg albumin, horse serein albumin, and milk albumin—had been successfully crystallized by 1899. Further successes were slow to appear.

Inspecting the optical and geometrical properties of crystals also gave limited insight into the chemistry of their constituents. Given the facts of polymorphism and isomorphism in crystals, it was unclear how reliable crystal geometry was for understanding chemical structure, regardless of improvements in crystallization techniques. In retrospect, one simple reason for the failure of crystallographic insights into chemical structure is the size and complexity of the molecules in question. An individual arrangement of thousands of atoms, capable of assuming different structures at different scales, was far beyond anything chemists were capably theorizing at the time. But there were also more foundational difficulties. Doubts about the purity of the crystals were never fully quieted. Worse, the scholarly community was riven over whether proteins were ultimately of the same nature as the atomic structures studied by inorganic chemists.

In 1861, Thomas Graham had published a paper that marked the beginning of a sustained movement in physiological chemistry. In a paper on the diffusion of liquids in substances, Graham distinguished two forms of matter: the crystalloids, capable of taking on a crystalline form, and the colloids, for which the "condition of matter" was marked by "physical aggregation […] with chemical indifference" (1861, p. 183). These two forms, he wrote, "appear like different worlds of matter […] The distinction between these kinds of matter is that subsisting between the material of a mineral and the material of an organized mass" (p. 220). Animal substances were a clear instance of such a non-mineral 'organized mass.' Physiological chemists under Graham's sway came to regard their research subjects in colloidal terms.

This perspective meshed with an emerging view of living cells as "wholly or almost wholly composed of undifferentiated protoplasm" (Foster, 1880, p. 1), characterized by vital properties such as contractility, irritability, and assimilative and secretory capacities.[88] Graham provided a physico-chemical basis for this view by construing substances like animal proteins as colloids. This meant he had to massage the phenomenon of protein crystallization into the colloidal framework:

[88] Early influential protoplasmic theories of life were introduced by Max Schultze in the late 1850s and further popularized by Ernst von Brücke and Thomas Henry Huxley in the early 1860s.

> extreme departure from its normal condition appears to be presented by a colloid holding so high a place in its class as albumen. In the so-called blood-crystals of Funke, a soft and gelatinous albuminoid body is seen to assume a crystalline contour. Can any facts more strikingly illustrate the maxim that in nature there are no abrupt transitions, and that distinctions of class are never absolute? (Graham, 1861, p. 223).

By the twentieth century, even physiologists who worked with crystals, like Reichert, were under the colloidal influence. He wrote of the cellular protoplasm as "consisting essentially of an extremely complex solution of interacting and interdependent colloids and crystalloids" (Reichert & Brown, 1909, p. iii) and claimed that haemoglobin is "amorphous in the corpuscles, or is 'colloidal,'" under certain conditions (p. 338). This, after all, was "the typical condition of living matter" (p. 337).[89]

Facing difficulties in understanding vital phenomena in terms of large molecular structures, physiologists began to see more promise in colloidal theories of the cell. These offered numerous interpretations for phenomena produced through new filtration methods and visible under the ultramicroscope (Deichmann, 2007).[90] Carl Nägeli, an early colloidist working in Munich, used the term 'micelle' to refer to constituents of biological colloids, hypothesized to be the aggregative units that made up the elements of protoplasm, including proteins. These instrumental and conceptual shifts, paired with the belief that colloidal substances had different chemistry from crystalloids, contributed to crystallography's decline (Sumner, The Chemical Nature of Enzymes, 1964).

The domineering approach of some colloidists is displayed in 1917's *The World of Neglected Dimensions*, a summary of lectures given during a tour of North American universities by the preeminent German chemist and Nobel laureate Wolfgang Ostwald. In this work he proclaimed that

> the physical and physico-chemical conditions necessary for life cannot be more accurately or more concisely summed up than in the words *all life processes take place in a colloid system*. The colloid state is the means of integrating biological processes. More correctly expressed, only those structures are considered living which at all times are colloid in composition (Ostwald, 1917, p. 155).

[89] These statements bear the influence of the distinction between "living" and "dead" proteins. This originated with Pflüger in the 1870s, who postulated that proteins removed from the cell underwent a chemical transformation. In the following decades theorists such as Bunge and Neumeister voiced a related skepticism that the chemical basis of life could be understood outside the cell (Fruton, 1999). On a grand scale, such debates were reverberations of foundational disputes between physicalists and vitalists.

[90] The ultramicroscope allowed for observations of particulate elements in colloids, whose shape and organization could be more readily studied (Cf. (Zsigmondy, 1966)).

For Ostwald, colloids were systems existing in an intermediate state of dispersion between coarse solution and molecular solutions. Their properties thus resulted from aggregative behavior at a higher scale than molecules. Importantly, Ostwald claimed that there was no definite connection between chemical constitution and colloid state (p. 38). Instead, their behaviors were to be studied kinetically, in terms of "characteristics of gelation, of swelling, of coagulation, etc." (p. 82). For example, muscular contractions could be explained in terms of alterations in the dispersive properties of colloids (Deichmann, 2007). Likewise, the ability of haemoglobin to carry oxygen was hypothesized by Ostwald to be a form of adsorption—a mechanical process to which colloidists routinely appealed (Fruton, 1999).

Applying hindsight, a student of American biochemist James Sumner viewed this as a reduction of proteins to "vaguely defined colloids with no exact internal structure or molecular weight" (Dounce & Allen, 1988, p. 318). But in the first decades of the 20th century, proponents of the colloidal theory held significant sway over biochemists and physiologists. In a 1927 speech on catalysis, Munich chemist Richard Willstätter credited Ostwald's approach to the topic with stimulating new research in the field. Willstätter had developed techniques of purifying enzymes through adsorption, and through this technique had produced an enzymic sample that tested negative for proteins. By the time of this speech, he conceived of enzymes as a mixture of molecular and colloidal elements, "a specifically active group and a colloidal carrier," with alterations in the potency of the active group resulting from variations in the colloid complex (Willstätter, 1927, p. 1374). The difficulty of separating enzymes from proteins was attributed to the tendency of substances of high molecular weight to cling together.

These statements came one year after a publication in which James Sumner claimed crystallized enzyme urease to be a protein. In 1917 Sumner, recently hired by Cornell, had embarked on purifying urease. He had studied under Otto Folin at the Harvard Medical School, a leading researcher on the chemistry of blood and urine. Following this line of inquiry, Sumner engaged in biochemical studies of muscle, blood, and urine, from which he developed an interest in urease (Sumner, 1946). Jack bean meal had recently been found to be incredibly rich in the enzyme, which Sumner suspected to be a globulin and

hence relatively simple to purify (Sumner, 1926). After poor results with adsorption techniques, he focused on one of the few alternative methods known to be effective for purification—crystallization. His work helped renew interest in the application of this technique to biological materials. Much of this interest was further prompted by the discovery of a groundbreaking technique for analyzing crystal structure with X-rays.

4.2. The development of X-ray crystallography, 1895-1935

It is now commonplace to model X-rays as comprised of a continuum of periodic transverse waves. This model remains a principal tool for protein structure determination, despite some fifty years of attempts at ab initio protein folding (Mitchell & Gronenborn, 2017). Yet this model had very few advocates for the better part of two decades following the discovery of X-rays. In this section, I discuss the relationship between the acceptance of this model and the body of empirical results accumulated over this time, culminating in Lawrence Bragg's explanation for Friedrich and Knipping's diffraction results as the effect of a wave continuum. I then review the development of X-ray crystallography that followed, up to the invention of Fourier methods for direct structure determination.

4.2.1. The technical groundwork

In a well-known episode from 1895, Wilhelm Röntgen discovered that a high voltage cathode ray tube, completely shielded with black cardboard, still produced a fluorescent glow in a film of photoreactive material. Cathode rays, produced when an electrical field is run through a vacuum tube, had been a subject of physical research since the 1860s, and several variants of specially outfitted vacuum tubes had become recognized laboratory instrumentation by this time. Röntgen had acquired several tubes, including some

designed by Crookes (Fig. 16, left), and possibly a high-quality tube designed by Philipp Lenard, a correspondent who had recently worked on cathode rays with Heinrich Hertz (Glasser, 1936; Seliger, 1995).

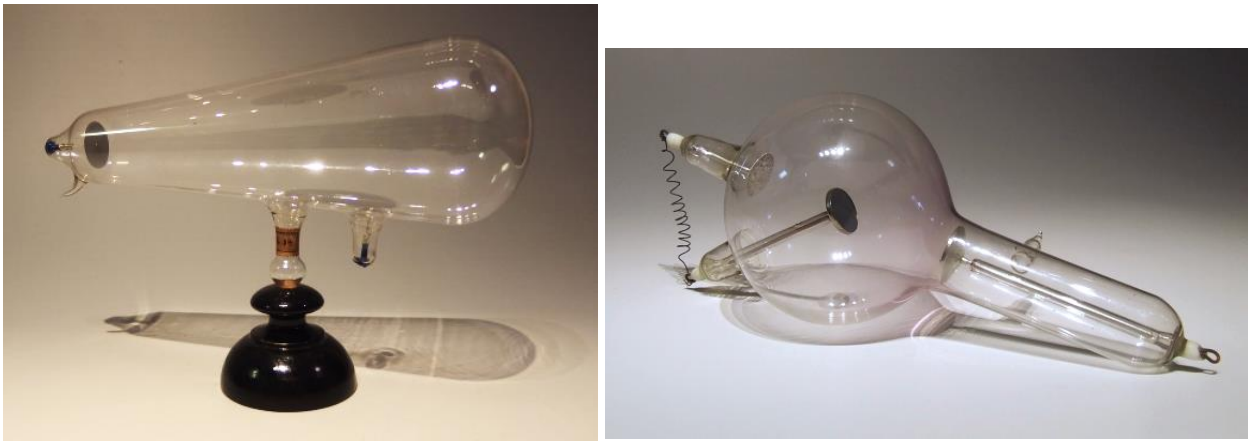


Figure 22: Left: a Crookes-Hittorf tube from 1896, with cathode is on the left side of the image. Right: a dual anode tube (c. 1910), with the central anode plane angled 45 degrees to the cathode. These were sometimes referred to as "focus tubes" with a concave cathode invented by Elihu Thomson to purportedly focus the rays onto a small point of the central anode. Images courtesy of Henk Dijkstra of the Cathode Ray Tube site.

After a series of experiments, Röntgen announced in his paper of the same year that he had discovered "a new kind of rays." The term *strahlen*, or "rays," was already applied to visible light, ultraviolet, and the electrically charged canal or cathode rays. Characteristic among these is the tendency to travel in straight lines unless interrupted. Hence Röntgen wrote, "I find justification for using the name "rays" […] in the very regular formation of shadows that are produced if one brings more or less transparent materials between the apparatus and the fluorescent screen (or the photographic plate)" (1995, p. 374). But these rays were different from others. Foremost among their novel behavior was the ability to differentially penetrate a variety of materials of the same thickness according to their density. Unlike transverse light waves, they could not be reflected or refracted by the same materials, nor was Röntgen able to produce any polarization or interference phenomena (as with diffraction patterns). Unlike cathode rays, they were not deflected when a powerful magnet was placed alongside their trajectory and their absorption was not strictly proportional to the density of matter. These effects thus warranted a superficial characterization of X-rays as highly penetrating linear propagators with no apparent charge or characteristically transverse-wavelike behavior.

Röntgen interpreted his data as the effect of a ray. This made the data into an empirical result. A photographic plate showing squares of varying darkness (data) could be interpreted as the effect of rays

differentially penetrating square sheets of metal interposed between the tube and plate (empirical result). This interpretation was justified by appeal to analogous empirical results characterizing more familiar ray phenomena, primarily the shadow-casting behavior. The tactic was perfectly in step with Röntgen's time. Cathode rays and canal rays had been named in the first two decades after he received his physics degree. In Germany, where the electromagnetic ether theory reigned, these were regarded by analogy to light. In the years following 1895 new forms of radiation were uncovered—α-, β-, and γ-rays, along with the alleged N-rays and various others historical jetsam. This characterization was therefore an act fully recognizable to Röntgen's scientific milieu, warranted by reference to known empirical results.

In a subsequent paper, Röntgen detailed what would become a standard technique for isolating X-rays. He built a box out of lead shielding and thick plates of zinc—a known absorbent of ambient electricity—save for an airtight aluminum window through which X-rays could pass. Using this set-up with an X-ray tube, he found that rays entering the box would produce a charge in neutral air, as observed by a discharge apparatus. By this point Röntgen had designed a modified vacuum tube in which the cathode rays would impinge upon a tilted anode made of a small sheet of charged metal (Fig. 22, right). He found that different anode materials resulted in "more or less" rays being emitted, with platinum producing the strongest signal. Two years later he published findings that the degree to which X-rays would penetrate materials (or their "hardness") depended on the pressure of residual gases in the vacuum tube (Röntgen, 1897). Further, he and Georges Sagnac (Sagnac, 1899) found irradiated metals could produce fluorescence on a plate that was shielded from the X-ray source, thereby inferring that these metals were induced to emit their own "secondary" rays. Thus the qualities of the rays appeared to be affected in multiple ways, both by properties of the incident materials and those of the source tube.

X-ray technology was adopted rapidly, spawning a diversity of tubes (Fig. 23). For many practical purposes, there was little need to understand the nature of the rays beyond these earliest findings. A 1910 textbook by Mihran Kassabian, a pioneer on the rays' medical applications, quotes heavily and almost exclusively from Röntgen's first papers in its summary of the basic X-ray properties (Kassabian, 1910). On the other hand, the book attests to an improved body of technical skills and knowledge of the

experimental apparatus. One early contribution came from the engineer E. Wilbur Rice, who found his image focus noticeably improved after introducing a lead plate with a small aperture in the immediate path of the rays ((Thompson, 1896), Sec. 106). This addition—essentially a beam collimator—was soon adopted in virtually all X-ray experiments. In the same year, Stine tested one of Röntgen's claims about the source of the rays by repeatedly irradiating and comparing the results of photographic plates positioned at different points surrounding a tube. This confirmed that the source of the rays was the point where the cathode rays struck the material opposite them, not necessarily the anode (Thompson, 1896, Sec. 108).

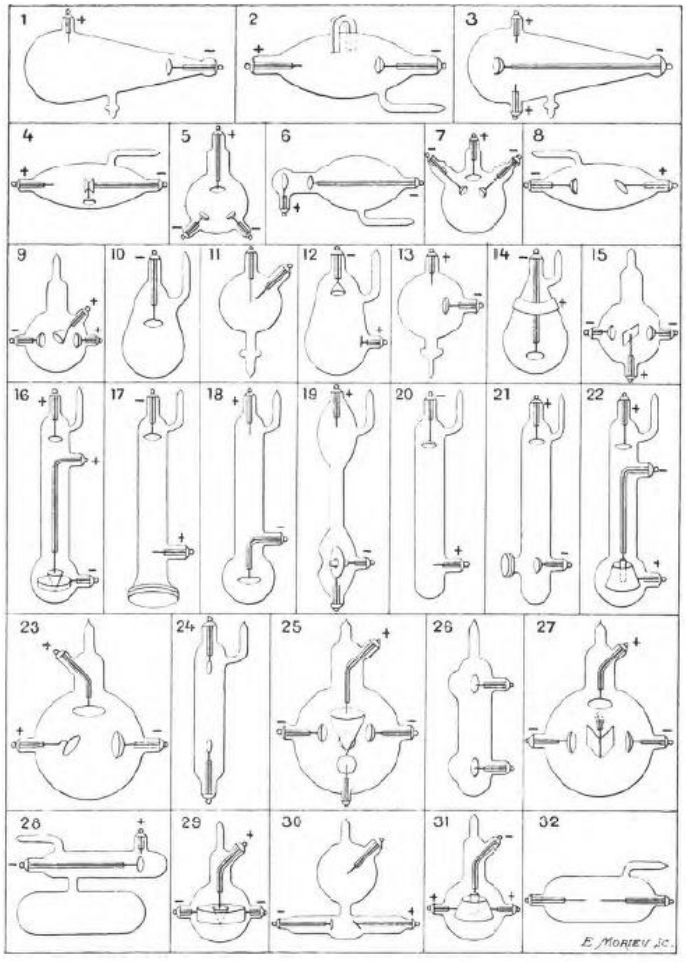


Figure 23: An assortment of tubes shown in a popular science article by the instrument maker and experimentalist Gaston Séguy (1897), including designs by Crookes (nos. 1, 2, 20) and Röntgen (nos. 24 and 32).

Despite widespread manufacture, X-ray tubes were often unreliable. The quality of the vacuum was inconsistent, not just between tubes but within the same tube over a period of use.[91] The gradual decrease in tube pressure hardened the X-rays emitted, which led to a host of additional problems. Harder rays provided less contrast in the resulting image, could puncture the glass of the tube, and led to a build-up of static electricity on its surface capable of damaging the tube (or its user) in discharge. Two ways of tackling these issues stand out in Kassabian's textbook. The first is the use of a regulator—a metallic wire "wick" extending into the tube from outside through an airtight plug that would release hydrogen into the tube when heated. This "softened the tube," in the parlance of experimentalists (Bragg W. L., The Development of X-ray Analysis, 1975). The second involves a variety of eyeball tests, examining the coloration and distribution of gas within the charged tube to evaluate its quality (Fig. 18).

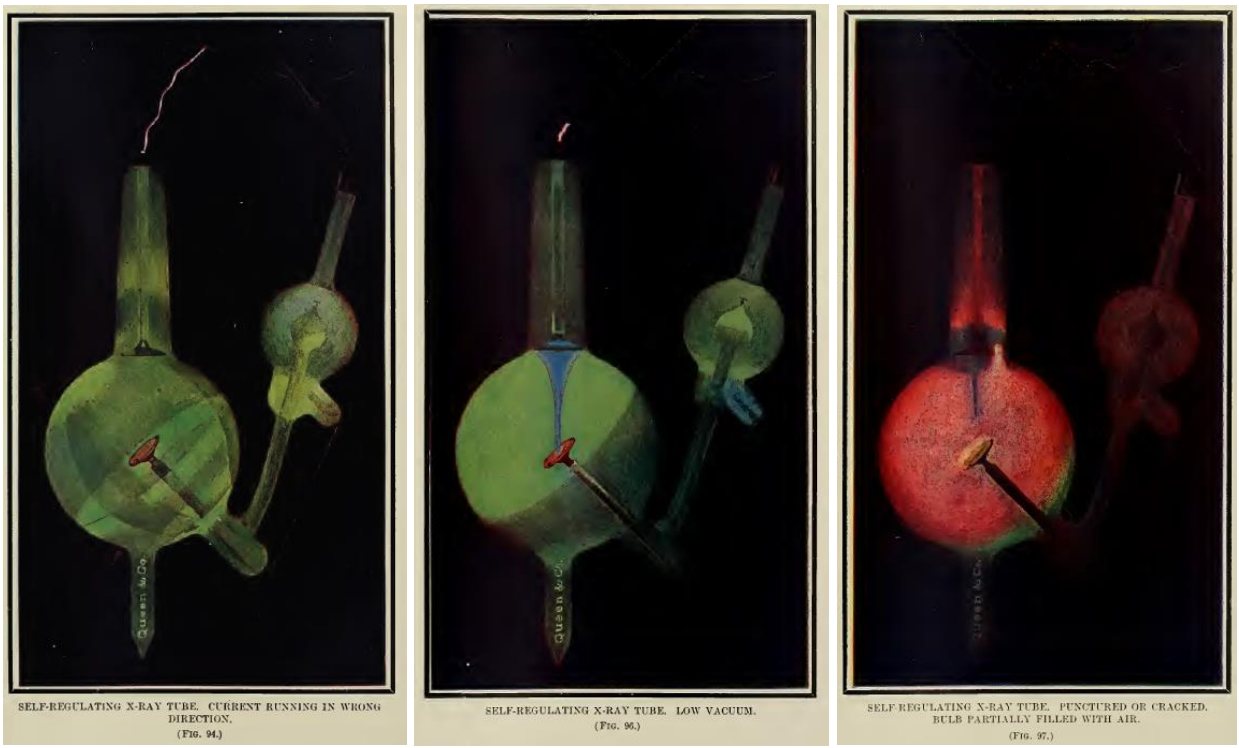


Figure 24: Plates from Kassabian (1910) showing, from left to right, a backward-running current producing illumination on the wrong side of the tube, low vacuum (i.e., high gas pressure) producing a blue light, and a cracked bulb.

[91] Mould (1995) recounts an 1896 examination of the life span of a single tube by London hospital physicist Charles Phillips.

In short, experimentalists developed a body of knowledge oriented around the operation of the X-ray apparatus. In addition to the ray properties summarized by Röntgen, this included a refined understanding of the following factors:

- Conditions for preparation and production (source of emission, vacuum pressure required, induction coil strength),
- Factors modifying X-ray properties (vacuum pressure-ray hardness relation, use of regulators, anode materials-beam intensity relation), and
- Factors affecting X-ray data (lead collimator, ray hardness, interactions with gas, interactions with photoreactive materials).

The roles of these factors could be summarized in a set of generalizations, such as those found in textbooks, that collectively provide a causal account of the experimental apparatus. Technical work guided by this theory of technique secured the reliability of empirical results. This is because the theory itself did not extend far beyond a summary of accumulated empirical results, with which the behavior of the apparatus was expected to cohere when producing further data. If the data was of poor quality, it provided a range of options to test and amend this: was the tube properly positioned? The distance between the cathode and anode correct? The vacuum pressure too low? A Duhemian web, embodied in the working knowledge of the apparatus, was woven around the core characterization of X-rays established by Röntgen.

It could be said that this theory is representational, in a sense that would apply to any inferential system of concepts with empirical content. However, the knowledge described here is not dependent on a representational model. There is no single inferential tool that captures the experimental set-up and is manipulated as something over and above the setup itself. Most causal generalizations within the theory of technique take the form of a dependency relation between the manifest behavior of one component and another within this setup: overheating the regulator results in the tube glowing blue; a blue glow result in a lack of photographic image. Unlike the path of model-based inference, when an inferential chain ceases within the theory of technique, there is no further question about connecting this result to something else; the result is an imperative to act. Still, these technical activities are anchored to a minimal, superficial

characterization of a target phenomenon insofar as they constitute an experiment *on X-rays* qua highly penetrating linear propagator, insofar as heating the regulator changes the penetrating property *of these rays*, and so on.

4.2.2. Competing characterizations

While this technical skillset developed, theoretical clarity was lagging. Kassabian's review of X-ray properties nears its conclusion by stating, "The nature and origin of the Röntgen rays is as little understood to-day as when first discovered" (Kassabian 1910, p. 151). This was not for a lack of theoretical models. Given their production by cathode ray tubes, it was natural for physicists to seek electromagnetic models of X-rays. This led Röntgen to hesitantly suggest his non-diffracting rays might be the sought-after longitudinal ether waves. Another theorist, Schuster, soon undermined the evidence favoring this model, arguing that high frequency transverse waves may also fail to diffract, refract, or reflect under the same conditions (Wheaton, 1983). For a moment, the transverse wave model drew advocates, only to lose them to a new contender—the ether impulse. British physicists at the time suspected cathode rays consisted of charged material particles, a kind of intermediary between the purely wavelike elastic ether, and the mechanics of neutral matter. Electromagnetic theory predicted that a charged particle, when rapidly decelerated to a stop, would produce a radiative disturbance in the ether. In 1896, Stokes modeled this effect as a single spherically expanding sinusoidal "pulse" in the ethereal medium, like the shape traveling down a cracked whip. The X-ray beam, he proposed, "consists of a vast succession of independent pulses" of this nature (1898, p. 54). The paper introducing this model concluded with a theoretical explanation for the lack of observed diffraction fringes from X-rays. Being single pulses, they lacked the periodic oscillatory structure responsible for the production of optical diffraction effects.

The following year, J. J. Thomson published work generally accepted as proof of the material nature of cathode rays, which he promptly connected to the production of X-rays following Stokes's lead

(Thomson, 1898). The German physicist Arnold Sommerfeld built on this work in 1901, providing a rich mathematical explanation for the appearance of a diffraction effect observed by Haga and Wind (1899). The Dutch team had sent X-rays through an initial slit followed by a trapezoidal slit tapering from 14 μm at the top to 2 μm at its base (Fig. 25, left). Using a square-shaped pulse model (Fig. 25, right), Sommerfeld (1901) showed these pulses would produce no diffraction fringes but subtly broaden the initial slit line in a manner consistent with Haga and Wind's image. With advocates among elite physicists in Britain and Germany, the ether pulse model prevailed over others.

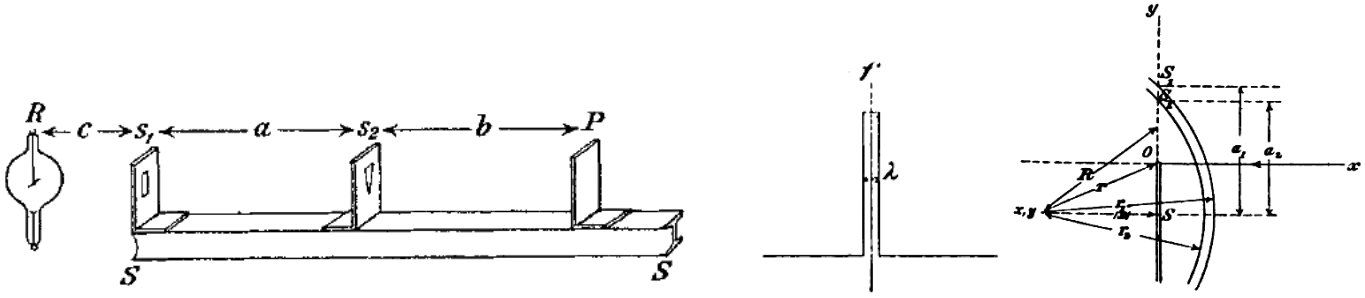


Figure 25: Left: the Haga and Wind diffraction experiment, as reproduced by Walter and Pohl (1909). Right: the radiating pulse shape posited by Sommerfeld to explain the Haga and Wind result.

A curious interplay of support for model and experiment was at work here. Haga and Wind's photograph, like many of its time, was of poor quality and difficult to interpret (Fig. 26). It was produced from over 100 hours of exposure, resulting in significant noise and, most likely, inconsistent ray quality. It was published against a background of failed attempts to demonstrate X-ray diffraction (e.g., Thompson 1896, Sec. 110 recounts attempts by Stine and Perrin) and its interpretation was disputed. A repetition of the experiment ten years later by Walter and Pohl (1909) was said to show no effect—prompting replies from Wind and Sommerfeld. As Bruce Wheaton recounts, "by itself the photograph submitted to support this claim would not have been convincing" (1983, p. 31). The main argument in favor of Haga and Wind's result was that they controlled for optical effects blamed for prior failures, though in Wheaton's assessment the real force behind its positive reception was coherence with Sommerfeld's treatise. At the same time, empirical results influenced what was included in Sommerfeld's model and what was not. For instance,

when calculating behavior at the edge of a slit, he assumed there was no contribution from pulses reflected off the material, as no results had shown this to take place.

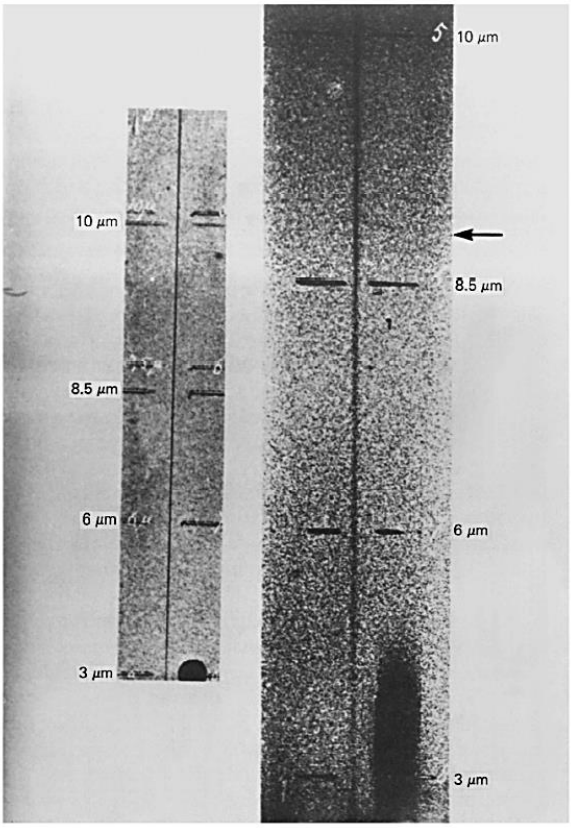


Figure 26: Photographic image from Haga and Wind's (1899) experiment, with width of the slit marked on the side. They emphasized the apparent blurring and widening of the line when the photographic plate was placed at a distance behind the second wedge-shaped slit (right image), compared that from a single slit (left). Note also the fading of the right line image toward the bottom.

Fundamental research slowed on the continent due to the low quality of photographic data and lack of tools for analyzing them. Meanwhile, researchers uncovered new results in Britain by exploiting the ionizing power of X-rays. Charles Barkla was foremost among them; he demonstrated that the rays could in fact be polarized and added further content to the empirical distinctions between ray types (Fig. 27 shows a typical set-up). This came through a demonstration that the secondary rays emitted by metals included a kind that was completely uniform in its penetrating power. These were called “homogeneous,” by contrast to the mixtures of harder and softer rays emitted by standard tubes. Both results secured further confidence in characterizing the rays as ether phenomena. Polarization was a typical behavior of light-like phenomena

and the nature of secondary rays was highly reminiscent of effects produced by incident light on fluorescent matter.

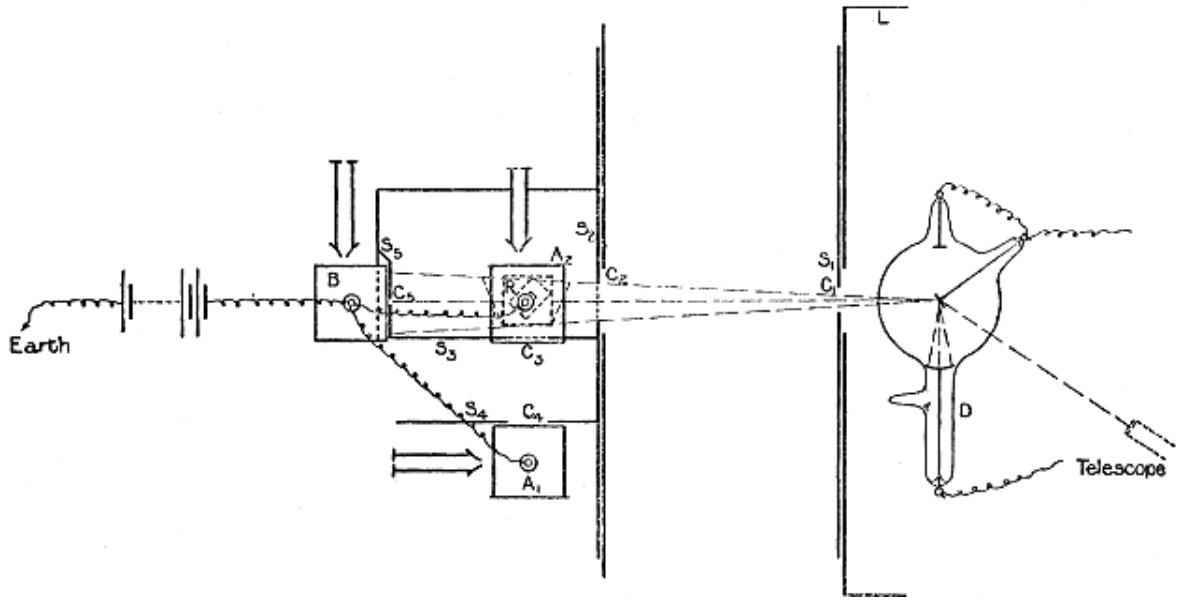


Figure 27: Schematic of Barkla's (1905) first polarization experiment. Electroscopes attached to ionization chambers at A_1, A_2, and B measure the strength of x-rays scattered forward and at right angles to the beam path. The experiment tests an analogy to light, for which polarized components preferentially scatter at right angles, giving a weaker signal than the forward scattering.

Opposition to this model came from Adelaide, Australia. In 1907, William Henry Bragg (WHB)[92] proposed a corpuscular theory. According to WHB, X-rays and the similar-acting γ-rays consisted of "neutral pairs" of positive α and negative β particles.[93] While both models could account for the penetrating power of X-rays—WHB by appeal to the pairs' neutrality, Barkla by large molecules' lack of response to impact from minute pulses—WHB noted a special difficulty for the ether pulse theory in ionizing effects. As these pulses radiated, the energy they carried would be distributed over a thin shell of a larger and larger sphere, rapidly becoming too diffuse to dislodge an electron from a gas molecule. Neutral pairs carried enough concentrated energy for isolated impacts, which cloud chamber photographs appeared to depict. WHB created further trouble for the ether pulse account by reporting a weaker X-ray signal from back-

[92] For clarity, I will hereon refer to William Henry Bragg as WHB, and his son William Lawrence Bragg as either Bragg or Lawrence Bragg.
[93] For an account of WH Bragg's corpuscular theory, see Steuwer (1971).

scattered X-rays compared to forward-scattered (Fig. 28), an asymmetry that was not predicted by electromagnetic theories.

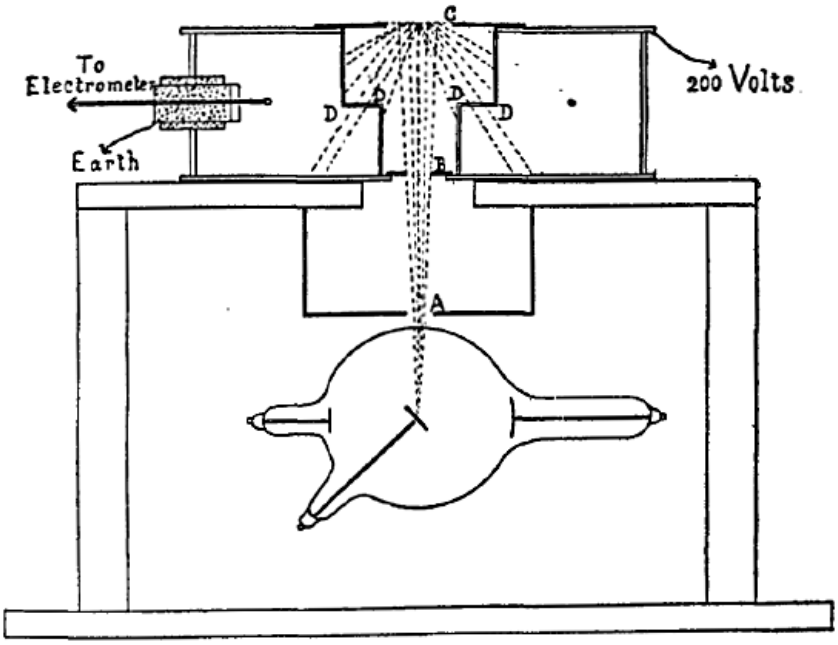


Figure 28: WH Bragg and Glasson's (1907) experiment comparing ionization measurements of back-scattered rays (produced when a metal sheet is placed at C) to forward-scattered rays (sheet placed at B).

A feud between Barkla and WHB emerged, each vying to illuminate faults in the other's model. WHB struggled, for instance, to explain phenomena like the production of secondary rays on his account, for which the fluorescence analogy came naturally to ether pulse modelers. Neither, however, defended a periodic wave model. Material particles were taken to behave mechanically, whereas ether pulses were independent and discontinuous disturbances in the ether with none of the extended structure required for periodicity. In fact, since the physical pulses lacked a periodic structure, the related notions of spatial frequency and wavelength could not apply. This led to confusion among Sommerfeld's readers, some of whom took his work to provide an estimate of the wavelength of X-rays based on diffraction results.[94] Though he used the symbol for wavelength, λ, this referred only to the *width* of single pulses (as seen in Fig. 25, center). By taking their Fourier transform, a pulse could be viewed as a spectral sum of

[94] This includes Lawrence Bragg's (1975) recounting of Sommerfeld's derivations.

"wavelengths," but for Sommerfeld this representation was non-physical; diffraction for X-rays could not be explained as a pattern of destructive and constructive interference.

This explains Sommerfeld's lack of enthusiasm in 1912 when a colleague in Munich sought out his laboratory assistants, Friedrich and Knipping, for an X-ray interference experiment. Recent results had begun to suggest that X-rays and light, the latter typically modeled as transverse waves, were not as different as first thought.[95] A conversation with Paul Ewald inspired Max Laue to see whether the atoms of a crystal could function as a diffraction grating for short-wavelength X-rays. Munich had been home to Leonhard Sohncke, a theorist of crystal symmetries, and it was commonly accepted that crystals were comprised of a regular, repeating molecular structure (Forman, 1969), so the experiment appeared well-motivated. But to Sommerfeld, the vanishing line image at the base of Haga and Wind's photograph signified that X-rays could not pass through a 2 μm slit, much less a gap between crystal atoms. Further, any X-ray pulses that might pass through would be of varying widths and scatter in every direction, producing only a uniform darkening of the photograph (Wheaton, 1983). He strongly discouraged Laue's experiment on these grounds.

A model of X-rays as comprised of many periodic waves was not widely accepted among theorists. Yet Laue proposed that the homogeneous secondary rays emitted by an irradiated crystal might have a more uniform structure capable of producing a diffraction pattern. This hunch was based on Ewald's work (following Lorentz), which studied the optical effects of crystals modeled as a lattice of resonators that would collectively vibrate when stimulated by incident light (Ewald, Fifty Years of X-ray Diffraction, 1962). Laue, Friedrich, and Knipping selected a crystal of copper sulfate, believing this would maximize the homogeneity of the secondary rays, and set to work (Fig. 29). They initially placed photographic plates on each side of the crystal, parallel to the beam line, under the assumption that its molecules would emit spherical waves.[96] When they failed to produce a clear image, other arrangements were tried until they

[95] For instance, two years prior researchers found that the photoelectric effect had the same forward scattering asymmetry that WHB found in X-rays (Wheaton, 1983).
[96] This assumption was also due to Ewald's resonator model.

found that a plate positioned behind the crystal yielded a crude, but roughly symmetrical two-dimensional pattern. When the plate was moved back, the image spread out. When the crystal was ground up, the data was uninterpretable. The effect thus appeared to spread out from internal structure of the crystal. A repetition with zincblende crystal irradiated for many hours produced more clearly resolved data (Fig. 30, left).

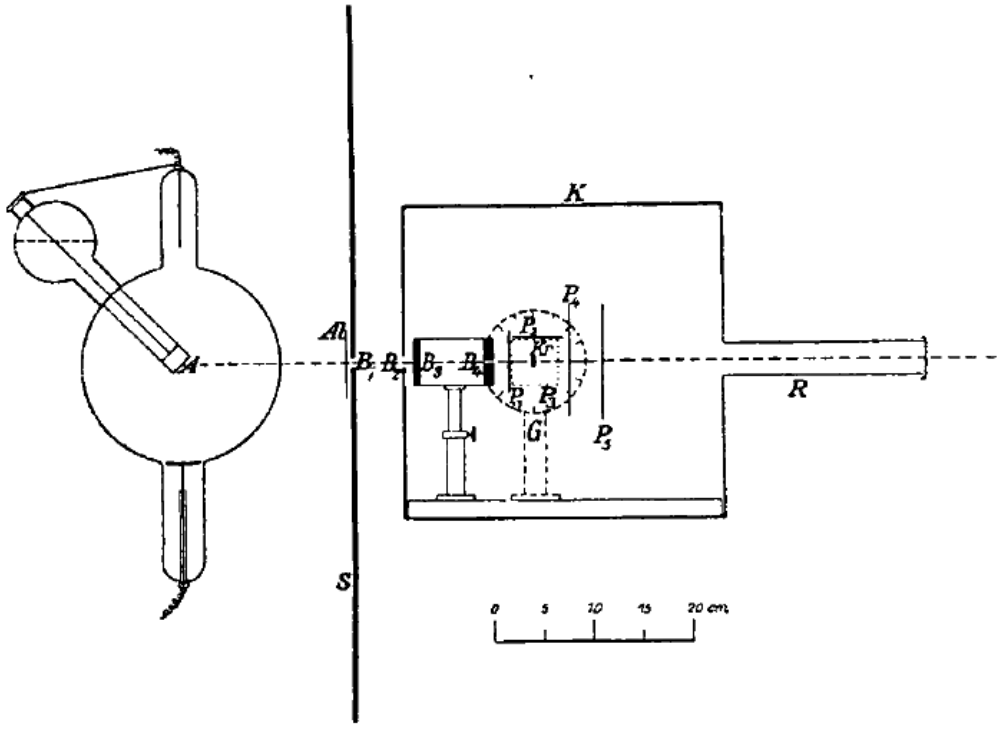


Figure 29: Friedrich and Knipping's experimental set-up (Friedrich, Knipping, & Laue, 1913). The crystal Kr is located within a lead box K. After passing through an aluminum window and several collimators (to shield off secondary rays), the X-ray beam strikes the crystal and is scattered onto the plate P_4 placed behind it.

Laue promptly developed a mathematical model of diffraction in three dimensions, seeking to interpret the pattern of spots as the maxima of interfering secondary waves from the crystal lattice. Using this model, he showed the proportions between (i) the coordinates (x, y) of each spot on a plate at a distance z from the crystal should match those between (ii) the scattering angles (α, β, γ) of the contributing wave in each dimension. Laue used these to derive fixed ratios for integer parameters (h_1, h_2, h_3) that could be used to determine a wavelength λ responsible for this spot via the "Laue conditions":

$$h_1\lambda = a\alpha \;;\; h_2\lambda = a\beta \;;\; h_3\lambda = a(1-\gamma), \;\; where \;\; \alpha{:}\,\beta{:}\,\gamma :: x{:}\,y{:}\,z$$

One merely had to take the values x, y, z for recorded spots, find the corresponding ratio α:β:γ, and then derive the ratio of α:β:(1 - γ) to see how the h's were related. However, Laue's method did not uniquely determine his results. For any ratio α:β:(1 - γ), infinite triplets of integers h_1, h_2, h_3 might satisfy it. To account for this, Laue decided to stick to small integer values of h and assumed that the secondary radiation from the crystal consisted of only a small number of wavelengths. He provided five that produced a similar pattern to that found in the photograph from zincblende, ZnS (Fig. 30, right).

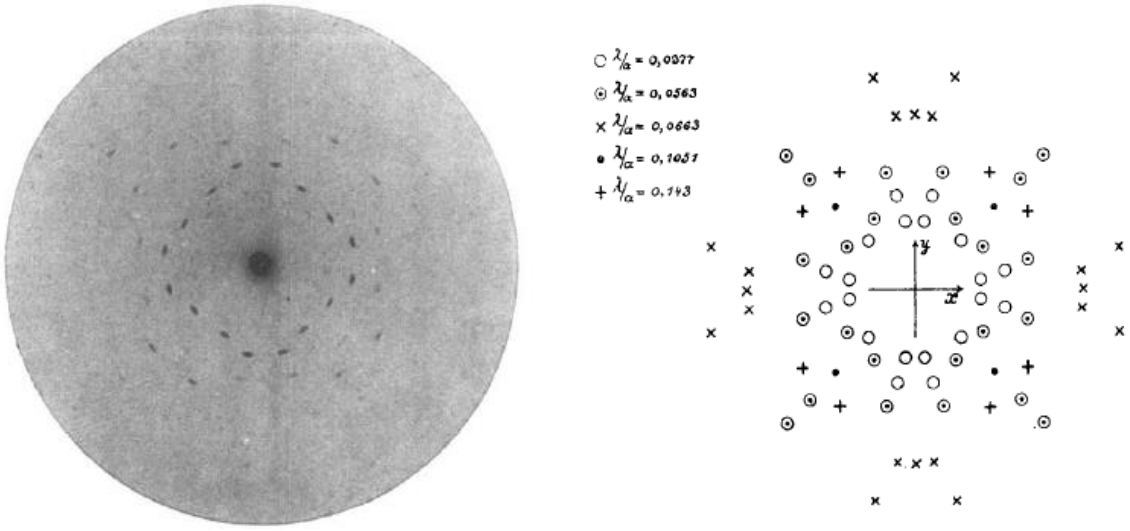


Figure 30: Laue's derivation (Friedrich, Knipping, & Laue, 1913). Left: the diffraction pattern produced by zincblende crystal. Right: Laue's reconstruction of the data using five selected wavelengths.

Even with these results on hand, there was resistance to a model that allowed for X-ray interference. Barkla expressed strong skepticism in a letter to Rutherford (Forman, 1969). Laue followed Sommerfeld even after his experiment, claiming in 1913 that an incident beam consisting of a continuum of wavelengths would only darken a photograph (Ewald, 1962).

William Lawrence Bragg, the elder WHB's son, was unsatisfied with this result. By assuming only five wavelengths contributed to the recorded result, Laue had arbitrarily restricted the freedom provided by his model. Further, these same values chosen by Laue also predicted some spots that were absent from the photograph. These left the adequacy of Laue's model in question. As Bragg put it, "there are a great many combinations of h_1 h_2 h_3 which might give spots on the photographic plate which are in fact, not there, and there is no obvious difference between the numbers h_1 h_2 h_3 which correspond to actual spots, and those

which are not represented" (1913, pp. 44-45). While studying physics at Cambridge, Bragg encountered models developed by William Barlow and William Pope, two of England's foremost theorists on symmetry and atom packing in crystals. These showed that within any crystal one could locate evenly spaced parallel planes of repeating atomic arrangement. Bragg raised the possibility that X-rays might be reflected off these planes and was encouraged to attempt it with mica. A report of his success was published in *Nature* that year, noting that "variation of the angle of incidence and of the distance of plate from mica left no doubt that the laws of reflection were obeyed" (Bragg W. L., 1912, p. 410). These findings led to a model for diffraction that remains in use today.

Bragg's solution exploited a mathematical equivalence, well-known in British optics, between the diffraction of light modeled in one way as discontinuous pulses and in another modeled as a continuous range of wavelengths. The conditions under which incident X-rays would form a constructive maximum (i.e., a diffraction spot) could then be derived as a relation—later dubbed "Bragg's law"—between their constituent wavelengths and the angle at which they reflect off parallel planes of the crystal. In other words, a portion of each incoming ray would be reflected off an atom in a plane while another part may pass through. If the distance between planes, the angle of the incident ray, and its wavelength was correctly balanced, a ray reflected off a lower plane would have traveled just the right distance to be in phase with those reflected from a higher plane, resulting in constructive interference. Different atomic planes within the crystal will be tilted at different angles with respect to the beam line, depending on the crystal's symmetries, and these will select for reflections of different wavelengths within the ray, forming a pattern of distinct spots corresponding to the internal structure of the crystal. Assuming a particular symmetry group for the zincblende and using this model, Bragg was able to account for Friedrich and Knipping's data, spot for spot. In a later summary, he claimed to show "that Laue's pictures were made by a continuous range of X-ray wavelengths, a kind of 'white' radiation" (1975, p. 25).

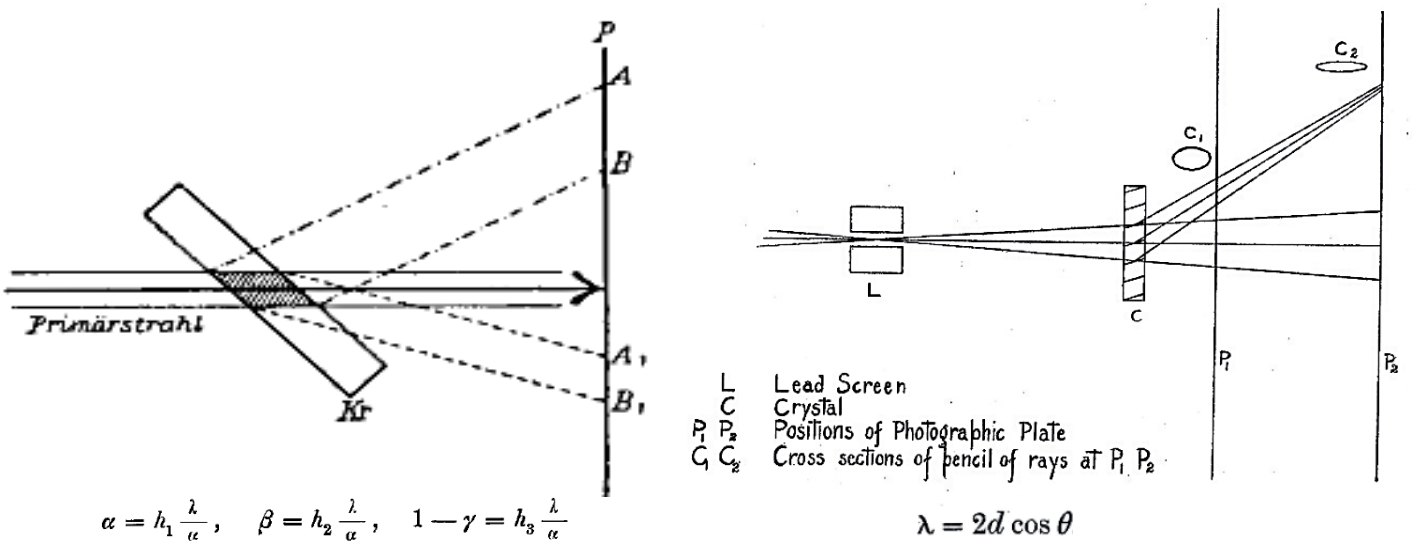


Figure 31: Two models of the diffraction experiment. Left: Laue's diagram showing the primary rays producing secondary radiation from the crystal and forming spots off the beam axis at angles corresponding to their wavelength. The core mathematical contribution of this model are the Laue conditions, giving three parameters (the h terms) relating scattering angle (greek letters) and inter-atomic distance, a, to ray wavelength (Friedrich, Knipping, & Laue, 1913). Right: Bragg's diagram showing X-rays of a preferred wavelength reflected off crystal planes to the same point. Bragg's Law, below, relates the wavelength to the angle reflection and the distance, d, between crystal planes (Bragg 1913).

Bragg's model had helped initiate new research programs in crystallography and spectrometry—the first canonical applications that required knowledge of X-rays beyond Röntgen's initial results—which he and his father avidly pursued. Compelled by these results, physicists across Europe rapidly adopted this view of X-rays as comprised of a continuum of wavelengths. In a review at the end of 1913, Laue and Sommerfeld's research assistant, Friedrich, claimed this work showed "the presence of a continuous spectrum" emitted from the X-ray tube (1913, p. 1081). Following Bragg's announcement, Barkla and a collaborator sought to replicate crystal diffraction effects (Allen, 1947). In an early statement of their findings, the former skeptic wrote, "we cannot at present suggest any probable explanation except the very obvious one of interference" (Barkla & Martyn, 1913, p. 647). The Braggs, however, insisted on the incompatibility of further diffraction experiments with the ether pulse model: "Since the reflection angle of each set of rays is so sharply defined, the waves must occur in trains of great length. A succession of irregularly spaced pulses could not give the observed effect" (Bragg & Bragg, 1913, p. 436). Barkla ultimately internalized this lesson; his Nobel speech of 1920 contains no references to ether pulses, instead characterizing primary rays as consisting of a range of wavelengths (Barkla, 1967). Barkla's sparring partner, the elder Bragg, likewise admitted defeat for his prized model. Commenting on the papers of Laue

and his son, WHB wrote, “If the experiment helps to prove X-rays and light to be of the same nature, then such a theory as that of the ‘neutral pair’ is quite inadequate to bear the burden of explaining the facts of all radiation” (Bragg W. H., 1912, pp. 28-29).

4.2.3.Reception of X-ray diffraction models and X-ray techniques

This historical process shows a mutual interplay between evaluations of models and empirical techniques. A skilled experimentalist would recognize the procedural steps taken by Friedrich and Knipping as part of an accepted theory of technique. The alignment of dual anode tube, aluminum window, multiple collimating apertures, lead shielding, and photographic plates (Fig. 29) were the standardized repertoire for producing reliable X-ray results, developed through an accumulation of causal generalizations over laboratory arrangements. These were the basis for characterizing X-ray phenomena in greater depth during the period leading up to 1912. They established a clear target for model-based inferences, which were in turn licensed by their consistency with electromagnetic or mechanical principles. As inferential tools, these models allowed for predictive inferences beyond familiar results. With motivation from models, experimentalists drew on their theory of technique to construct novel data-gathering setups, as seen in the experiments undertaken by competitors WHB and Barkla. New characteristic behaviors, once established in the lab, could then be added to the developing theory of technique. These, in turn, placed new demands on models designed to account for an expanding range of empirical results.

Bragg’s model would not have been accepted by his scientific community if it were not for Friedrich and Knipping’s technical work, in conjunction with a growing body of results demonstrating analogical behavior between X-rays and light. There was hardly any enthusiasm for this model within the community of X-ray researchers prior to 1912. Even after the Munich experiments were publicized, other models were first suggested. Pulse theorists like Barkla questioned whether the data was due to wave interference at all. Yet, given the reliability of their techniques, there was little doubt that their data were

the effect of *some* interactions between X-rays and crystals. Both Laue and Bragg depended on this fact in applying their models. They were confident in abstracting from everything but the relations between the rays, crystal, and photographic plate because Friedrich and Knipping had isolated the interaction between these components in their data. But the model presented by Laue could only reproduce their data by admitting a degree of arbitrariness in its method of application. By contrast, Bragg's model uniquely determined the structure of the data on hand without predicting additional points that were not found in the recorded pattern. These advantages were easily recognized. In fact, it was largely because of Bragg's work that Friedrich and Knipping's data were fully accepted as evidence *of diffraction*. Broader reception of the wave continuum model of X-rays soon followed. In this way, Bragg's model played a constitutive role in the scientific community's characterization of its target phenomenon. Its success depended on confidence that the data were reliably connected to a particular interaction and its ability to simulate the process isolated by these techniques.

4.2.4. The growth of X-ray crystallography as an empirical program

Though Bragg's wavelength continuum model was not the final word on the nature of X-rays, it was able to encapsulate a wide range of characteristic empirical behaviors discovered by Röntgen and others (Fig. 32). In contrast to prior debates between ether and corpuscular models, a plateau of stability had been achieved between this model and the diverse X-ray phenomena. Researchers now had a firm grip on the manufacture of X-ray effects and an encompassing model for how these were produced.[97] A rapid shift in research priorities ensued for many of the central commentators on the 1912 experiment. This stability put them in a position where they could use the X-ray as a tool for investigating the properties of crystals. What had been a means in Friedrich and Knipping's experiment—the crystal as a component of an experimental

[97] To take one example, the distinction between homogeneous and heterogeneous X-rays, previously used to distinguish rays of uniform penetrating power and those formed by a mixture of differently penetrating rays, was now explicable in terms analogous to monochromatic and mixed waves of visible light.

apparatus for the study of X-rays—had now become an end of inquiry. This spurred further investigation of the interaction between X-rays and crystal structure, with a corresponding growth in diffraction methods.

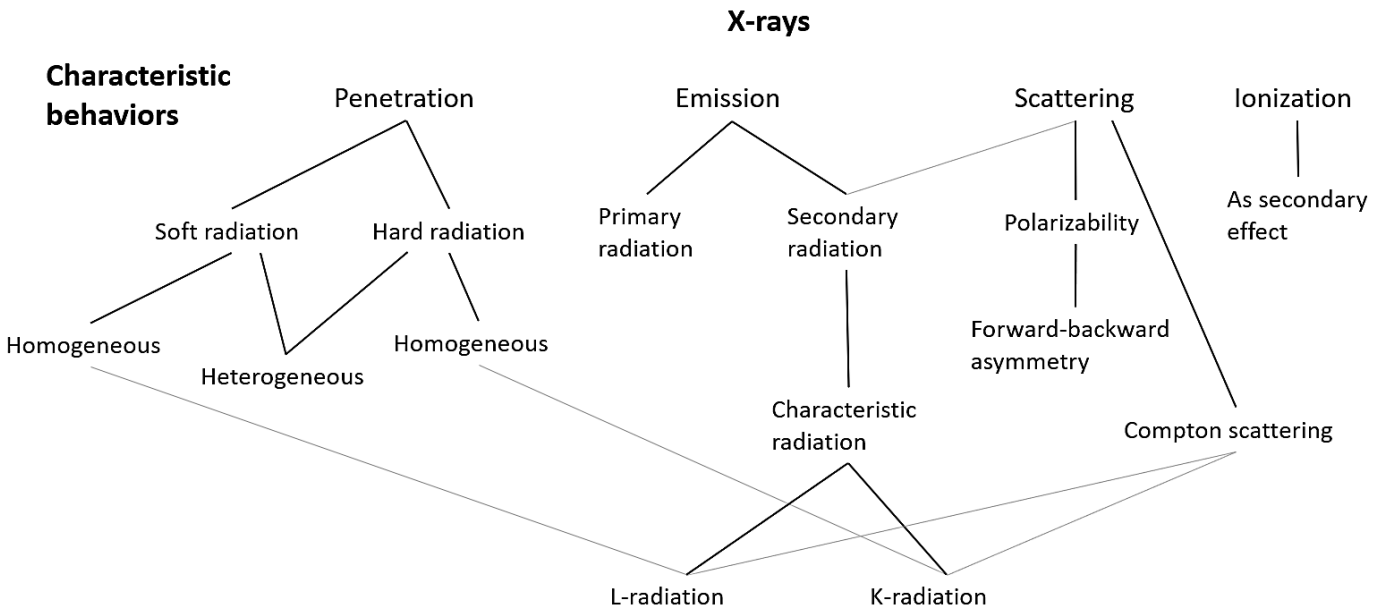


Figure 32: Some refinements of the characteristic empirical behaviors of X-rays from Röntgen into the early 1920s. Vertical lines indicate a genus-species relation between concepts. The gray are included to illustrate how some concepts unified several different forms of behavior. The power of mathematical models lay partly in their ability to encompass these differences.

In simulating his results, Bragg had assumed that the atoms of Friedrich and Knipping's zincblende crystal were arranged as a face-centered cubic lattice. X-ray diffraction data offered, as his father put it, "an illustration of the arrangement of the atoms in the crystal (Bragg, W. H., 1912, p. 28). This was one of the first empirical methods grounded in physical theory for determining the structure of molecules. But it quickly became apparent to researchers that a better understanding of the interaction between X-rays and crystal structure was needed to accurately interpret diffraction data. A series of associated problems would dominate the early years of X-ray crystallography from the publication of Lawrence Bragg's paper into the mid-1920s.

One question concerned the source point, in the crystal, of diffracted waves. In 1913, Bragg published a paper extending his analysis of zincblende structure to other molecules in the halide family—potassium chloride KCl, potassium bromide KBr, potassium iodide KI, and rock salt NaCl. There he

claimed that his results "would seem to indicate clearly the association with each diffracting centre of a single atom" (Bragg W. L., 1914a, p. 64). Noting the variations in the diffraction data, in particular the relative size of the dots in pictures of each, alongside prior results showing that absorption of X-rays increase linearly with the atomic weight of a material, he reasoned that "the weight of the atom in the main defines its effectiveness as a diffracting centre, and that two atoms of equal weight are equally effective" (p. 63). These provisional assumptions—that atoms were the source of diffraction and that their diffracting power was proportional to their size—would guide a line of crystallographic research that sought to measure the distances between atoms, the size of atomic radii, and to test different models of atomic structure.[98]

If X-rays were reflected from atomic centers, reflection intensity varied according to atom size, and reflecting planes of atoms were equally spaced according to Bragg's crystal parameter *d*, then three-dimensional atom-by-atom models of crystals were in principle testable through a process of determining the orientations of the different reflecting planes of a crystal over a range of glancing angles (the angle between the incident rays and crystal face) and measuring the intensity of reflected rays.[99] This would then allow for the measurement of distances between atoms in the crystalline molecule, and might even support inferences about the size of atomic radii.[100] As Bragg's Law made clear, however, precise determinations of the interplanar distance *d* required knowledge of the wavelength of the incident radiation, and this in turn required a better understanding of the structure internal to the diffracting centers—that is, atoms. Bragg recognized this early on: "As long as the complexity of the unit associated with each point of the lattice is unknown, the absolute wavelength [of the incident radiation] cannot be calculated" (1914, p. 70). A more detailed model of the structure internal to atoms was thus one path to more precise determination of crystal

[98] Talk of atomic "size" and "number" had become roughly synonymous within this community, following Moseley's X-ray spectroscopic work in 1913, which established a relation between the characteristic X-ray absorption/emission frequency of elements and their atomic number.

[99] In the Braggs' early crystallographic work this was done using an ionization spectrometer rather than X-ray photographs, some aspects of which are described below.

[100] In 1920 Bragg suggested "the distance between neighbouring atomic centres in a crystal is the sum of two constants characteristic of the atoms." Following his old teacher Barlow's theory, he writes, "The crystal may be imagined as an assemblage of spheres packed together, the constants then representing the radii of the spheres" (Bragg W. L., 1920, p. 37). These notions were pursued by researchers such as Victor Goldschmidt who in 1926 systematically substituted atoms of one kind for others in a crystal preparation. By this method he began to develop a table of atomic radii, regarded as basic building blocks of crystal structure.

structure. As WHB wrote, “We may even conceive the possibility of discovering from their relative intensities the actual distribution of the scattering centres, electrons and nucleus, in the atom” (1915, p. 111).

As early as Moseley’s 1913 study of elementary X-ray emission spectra, there were efforts to compare X-ray data with predictions from Bohr’s (and others’) atomic models. In 1914, Charles G. Darwin used electromagnetic wave theory to derive the energy reflected by an irradiated atom from the assumption that electrons clustered at the center of the atom. He found this differed from a general empirical formula due to WHB by a factor of $1/\sin\theta\cos\theta$ and concluded this was due to some electrons being far from the nucleus. Following suit, Compton (1917) calculated the reflected energy for different distributions of electrons and found a Bohr-type shell structure fit best with WHB’s formula, in contrast to a model by Crehore that located electrons tightly around the atom’s center. Four years later, Lawrence Bragg, Reginald James, and C. H. Bosanquet (1921; 1922) pursued these claims with more empirical specificity by measuring reflected intensities from the chlorine and sodium atoms of rock salt and comparing these results to the intensity curves predicted by three models of electron distribution within the atom (Fig. 33).

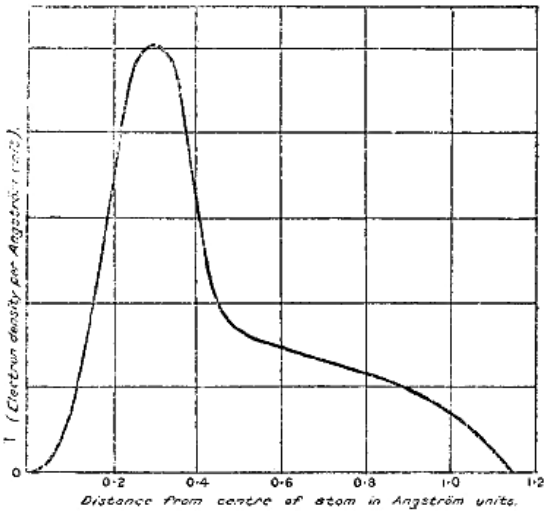


TABLE I.—Sodium.

	Sin θ.	0·1.	0·2.	0·3.	0·4.	0·5.
F	Observed	8·32	5·40	3·37	2·02	0·76
	Shells {0·29 Å, 0·76}	8·56	5·59	3·33	2·19	0·93
	Smooth Curve ...	8·57	5·40	3·20	1·91	1·00

Figure 33: Left: empirical measurement of the distribution of electrons around the atomic center of sodium. Right: comparison to predicted values (Bragg, James, & Bosanquet, 1922)

They found their results agreed best with electrons arranged in a series of spherical shells that oscillated around a mean position.[101] Similarly, WHB (and later figures such as Pauling) used empirical data to probe the nature of bonds between atoms, noting these had a non-uniform structure in diamond: "The structure of the diamond cannot be explained on the hypothesis that the field of force round the carbon atom is the same in all direction," he wrote, "It is necessary, therefore, to suppose that the attachment of one atom to the next is due to some directed property, and the carbon atom has four such special directions" (Bragg W. H., 1921, p. 303). In 1925, Douglas Hartree introduced a more sophisticated method for calculating the amplitude of a wave scattered from an atom, under varying glancing angle, based on the dimensions of their core electron orbits and approximate time they are occupied. He proposed quantifying the scattering power for a given atom in terms of "the ratio of the amplitude of the wavelet scattered by the atom to that scattered by an electron" (Hartree, 1925, p. 297), a contribution Lawrence Bragg hailed as "a great advance on the simple assumption of proportionality between scattering power and atomic number" (1925a, p. 298).

Thus, by the mid-1920s, diffraction data was taken not only to reflect the size and position of atoms, but to vary according to the detailed electronic structure of individual atoms and the bonds between them. At each step, empirical methods marched in tandem with modeling claims. Following Hartree's proposal for modeling scattering amplitudes, for example, it became routine to measure the so-called f-curves of elements and compounds, tracking atomic scattering power (measured by reflected X-ray intensity) under the variation of glancing angles and under variation of incident X-ray wavelength. Such measurements underlie what came to be an essential parameter involved in structure determination, commonly known as the atomic form factor. Depictions of crystal structure evolved in turn (Fig. 34). In the 1910s atom positions were plotted as points on a unit cell, by the early 1920s "space packing" imagery had come into use, and by the later part of this decade these had merged in images showing spheres on crystal cells that included bonding information.

[101] Individual elements could be analyzed in this way under the assumption that rock salt crystals under certain orientations had layered planes comprised of a single element.

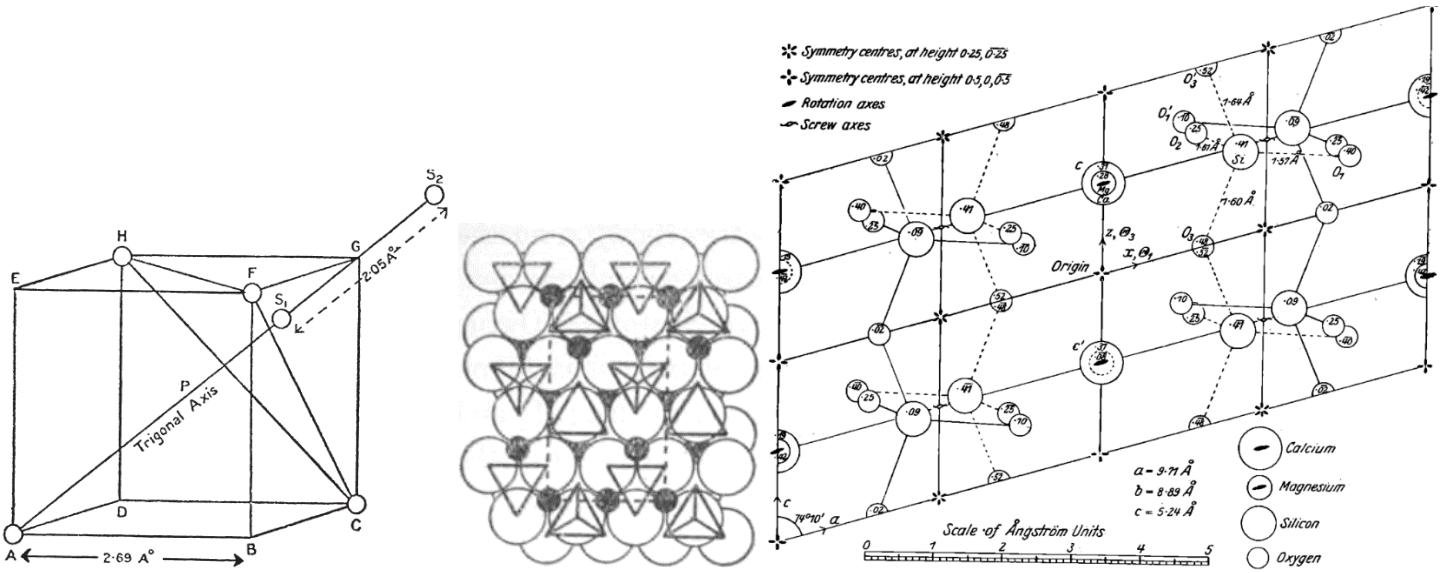


Figure 34: Representations of crystal structure. Left: a lattice model used since the mid-1910s (W. L. Bragg, 1920). Center: a packing model used through the 1920s (Taylor & West, 1928). Right: an atomic bond model that became more common in the late 1920s (Warren & Bragg, 1929).

Other questions were posed but remained unresolved. In 1912, the self-taught crystallographer Alfred Tutton noted that the zincblende photograph from Friedrich and Knipping had symmetries that corresponded to the full cubic system, despite the zincblende molecule having a different form of internal symmetry. He concluded that the photo contains information about the symmetry of the whole crystal, and not stereochemical information. This left open the matter—posed as far back as Haüy—about the relation between the internal structure of a given molecule and that of the crystal it forms en masse. Various proposals were entertained. Shearer (1923) hypothesized that "all the symmetry of the molecule is reproduced in the crystal" and thus the crystal must have at least the amount of symmetry as its constituents. William Astbury and WHB (1925) were more cautious, noting that all the crystallographer has "direct" access to is the symmetry of the whole crystal, including the molecule's environment and points of contact with neighbors; "Whether this can be considered as extending right down into the heart of the molecule is another question" (p. 30).

Related questions on large-scale crystal structure proved more tractable as the theory of technique developed. It could be asked, for instance, how well laboratory crystals conformed to the perfect symmetries of the geometric theory. This question had direct bearing on empirical results, as WHB noted with respect

to a particular measurement procedure. In his spectrometric method, popular in Great Britain, the experimenter would slowly rotate an ionization chamber around an irradiated crystal at a set radius. The chamber's electrometer would show a positive read-out at angles where crystal planes were aligned and constructively reflecting X-rays, and this read-out was recorded as a curve that varied with glancing angle (Fig. 35).

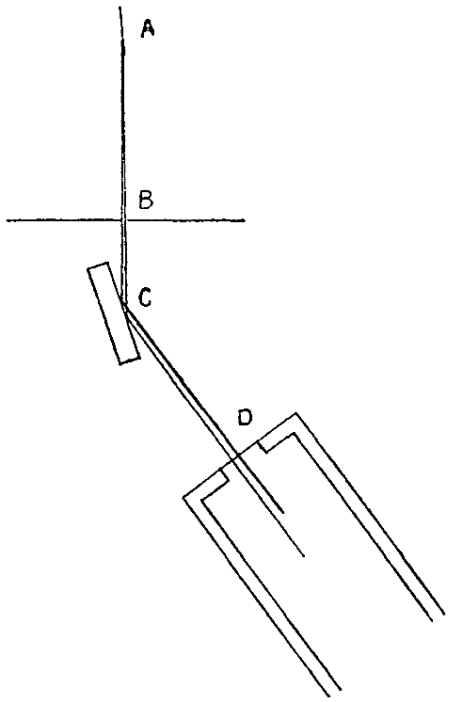


Figure 35: A schematic of ionization spectrometry from WHB (1914). A is the X-ray source, C is the crystal face, and D is the window of the ionization chamber.

Prior practice had involved quantifying the intensity of the reflected rays by recording the peak value of this curve. The elder Bragg commented:

> This method would be as good as the other if crystals were perfect. But they are not. Rocksalt, for example, is very far from being uniform: it consists rather of an agglomeration of smaller and more perfect crystals put together in imperfect alignment […] Since reflexion does not occur within close limits but is spread over a wide range, the highest point of the peak is much lower than it would be if the crystal were perfect, and is no true measure of the intensity of the reflection" (Bragg, W. H., 1914, p. 106).

Instead, WHB recommended his "sweep" method of measuring intensity, which involved integrating the electrometer read-out curve from its initial point of excitation to the angle at which the signal died out. This operationalization of intensity would account much better for crystal imperfections.

Crystal imperfections were one of many factors found to contribute to intensity readings. Shortly after Laue and Lawrence Bragg's initial publications, it was realized that the thermal motion of atoms would cause their positions to deviate from the ideal points of a lattice, "blurring" the area of interference and reducing overall intensity. Using a model of independently vibrating atoms, Debye (1913; 1914) derived a temperature-dependent correction called the "heat factor" that was found by WHB to improve agreement with experiment when applied to predicted intensities (Bragg W. H., 1914). Similarly, the assumption that atoms take up a volume and do not diffract as points meant a crystal's planes of reflection had some width. This meant the total reflection intensity would be spread over a small angle corresponding to this planar thickness, a matter pointed out by Stark (1913) and subsequently dubbed the Lorentz factor. Ewald (1913) demonstrated by experiment that differently polarized reflected waves would not interfere and called for a further correction, the polarization factor. Lawrence Bragg (1914b) reasoned that X-rays were gradually absorbed as they penetrated the crystal such that deeper layers of atoms contribute less to the reflected intensity. This led to a further term known as the extinction coefficient. By the early 1920s, he and others recognized the need for an additional "absorption coefficient" unique to each material, which might account for so-called secondary extinction of reflecting power due to imperfections in crystal structure.

Each of these factors affecting intensity readings were incorporated into the central kinematic formulae of the early theory of X-ray diffraction theory. Laue's inaugural theory of 1912 introduced a key expression for the combined contribution of diffracted waves at a point (α, β, γ) arising from the phase difference of waves emitted from a collection of atoms with coordinates (*x, y, z*) at a distance *r* from the center of a crystal. Laue drew on an expression from Helmholtz for the excitation at a distant point resulting from a spherical wave, which he presented in terms of the familiar mathematical form for a wave: an amplitude term multiplying a complex phase term e^{ix}:

$$\sum \Psi e^{-\frac{2\pi i}{\lambda}(r+x\alpha+y\beta+z\gamma)} /r$$

Beyond some vacillations in the mid-1910s, the general form of this expression remained largely unchanged into the 21st century. Thus Ewald (1914) recalls Laue's expression with some notational differences (e.g., eliding the amplitude term Ψ),

$$\sum = \sum_k e^{2\pi i(h_1\xi_k+\eta_k+\zeta_k)}$$

Fifteen years later, Lonsdale (1929) used a similar expression to investigate the structure of benzene,

$$S = \sum A_i e^{2\pi i\frac{\boldsymbol{x_i}}{d}}$$

and Patterson (1935) used a similar version, with notation familiar to contemporary practitioners:

$$F(\bar{h}) = \sum_{r=1}^{N} f_r e^{2\pi i(h\boldsymbol{x}_r/a)}$$

(where *d* or *a* is the spacing of the repeating unit cells of the crystal). The most significant changes over this period were not in the form of the expression, but in the interpretation of the wave amplitude factor—written as Ψ, *A*, or *f* above.[102] Laue initially regarded this as a function of the wave's direction, but by 1916 described it more neutrally as a "diffraction factor." Darwin (1914), following Bragg's reflection model, incorporated an "arrangement-dependent quantity" into the amplitude factor. By the mid-1920s, following empirical investigations of the source of diffraction in the crystal, the amplitude component was recognized as a quantity associated with the scattering power of a given type of atom (as in (Ewald, 1925)). By the mid-1930s it was called by its contemporary name, the *atomic form factor*. The quantity representing the sum of individual contributions to the intensity reading at a point—variously written Σ, *S*, or *F*—had been recognized since at least 1914 to reflect the arrangement of atoms in the crystal and so came to be known as the *structure factor*. Its central role in X-ray crystallography is summarized by Bertram Warren and Bragg, who wrote that "*F* measures the contribution of the whole unit cell to the amplitude of a wave reflected from the [lattice] planes (*hkl*) […] Each value *F*(*hkl*) is a non-dimensional ratio, characteristic of

[102] The other significant change involved taking *F* seriously as a complex number, and thus as a carrier of phase information, the implications of which researchers such as Bragg did not fully grasp until the late 1920s.

the crystal. The values of *F* provide the data for the determination of the structure" (Warren & Bragg, 1929, p. 323).

In contrast to the stability of the structure factor expression, the theoretical formula for the quantity directly measured by X-ray crystallographers—the intensity of waves diffracted at a given angle—went through several revisions. Laue originally wrote this quantity in 1912 as

$$\frac{|\Psi|^2}{R^2} * \frac{\sin^2 MA}{\sin^2 \frac{1}{2}A} * \frac{\sin^2 NB}{\sin^2 \frac{1}{2}B} * \frac{\sin^2 PC}{\sin^2 \frac{1}{2}C}$$

the first term giving the amplitude-dependence of the intensity (wave optics had established that intensities are proportional to the square of an amplitude, expressed here in terms of the dependence of the intensity on Ψ, the wave amplitude), with the following terms expressing the net interference from combined phases. Two years later WHB drew on the novel reflection theory of diffraction for a simpler expression. He expressed the intensity of waves reflected at a glancing angle θ at a given point on a photographic plate as follows:

$$I_\theta = \frac{A}{\sin^2 \theta} e^{-B \sin^2 \theta}$$

The exponential term was an addition to Laue's expression, standing for Debye's temperature factor. Compton (1917) followed this trend by inserting the polarization factor into an otherwise identical expression:

$$E_r = \frac{C(1 + \cos^2 2\theta)}{\sin^2 \theta} e^{-B \sin^2 \theta}$$

This intensity or "total reflection" formula continued to grow as further corrections were introduced. Drawing on ab initio calculations from Darwin (1914), Lawrence Bragg et al. (1921) derived the following expression for the total energy reflected from a crystal:

$$\frac{E\omega}{I} = \frac{N^2\lambda^3}{2} * \frac{1}{\mu} * \frac{1}{\sin 2\theta} * F^2 * \frac{e^4}{m^2c^4} * \frac{1 + \cos^2 2\theta}{2} e^{-B \sin^2 \theta}$$

The same polarization and temperature factors are seen on the far right, to which the following considerations are added: the wavelength λ of the incident radiation, the number of atoms *N* per unit volume,

intensity of the incident beam I, the velocity ω at which the irradiated crystal is rotated, the absorption coefficient of the crystal μ, the Lorentz factor (the $\sin(2\theta)$ term), and a factor due to Thomson approximating the scattering power of an electron (third from the right). Bragg and his coauthors noted that all terms in this expression except F^2, the squared structure factor, could be measured or approximated for simple compounds such as rock salt. A path was thus open for deriving a quantity from intensity measurements, F, that corresponded to the characteristic atomic structure of crystals.[103]

To arrive at this equation, Bragg et al. had assumed that each unit cell of the crystal was slightly misaligned with its neighbors. This was presented as a theory of a fully "mosaic" or "ideally imperfect" crystal. In a subsequent paper, Bragg, Darwin, and James (1926) compared this expression with the total reflected intensity from a perfect crystal (a quantity proportional to F rather than F^2). These two expressions, they reasoned, gave the upper and lower limits for X-ray intensities due to crystal scattering. In 1934, James compared these predictions to actual measurements and found that all crystals were somewhere between the perfect and ideally imperfect limits, but most tended toward a more mosaic structure. Twenty years after WHB raised the issue, crystallographers had found a way to leverage intensity formulae to give a general measure of imperfections within crystal symmetries.

Other imperfections were not intrinsic to the crystal, but recognized as a result of its preparation. Bragg, James, and Bosanquet (1921) insisted on the importance of mechanical grinding to flatten the faces of crystals instead of relying on cleavage alone, going to the length of measuring the effect of cleavage errors on intensity measurements. This was one of many developments in crystallographic procedure facilitated by an understanding of the experimental apparatus. For instance, researchers recognized that the anticathode in the X-ray tube—the component that cathode ray electrons struck to generate X-rays—would emit radiation of different penetrating power depending on its constituent metals. Starting with Lawrence Bragg and WHB (1913), they developed means to estimate the characteristic wavelengths emitted from the

[103] While lattice spacings could be inferred from properties of diffraction photographs—which was often sufficient for solving simple crystals—the only quantitative means available for intensity measurements was the ionization spectrometer developed by WHB as described above.

anticathodes of different types. This allowed experimentalists to better measure the effect of different wavelengths of radiation on the scattering behavior of atoms. Similarly, studies were performed to measure the effect of feeding different voltages to the X-ray tube on atomic scattering power (as in (Wyckoff, 1930)). Second-order knowledge of this type, developed through investigation of the X-ray apparatus, clearly informed the presentation Bragg, James, and Bosanquet (1921) gave of their experiment. They describe how their ionization chamber was filled with methyl bromide and treated with a potential of 320 volts between its outer walls and inner electrode in order to prevent recombination of ions during measurement. This electrode was connected to a Lutz-Edelmann string electrometer adjusted to read centivolts. The incident beam was generated from a Coolidge bulb—known for operating at a consistent intensity—charged by an induction coil running at 50,000 volts, its anticathode consisting of "a button of rhodium embedded in a tungsten block" (p. 267).[104]

As technical knowledge developed, so did the number of methodological choices available to crystallographers. A range of data production methods were developed. The old methodological differences between continental and British X-ray experiments[105] lived on in the contrast between the production of "Laue diagrams" from photographic plates and use of the ionization spectrometer by researchers like the Braggs, Moseley, and Darwin. To these were added the powder diagram method of Debye and Scherrer (1916) in which crystals were ground up and a photo was produced that captured reflections from all planes at once; the rotation method of Polanyi (1921)) in which a crystal was rotated within a cylinder of photographic film, and subsequent methods (such as (Weissenberg, 1924)) for coupling rotations with moving film. By 1927 the methods and applications of X-ray crystallography had multiplied enough to deserve review of their costs and benefits. This was provided by WHB's student John Desmond Bernal. As

[104] Intensities were measured using a "null method" variation on WHB's sweep approach: when the electrometer reading began to rise at a certain angle, experimenters would slowly increase the voltage applied to the outer wall of the ionization chamber until the meter returned to zero, then they would rotate the crystal more and repeat the process, until no further signal was detected. The total energy reflected from a plane was then quantified in terms of the net voltage applied over the angle of rotation. As seen in the above equation, this was divided by the intensity measured from the incident beam alone in order to determine the absolute intensity due to reflection.
[105] See section 2.2.

he put it, “it becomes profitable at this stage to treat of instruments made for this specific purpose rather than to leave to each investigator the burden of creating special apparatus for his own needs.” (Bernal, 1927, p. 273). To this end, he presented an account of the general experimental procedure for X-ray crystallographers:[106]

> (1) Measurement of spacings of planes of known indices leading to the determination of the size and shape of the unit cell.
> (2) Measurement of crystal density from which with (1) the number of molecules per cell can be found.
> (3) Determination of the symmetry class by the methods of ordinary crystallography.
> (4) Determination of the indices of absent reflections leading with (3) to the determination of the space group; and this combined with (2) giving the molecular symmetry.
> (5) Measurement of the intensities of the reflections from planes of known indices, and of the correcting factors necessary to obtain the structure factors, from which, together with (4), the complete structure may be derived (Bernal, 1927, pp. 275-276).

alongside a comparative table for the major methods in use (Fig. 36):

[106] A similar series is found in Bernal and Crowfoot (1933), albeit with more emphasis on the use of Fourier methods.

	Laue Photograph	Rotation Photograph	Ionization Spectrometer	Powder Photograph	Ionization Spectrometer (powder)
Applicability to crystals of different sizes	Medium 0.1–.01 cm	Medium 0.1–.01 cm	Large > 0.1 cm with developed or ground faces	Small < .01 cm	Small < .01 cm
Applicability to crystals of different symmetry	All	All	All, but troublesome with mono- and triclinic crystals	Only cubic tetragonal, trigonal and hexagonal crystals	Only cubic tetragonal, trigonal and hexagonal crystals
Number of reflections observed	Very many, usually hundreds	Many, up to two hundred	Any number, but troublesome and slow, usually about forty	Usually not more than forty, very crowded	Very few
Determination of indices of reflecting planes	Simple and certain	Troublesome, but certain with care	Certain, but troublesome, especially for general planes	Almost impossible, except for planes of low indices	Almost impossible, except for planes of low indices
Measurement of spacing	Impossible	Highest obtainable accuracy	Accurate	Fairly accurate but for overlapping	Fairly accurate but for overlapping
Measurement of cell size	Only indirectly from other methods	Simple and certain	Liable to errors in complex crystals	Fairly good, where indices can be found	Fairly good
Determination of symmetry class	Best X-ray method	Poor indirect method	Possible from intensity measurements	Impossible	Impossible
Determination of space-group	Only indirectly, liable to error	Best method	Certain only if enough planes be observed	Possible if enough indices can be determined	Possible
Accuracy of intensity measurements	Very poor, too many corrections necessary	Better than Laue method	Standard method, but extinction corrections necessary	Best photographic method	Most accurate, less corrections, but limited application

Figure 36: Bernal's (1927) comparative table of X-ray crystallography data production methods.

We see, for example, that powder methods were preferable for determining a property like the cell size and spacing of the smallest crystals, but presented major difficulties for deeper structural insights, like crystal symmetry class. The use of static Laue photographs was much better for this task, but only allowed for very rough qualitative estimations of intensity (usually based on the size of the spot). Ionization spectrometry was the standard method for this latter task, but was time consuming (only able to measure one angle at a time) and ran into difficulties with complex crystals.

A set of procedural norms, not unlike those implicit in Liebig's *Instructions*, provided a means of evaluation: range of application, expediency, and accuracy (understood in terms of the resolution or liability to error of data and in terms of the inferential distance between data and a desired result). Alongside the growth in techniques, there hence emerged increased reflection by practitioners on the in situ strategies a crystallographer might employ in obtaining reliable X-ray data. Coverage of such advantages and deficits became routine material for the textbooks that followed. A 1942 text by Martin Buerger covers all methods

reviewed by Bernal in greater detail, including their different practical aspects. Its chapter on the rotation method includes sections on collimator system design, camera design, choice of radiation, choice of crystal size, advice on selecting and mounting crystals, along with techniques for reducing sources of error, imposing lattice coordinates on photographs, and further interpreting and correcting data.

4.2.5. The search for a direct method

Shifts in the measurement of crystal properties demonstrate the interplay between modeling, target characterization, and the technical understanding that comprises an empirical program. Measurement procedures were carried out and altered on the basis of implicit models of crystal and X-ray interactions. Conversely, models of this process were altered in response to empirical findings. This reciprocal influence is clearly on display when we consider the measurement of a property like the structure factor (*F* in the final equation of section 2.4), a crucial step to structure determination for all but the simplest crystals. Starting with Laue, this term was thought to reflect the amplitude of a wave emitted by a collection of resonating atoms, the square of which would contribute to the intensity recorded at a point on a detector. After Bragg applied his reflection model to the Friedrich and Knipping experiment, it was instead taken to correspond to the collective scattering power of the atoms within a unit cell of the crystal. He thus inferred the size of the atom from the size of the spot on a Laue diagram. WHB reasoned from this model back to its consequences for a technique: if X-rays were reflected from crystal layers, then minor imperfections in layer alignment would spread out intensity readings over a small angle. This motivated his method of integrating ionization curves rather than recording their peak for the measurement of intensities. In each case, we see a double inferential movement from a model of the data production process to a feature of the recorded data, and back again.[107]

[107] See Tal (2017) for a discussion of this in the context of calibration and measurement.

As further complexities were recognized in the X-ray and crystal interaction, the inverse inference from recorded data to a target model became correspondingly complex. Crystal lattice points were three-dimensional, vibrated from thermal motion; the interaction between incident X-rays was affected by polarity; X-ray energy was absorbed by crystal atoms… all of these factors were expected to measurably affect intensity readings in ways that could not be corrected or improved through changes to the apparatus (grinding the crystals, increasing the voltage, choosing the right anticathode). They were added in piecemeal fashion to the intensity formula, burying the structure factor in a series of accompanying terms. This gave researchers a more accurate model of the crystal-ray interaction, but demanded more precision from the measurement process. The relative positions and sizes of atoms could not be simply read off the spots on a photograph, as Lawrence Bragg had done with the simple zincblende. Inferences from intensity recordings to structural details were prone to a range of uncertainties that gave the interpretation of diffraction data a qualitative character.[108]

Despite the foregoing developments, signs of frustration recur in the canon of X-ray crystallography: "After fifteen years of crystal analysis we are not in possession of a means of utilizing to the full the very accurate intensity measurements which can easily be made, and have to be content with qualitative tests. All who have worked in this field will appreciate the unsatisfactory nature of the arguments on which analyses are based" (Bragg, Darwin, & James, 1926, p. 307). Bragg was concerned with significant uncertainties corresponding to the double inferential movement described above: On the one hand, there was "the correct method of calculating the structure factor [*F*] for any given atomic

[108] Though things had improved somewhat by the end of the decade, complaints of this kind were routine in the literature of the 20s (Cf. Bernal 1927, p. 276 on the Laue method). Take Astbury: "it seemed very probable that not too much reliance could be placed on the evidence of intensity measurements. Even if these latter were taken with the greatest possible care, it is very doubtful if one is justified in comparing them rigidly with the theoretical values which might be calculated by using the simple Bragg theory. The complexity of the structure does not warrant this at the present stage of our knowledge concerning the scattering of X-rays. The labour of such calculations is very large, if precision is desired, and it is difficult to fix definitely the position of a small diffracting body in an elaborate structure. Barely more than two theoretical intensity ratios can be made to agree with experiment at the same time." (Astbury, The Crystalline Structure and Properties of Tartaric Acid, 1922, p. 510).

arrangement," and on the other, "the relation of the intensity to the structure factor [*F*]" (Bragg W. L., 1925b, p. 304).

In complex crystals it was recognized that one had to make assumptions about the arrangement of constituent atoms for any comparison between these inferences to be possible. Bragg was transparently vexed by this issue:

> Anyone who is familiar with X-ray analysis will realize the unsatisfactory nature of the assumptions usually made in calculating the intensity of reflexion to be expected from a given atomic arrangement, and the uncertainty of the results due to an ignorance of the laws governing the intensity. Experimental observations can be made with a high order of accuracy, but their interpretation is extremely difficult (1925a, p. 298).

Nevertheless, he expressed some optimism:

> Even though the dependence of intensity on structure factor is so very obscure, an accurate method of calculating the structure factor furnishes by itself a sufficient guide to the analysis of a structure. If the observed intensities show a parallelism to the calculated structure factors when both are arranged in order of increasing glancing angle, the assigned structure must be right (1925b, p. 305).

Bragg's confidence was vacillating. In the following years, as crystallographic ambitions grew and the targets of inquiry became larger and more complex, this "trial and error" method—assume a crystal structure, calculate the structure factor, then compared to empirical measurements—was increasingly viewed as unsatisfactory. This method had been employed since the beginning of X-ray crystallography.[109] The symmetry of simple crystals tightly constrained the number of parameters one had to consider ex hypothesi. But by the late 20s, this process became considerably more painstaking; where the 1912 zincblende photograph left room for one unknown parameter, Bragg's solution to the diopside crystal in 1928 required solving for fourteen (Perutz, How Lawrence Bragg Invented X-ray Analysis, 1990a). Bragg and others began to seriously ask whether some other means of inferring crystal structure were available. So began the search for a "direct method" for deriving the positions of atoms from experimental data.

[109] In a 1915 textbook co-written by father and son, the Braggs describe the same method: "Placing the sulphur atom at different points along the diagonal, it is possible to calculate theoretically the relative intensities of the spectra, in the way employed in the case of rock salt and zincblende. These theoretical values are then compared with the actual spectra obtained, and it appears that there is only one position for the sulphur atoms which explains the facts" (Bragg & Bragg, 1915, p. 132).

As early as 1915, WHB had entertained the application of Fourier synthesis to X-ray crystallography, but thought it "premature to expect too much until all the other causes of the variations of intensity have been allowed for" (1915, p. 111). The situation was more promising a decade later. In 1923 the American physicist William Duane showed Bragg's Law of diffraction followed from a simple quantum model involving a transfer of momentum from an incident X-ray photon to a crystal lattice. Compton (1923) pointed out that Duane's momentum transfer hypothesis followed from a general law in quantum theory expressing momentum as an integral over position space. The following year, Epstein and Ehrenfest (1924) reasoned that the electron density within an infinite grating could be expressed by a sinusoidal function:

$$\rho = \sum_{n=0}^{\infty} A_n \sin 2\pi m x/a$$

Duane (1925) and Havighurst (1925) applied these ideas more directly to the case of X-ray diffraction, deriving a related expression for the distribution of the "diffracting power" within a crystal:

$$\rho = \sum\nolimits_h \sum\nolimits_k \sum\nolimits_l A_{hkl} \cos\frac{2\pi h}{a} \cos\frac{2\pi k}{a} \cos\frac{2\pi l}{a}$$

a sum ranging over the different reflecting planes (*hkl*) of the crystal, where A_{hkl} was the square root of the intensity of reflection from that respective plane. Havighurst showed that *A* was proportional to the square root of F^2 in the intensity expression from Bragg et al. (1921; see section 3.4) and used their measurements to plot positions for the atoms of rock salt. This showed that values for the structure factor *F*, derived from intensity measurements, could be treated as the Fourier transform of the spatial distribution of diffracting matter within the crystal.

These theoretical contributions quickly caught the attention of Bragg who, with West (1928), elaborated on an "optical analogy" with Abbe's theory of the microscope to justify the application of Fourier methods to X-ray data. This, they noted, held the promise of being a direct method, but it required that the electron density be treated as the transform of a quantity given by complex numbers, i.e., one that includes phase information. Bragg wrote that,

> All data necessary to build up the image must be available if we know the amplitudes and phases of the waves coming from the spectra, for these are their only characteristics. We

> may therefore consider the complex system of spectra in the S space as a sort of code message, containing complete information about the form of the object (Bragg W. L., 1929, p. 388).

Because the intensities of scattered X-rays only correspond to their amplitudes (because they are given by the structure factor multiplied by its complex conjugate), information pertaining to their phases was lost in recorded data—an issue crystallographers came to call the "phase problem." Since the results of a Fourier transform depend on phase, this produced a significant form of underdetermination. Researchers had not escaped trial and error: "An infinite number of images correspond to any given series of experimental results. In order to determine the structure uniquely, we apply the criterion that the correct image out of all those possible must resolve into masses which correspond in form, extent and number to the atoms in the unit cell. This introduces the 'trial and error' principle into the Fourier analysis" (Bragg and West 1928, p. 385).

Though the Fourier series was now recognized as an additional tool for parameter determination, further progress stalled until an innovation due to Arthur Patterson, which served as a culminating moment in the early development of Fourier methods. In a paper titled "A Direct Method for the Determination of the Components of Interatomic Distances in Crystals" (1935), Patterson applied Fourier methods to observed intensities $|F_{hkl}|^2$ rather than the bare structure factors F. He showed that the Fourier series corresponding to the transform of observed intensities reflected from a given crystal plane (hkl)—the "F^2-series"—is equivalent to the weighted average distribution of electrons about a given reference point within a crystal. Moreover, this latter distribution can be constructed by taking an integral over a series of density functions, each standing for a single atom located at x_r with form factor f_r. The result can be expressed in one dimension as follows:

$$\sum_{h=1}^{\infty} F^2(h) e^{2\pi i h u/a} \;=\; \frac{1}{a}\int_0^a \rho(x)\rho(x+u)dx \;=\; \sum_{r,s=1}^{N}\sum \{ \sum_{-\infty}^{\infty} f_r f_s e^{2\pi i\left(\frac{h}{a}\right)[u-(x_s-x_r)]} \}$$

This provided a way to give "physical meaning," as Patterson put it, to crystallographic data by noting a correspondence between the height of peaks in the F^2-series—obtained directly from the transform of intensity data—and the distances between atoms. Using this result, one could work backward from intensity measurements, analyzing the F^2-series into the individual atomic positions and form factors that contribute to it. Further ways of manipulating these series were needed to remove the effects of temperature, uncover smaller

peaks within the hills of larger ones, and so on. Yet, the theoretical stage was set for analyzing the molecular structure of proteins. Of course, such a task assumed that proteins had the *kind* of structure that was susceptible to X-ray analysis. This was point in tension with the colloidal theo

ry of proteins, but by the 1930s the notion that proteins had no definite structure had lost influence among most biochemists.

4.3. Merging X-ray crystallography with the study of biological molecules

4.3.1. Protein crystallization revived

James Sumner's[110] success in crystallizing the enzyme urease was one contributor to the downfall of the colloid theory. The theoretical context motivating biochemists like Sumner was notably different from that cited by Reichert and Brown. Sumner and his contemporaries do not show interest in the geometrical structure of the crystals as guides to chemical knowledge. Rather, crystallization was primarily employed as a means for isolating and purifying substances for further chemical analyses akin to those being carried out on small organic compounds at the time. This program is seen in the work of American biochemist Thomas Osborne, who studied plant proteins at the turn of the century. Writing then, he said, "the fact that these proteid substances can be artificially crystallized […] is important as presumably furnishing a means of making preparations of undoubted purity which will offer a secure basis for further study of their properties" (quoted in Fruton, 1999, p. 142). Though they were put to different uses, Sumner's methods for deriving crystals would be familiar to earlier crystallographers. He ground jack bean meal into a fine powder, tried different solvents—water, salts, glycerine, alkalis—to extract urease, and experimented with filtration methods. In 1922 Sumner was inspired by the urease purification attempts of his mentor Folin, who used 30% alcohol solution and cooled the precipitate to −5°C during filtration (Sumner, 1937).

[110] Prior discussion of Sumner found at the end of section 2.6.

Such technical choices are strongly reminiscent of 19th century crystallization techniques: Funke suggested the use of alcohol solvent in one of the first published methods; Bojanowski obtained crystals from lark's blood using 50% alcohol; Hoppe-Seyler effected crystallization using 80% alcohol, filtering the results at 0°C, and washing them with 25% alcohol and ice water; Böttcher mixed 50% alcohol with fresh blood and cooled the mixture; and so on. Some technological improvements aside, Sumner's instruments were also standard issue for crystallization: a refrigerator for cooling, a high-capacity centrifuge, a microscope provided by Simon Gage, and an assortment of glass stills, pipettes, and grinding instruments (Dounce and Allen, 1988). On April 29, 1926 Sumner tried precipitating urease in 31.6% acetone instead. The reasons for this decision are unclear, though he may have gotten the idea from its prior use to extract urease from soy beans. Whatever his motives, it paid off. He later wrote, "In obtaining the crystals I felt much the same as a person does who is trying vainly to place in position a piece of a machine. Suddenly the piece slides in as if covered with butter. One knows that it is now where it belongs." (Sumner 1946, p. 119). The resulting crystals showed no contamination under the microscope, were strong in enzymic activity, and registered positive for proteins by the familiar Millon reaction.[111] This led Sumner to open his subsequent publication by claiming "a means of obtaining from the jack bean a new protein." (1926, p. 435).

Building on these results, John Northrop tested a number of proposed methods for purifying the enzyme pepsin at his Rockefeller Institute laboratory in Princeton. After attempts with fractionation, safranin precipitation, and dialysis, Northrop settled on the use of an alkali solvent and consecutive rounds of precipitation in magnesium sulfate and acid solutions—an improvement on a technique used by Pekelharing in 1896. Heating the product to 45°C, filtering, and slowly cooling the filtrate was found to induce crystallization. While a xanthoproteic test had signaled the presence of proteins and exhibited the proteolytic activity of pepsin, Northrop devised a battery of further tests to ensure that his crystals were

[111] The Millon test was one of three tests for proteins originating in the first half of the 19th century (1849 in its case). It signals the presence of proteins in an acid solution by forming a red color in response to mercurous nitrate. The earliest such test was the xanthoproteic test mentioned in footnote 21. In between was the Rose test or "biuret reaction" of 1833, in which a purple color formed in response to alkaline copper sulfate.

pure and not, say, a mixture of enzyme and separate protein. This was done by first showing that the crystals retained constant composition and properties after repeated crystallization, studying their solubility in a variety of solvents, comparing the diffusion coefficients of the protein and the material responsible for enzymic activity, and testing for production of antibodies upon injection into immunized rabbits. Having reviewed these results, Northrop wrote "It seems reasonable to conclude from these experiments that the possibility of a mixture must be limited to a mixture of proteins, so that the conclusion seems justified that pepsin itself is a protein" (1930, p. 765). With the publication of a similarly thorough test of crystalline trypsin in 1932, the perceived gulf between enzymes and proteins (reflected in Willstätter's view, Sec. 2.6) began to give way.

Northrop's remarks in his 1946 Nobel lecture reflect the acceptance of some novel methodological standards for crystallization work. "No one general method has been found which will lead to the isolation and crystallization of an enzyme, but certain general principles have been found to be of great assistance" (1964): The use of large quantities of material is advised, since dilute concentrations fail to yield crystals. Filtration by suction, previously employed by Sumner, is found superior to the centrifuge for separating precipitate from its mother liquor. Finally, fractionation using concentrated salts is advised over the use of dilute salt solutions, which destabilize proteins. These largely offer modest improvements and technological additions to 19th century approaches to crystallization, while drawing from a very similar stock of techniques and targets.

The rest of Northrop's comments focus on purification tests and further evidence, since 1932, that enzymes are proteins. This connection was a further blow against the colloid theory. By the later 1910s, even its most ardent proponents, such as Ostwald, had conceded that some substances within the living organism had definite molecular structures. Enzymes were widely regarded as substances of this type. The term 'enzyme' was due to Kühne, who distinguished them in 1876 in terms of their "fermenting" behavior. Importantly, this behavior could be exhibited outside of the living cell environment.[112] In 1894, Fischer

[112] After years of debate, experimental work by Buchner in 1897 settled the fact that fermentation could take place outside of cells.

hypothesized that this activity was due to the structural compatibility of an enzyme "lock" with its substrate "key."[113] If proteins were enzymes, and enzymes were capable of characteristic reaction behavior outside of the organism, then proteins must retain important structural features beyond the living environment. This did not sit well with the notion of colloidal proteins as a special living formation within the cellular protoplasm, nor the related distinction between "living" and "dead" proteins (Cf. footnote 34). Further, the colloidist's notion of proteins as mutable aggregates of protoplasmic micelles suggested that they would not possess a determinate molecular weight. This clashed with novel techniques for determining the mass of biological molecules. In the mid-1920s, Gilbert Adair's osmotic pressure methods and Theodor Svedberg's ultracentrifuge yielded more precise measurements that converted many of the remaining holdouts to the belief that proteins had a definite mass.[114] As evidence of determinate protein structure—and corresponding models—built up, the colloid theory continued to lose ground.

As for biochemists like Sumner and Northrop, work in their labs remained concentrated on the painstaking task of isolating and rigorously testing additional enzymes and viruses. For them, crystallization was a path to chemical inquiries and tests. But through their work, they developed a reliable means for producing stable objects that were soon put to markedly different uses.

4.3.2. Protein X-ray crystallography: first steps

In 1906, Fischer delivered a lecture to the German Chemical Society in Berlin on his biochemical studies of the preceding five years. He presented a thorough overview of the known amino acids, which he

[113] This was in contrast to theories like that of Carl Nägeli, who theorized enzymes were properties of substances that would act by means of molecular vibrations communicated by a substrate.

[114] Both measured the molecular weight of haemoglobin to be about 67,000 daltons. This was about four times the size of prior measurements by Hüfner and Gansser in 1907, Jacquet in 1889, and Zinoffsky in 1886, which hovered around 16,700 (prior measures were derived from Funke's analysis, placing the weight as low as 13,322). The discrepancy is most easily explained by the current view that haemoglobin is comprised of four heme groups, which were presumably separated by earlier methods, though there was also considerable bias against the view that organic molecules could be much larger than 10,000 daltons. For example, Fischer (1907) expressed skepticism about the recorded weights of 12-15,000, calling this "an alarmingly large complex." He wrote, "I believe, on the contrary, that they are mixtures of substances, the composition of which is in fact much simpler" (p. 8).

regarded as "the most important building blocks of the protein molecule." He then raised a basic question on which optical crystallography had stalled:

> Gratifying as this result may be, it solves only the smallest part of the problem that research into the chemical constitution of albumen puts before us; because the question is much more difficult: in which way and order are these pieces connected to each other in the molecule of the natural proteins? […] Research had come to a virtual standstill here, leaving some experts with doubts as to the solvability of the problem (1907, p. 9).

This question of structure was increasingly pertinent. In the mid-19th century, physiological chemists primarily classified substances in terms of their chemical constitution and the empirical properties they would exhibit in response to an array of reactive tests. By 1908, the biochemist Osborne would note in "Our Present Knowledge of Plant Proteins" that researchers were "well past the time when agreement in solubility, ultimate composition and color reactions, are to be accepted as evidence of the identity of two preparations of protein" (as in Mulder's study, Sec. 2.2). Deeper structural details were needed.

Fischer presented his synthetic method as a path forward. He had coined the term 'peptide' in 1902 from his effort to isolate homogeneous building blocks within the substances called peptones. Peptones, thought to originate from the action of gastric juices on proteins, attracted significant attention from physiological chemists in the late 1800s and until 1901 were hypothesized to provide the raw materials for the construction of proteins in the living cell.[115] Fischer developed a stepwise method to synthesize peptone-like molecules, consecutively adding amino acids into chains linked by "anhydride-like" bonds.[116] By 1907 he had synthesized over 100 different polypeptides, as he called them, using up to fourteen amino acids in a single compound. In 1924, one of Fischer's protegé's, Emil Abderhalden, built on this work to present an early molecular chain model of protein structure. In his model, the repeating unit was a class of cyclic compounds called diketopiperazines to which amino acids were attached. This assumption of recurring

[115] This hypothesis was due to Benjamin Moore and was disproved by Cohnheim and Loew, who gave clear evidence that enzymatic activity turned proteins into amino acids.

[116] These are bonds created by the formation of water between the carboxyl group (CO-OH) in one molecule and a hydrogen in another molecule. By 1907 it is clear Fischer had the CO-NH bond between a carboxyl and an amine group in mind. This grouping of elements had been postulated in 1892 as a recurring unit (CNOH) of proteins in Latham's (otherwise erroneous) model of protein structure as a compound of alcohols united to a benzene nucleus. In 1902 Hofmeister argued from chemical studies that the recurring unit was CO-NH-CH (Fruton, 1979). Still, there remained considerable uncertainty regarding the overall structural composition of these elements.

molecular structure was further buttressed by Svedberg's sedimentation studies with his ultracentrifuge, which he took to suggest that all proteins were made up of units of similar molecular weight (Kay, 1993; de Chadarevian, 2003).

Proteins had been theorized to play a central role in vital functions since the mid-1800s, a view that steadily gained adherents into the following century. According to Kay (1993), by the 1930s most life scientists were in some way committed to a "dominant explanatory framework that endowed proteins with determinative powers over heredity and related vital phenomena" (p. 104).[117] Protein chemistry was a central focus of funding bodies such as the Rockefeller Institute. Molecular models of the kind Abderhalden proposed for proteins were therefore of clear interest. Yet they encountered some basic limitations. For one, evidence for these models was produced biochemically through examination of the products of synthetic and cleavage-based approaches.[118] While these techniques allowed researchers to understand the basic atomic and molecular constituents of proteins, reason about the bonds holding them together, and identify correspondences between the chemical behavior of artificial and natural molecules, they offered limited insight into higher-scale geometrical structure.[119] Second, the synthetic approach favored by Fischer was time-consuming and much of its promise for determining protein structure relied on the assumption that proteins were much smaller than measurements suggested (Cf. footnote 59). Though limited in isolation, these models took on a more expansive use with the advent of X-ray analysis, informing the interpretations of early X-ray investigators such as William Astbury.

---

[117] Much of this was stimulated by perceived connections between enzymatic autocatalysis and replication of the molecular constituents of life. Growing evidence for the biochemical identity of enzymes and proteins (and, for a time, viruses) further encouraged this view. Cautious interest in this view is found, e.g., in (Svedberg & al., 1939, pp. 14-15).

[118] Fischer (1907) cites hydrolysis as the main method of protein cleavage. The action of proteolytic enzymes was beginning to be understood at this time, and would become another analytic resource.

[119] This is not to diminish the structural detail that was obtainable in this way. In fact, one key piece of evidence against Abderhalden's model was the fact that proteolytic enzymes were unable to cleave diketopiperazines. But the limitations are evident in biochemical studies of the internal structure of proteins carried out decades later, such as by Haurowitz (1949). Citing evidence that the chains of globular proteins undergo significant folding, he suggests various bonds that may form between amino acid residues to produce hypothetical folding structure. His only method for checking these claims, however, is by treating batches of polypeptides sequentially with reagents that cleave these bonds. Though this demonstrates the presence of the bonds and possibly their loci, reconstructing three-dimensional form from this process alone would be nearly impossible.

In 1934, Northrop shipped samples of crystalline pepsin to Astbury at the Textile Physics Laboratory in Leeds, UK. Since the early 1920s Astbury had been studying the crystalline structure of organic compounds using the new technology of X-ray diffraction. This was initially done under the advisement of WHB, whose equipment Astbury used to study simple acids at the Physics Research Laboratory at the University College of London. By the mid-to-late 1920s Astbury began applying these techniques to fibrous molecules, giving X-ray interpretations of the structure of animal hair keratin, wool, vegetable fiber, and muscular myosin. His initial studies were of low resolution, but clear enough for the discernment of two diffuse rings signaling repeating structures at approximately every 4.5 and 10 Angstroms (Fig. 37), which Astbury interpreted respectively as "backbone" and "side-chain" (or amino acid "residue") spacings within peptide chains (Bernal, Crowfoot, Evans, & Wells, 1935).

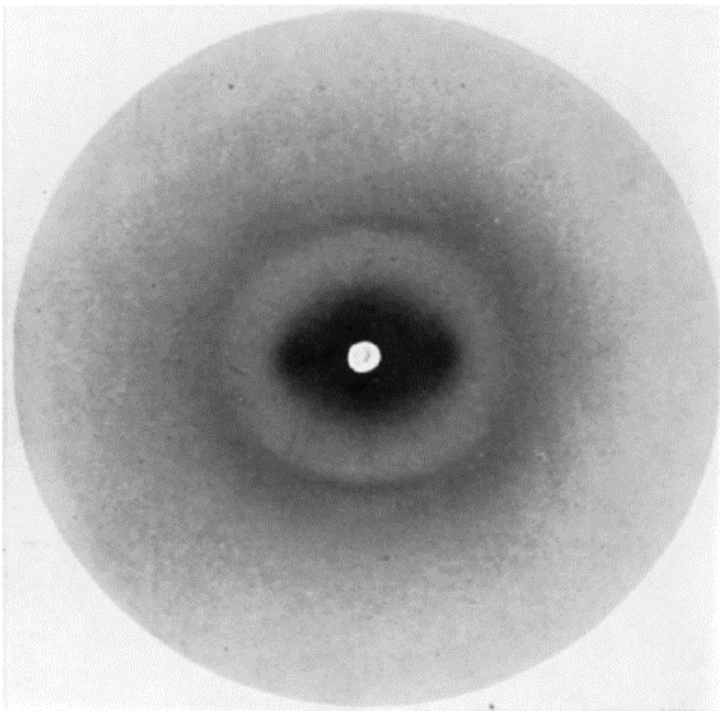


Figure 37: Astbury's diffraction pattern of stretched merino wool (Astbury & Street, 1932).

When confronted with crystalline pepsin, however, Astbury's fiber approach was of no avail. As he and a co-author noted,

> It is now some time since we first took X-ray powder photographs of crystalline pepsin kindly sent by Prof. J. H. Northrop, but no really satisfactory interpretation of these photographs presented itself because they show features which we have learnt recently

> to associate with the fibrous proteins: even single crystals, so far as we could judge with the minute crystals available, appeared to give results similar to those produced by many crystals in random orientation (Astbury & Lomax, 1934, p. 795).

The Debye-Scherrer powder method used by Astbury involves diffracting X-rays off a large batch of crystalline powder. The jumbled orientation of the individual crystals ensures that the rays reflect off every available plane of diffraction. But this method is best suited for simple structures and yielded little information when attempted with pepsin.

Proteins like pepsin required a specific kind of preparation for X-ray diffraction, as John Desmond Bernal and Dorothy Crowfoot (later Hodgkin) soon discovered. A decade earlier, Bernal had worked under WHB at the Davy Faraday Laboratory in London, where he toiled at determining the crystal structure of common metals. During this time, Bernal became skilled with the rotation method, in which a small crystal is mounted on a rotating arm in front of a narrow beam of rays. Bernal had attempted to diffract X-rays from dry specimens of non-fibrous proteins, including haemoglobin, using this technique, but "only obtained obscure bands." (Finch, 2007, p. 4). In 1927 he was appointed as the first lecturer in Structural Crystallography at Cambridge, a post created by the crystallographer Hutchinson who had previously advised Astbury and Bernal to work with WHB. Bernal was joined by Crowfoot five years later after a chance meeting on a train (Crowfoot, Dorothy Crowfoot Hodgkin - Biographical, 1972). Crowfoot was a skilled chemist with developing interests in X-ray crystallography. Together with Bernal, she published on compounds of biological interest such as liquid crystals and hydrocarbons, before turning to proteins. After receiving a sample of crystalline pepsin from John Philpot (working in Svedberg's lab in Uppsala), they soon discovered a basic effect that had impeded earlier research. Noting that "X-ray photographs taken of the crystals in the usual way showed nothing but a vague blackening," they set in search of a culprit (Bernal & Crowfoot, 1934, p. 794). Bernal and Crowfoot observed that the crystals rapidly lost their optical birefringence upon exposure to air yet produced a clear diffraction pattern when examined unfiltered in their mother liquor. They concluded that removing pepsin from a wet solvent environment distorted its structure, which explained the poor results up to then.

Once again, later uses of protein crystals were altering preparation techniques. This dynamic had always been at play: Reichert and Brown applied localized knowledge in altering their preparation techniques in order to obtain crystals that were satisfactory for viewing under the microscope. They avoided the use of "injurious" solvents or recrystallization procedures, and would add inert substances such as plasma or egg-white to retard substances whose rapid crystallization produced very small crystals. Likewise, it was noted earlier how spectrometric data was used to argue against repeated cycles of crystallization. In the same way, Crowfoot and Bernal had discovered that protein crystals that were satisfactory for microscopic or chemical analyses required further preparation to generate satisfactory diffraction data.[120] From this work, Bernal produced a high-quality diffraction pattern from pepsin in 1934. Like Astbury, he interpreted the regular spacings in this picture to indicate the regular molecular structure proposed by prior chemists like Fischer, Abderhalden, and Svedberg (de Chadarevian, 2003, Ch. 4).

Crowfoot continued her work on the crystallization and preparation of proteins for X-ray analysis, developing a sensitivity to background conditions. As a result, she was able to arrive at generalizations about the obstacles to protein X-ray crystallography. The most prominent in her work at this time concern the low reflecting power of crystals, warning against the use of powder methods, and the general requirement that crystals retain their water of crystallization (or mother liquor), which meant they must typically be examined "wet" in solution (Crowfoot, 1938; Svedberg & al., 1939). The rise of X-ray diffraction applications thus fed back on crystal preparation procedures, promoting renewed attention to factors affecting the sample states required for satisfactory data.[121] Early successes of Northrop and Crowfoot inspired a subsequent push for novel methods of protein crystallization, as seen in the vapor diffusion method used for lysozyme in 1937 (Abraham & Robinson, 1937).

Protein crystallography had taken on a new form by the late 1930s, compared to its nineteenth century practice. For one, it was no longer a given that practitioners were well-versed in crystallization

[120] Notably, their indicator that the crystals had undergone a structural change—loss of optical birefringence produced by double refraction—was an effect first associated with crystalline materials in Bartholin's 1669 essay on Iceland spar.

[121] Such considerations would later prove relevant to the determination of DNA's structure (Rich, 2003).

techniques. Instead, one could simply request samples from a group that specialized in the production of crystals, as Astbury did with Northrop's lab and Bernal and Crowfoot did with Svedberg's lab. It became possible for those working on the crystalline structure of biological specimens to know little about the details of their preparation—a state of affairs that resulted in a notorious mistake on the part of Watson (Cf. (Olby, 2012; Watson, 2001)). Along with protein crystallization procedures, attention to the microscopic details and relatively large-scale geometrical features of these crystals increasingly fell out of the picture. A look at the reviews of investigations produced by Crowfoot and others during the 1930s shows a much greater interest in the compositional details of the molecular constituents of crystals—determinations of their lattice parameters, bond orders, bond lengths, and bond angles as opposed to the angles, shapes, and colors of the crystals they form en masse.

The new crystallographers had found a way to largely bypass the observations of Reichert and Brown's crystals in pursuit of grasping their underlying constitution. As Bernal wrote, "What the X-rays have done is to move the structural formulae [of molecules] from the plane of hypothesis to that of measurement" (Bernal, 1930). Astbury claimed that "polypeptide chains in proteins first became 'real' in the light of the X-rays" (Astbury, 1939, p. 113). These researchers increasingly came to refer to their object of study as "molecular crystals" under the assumption that proteins were best viewed as molecular compounds, i.e., large structures held together by covalent bonds (Bernal, Crowfoot, Robinson, & Wooster, 1933). This view was reinforced by Bernal and Crowfoot's calculation of the molecular weight of pepsin based on their diffraction results, which was found to agree with Svedberg's measurements obtaining using his ultracentrifuge. The weight of nearly 40,000 daltons flew in the face of earlier assumptions that large, complex molecules could not remain stable in life processes. Such findings, taken with prior biochemical work, signaled the demise of the colloidal theory of proteins. Protein crystallographers were fully confident that they were studying macromolecules, and that their task was to seek a complete determination of their structure. This shift is captured in Astbury's observations that "Biology is fast becoming a molecular science, a desire to tread as far as possible the friendly ground of physics and chemistry and see where it leads" (1939, p. 113). For those engaged in this line of work, he reserved the term 'molecular biologist.'

4.3.3.Protein X-ray crystallography: core theory, motivating problems, and research community

In a 1930 article in Radiology titled, "The Place of X-Ray Crystallography in the Development of Modern Science," Bernal covered the technique's biological applications. Most of his discussion at this time revolved around the study of fibers and explanations of the contractile motion of muscles. By the end of the decade, on the other hand, the newly crystallizable proteins had become a subject of marked interest. A standard set of mechanical apparatuses were now available for their study (Fig. 38). In a series of reports on the state of the field from 1933-36, Crowfoot and co-authors comment on the increasing use of Weissenberg's goniometer for producing patterns, the integrating photometer for deriving intensity measurements,[122] and the recognition of Fourier analysis as the pathway to determining electron densities. X-ray crystallography encyclopedias, such as the International Tables for Crystal Structure Determination, were now available to researchers. The first volume of this work presented diagrams for the equivalent general positions of the 230 possible lattice space groups, together with structure factors and a list of characteristic reflections. The second volume provided tables for many of the trigonometric functions used in Fourier analysis and for many of the physical quantities encountered in structural studies. In the fifteen years following Bernal's 1927 review article, leading up to Buerger's 1942 textbook, an increasingly stable and standardized repertoire of techniques and instrumentation had become available to X-ray analysts. This facilitated the wider adoption of X-ray crystallography as a method, and so its emergence as a specialized field. In the mid-1940s Ewald proposed the formation of an international society of affiliated researchers. In 1948 the International Union of Crystallographers was officially founded, along with the journal *Acta Crystallographica*, as a central hub for crystallographic research. As this specialization took place in the 1930s and 40s, a small contingent of researchers became deeply devoted to the investigation of proteins.

[122] Astbury (1929) describes one such device and its underlying theory.

Shortly before the war, Crowfoot was told by Bragg that he looked forward to "the structures of proteins which we shall all live to see" (Hodgkin, Taking Penicillin to the R.I., 1990). Bragg had evidently caught her and Bernal's enthusiasm for the protein problem. Bernal and Crowfoot agreed that the basic units of proteins were amino acids linked by peptide bonds and that fibrous proteins were likely arranged in chains thereof. Things were less clear with proteins such as pepsin, insulin, and haemoglobin, which came to be called globular: "Peptide chains in the ordinary sense may exist only in the more highly condensed or fibrous proteins, while the molecules of the primary soluble proteins may have their constituent parts grouped more symmetrically about a prosthetic nucleus" (Bernal & Crowfoot, 1934, p. 794).

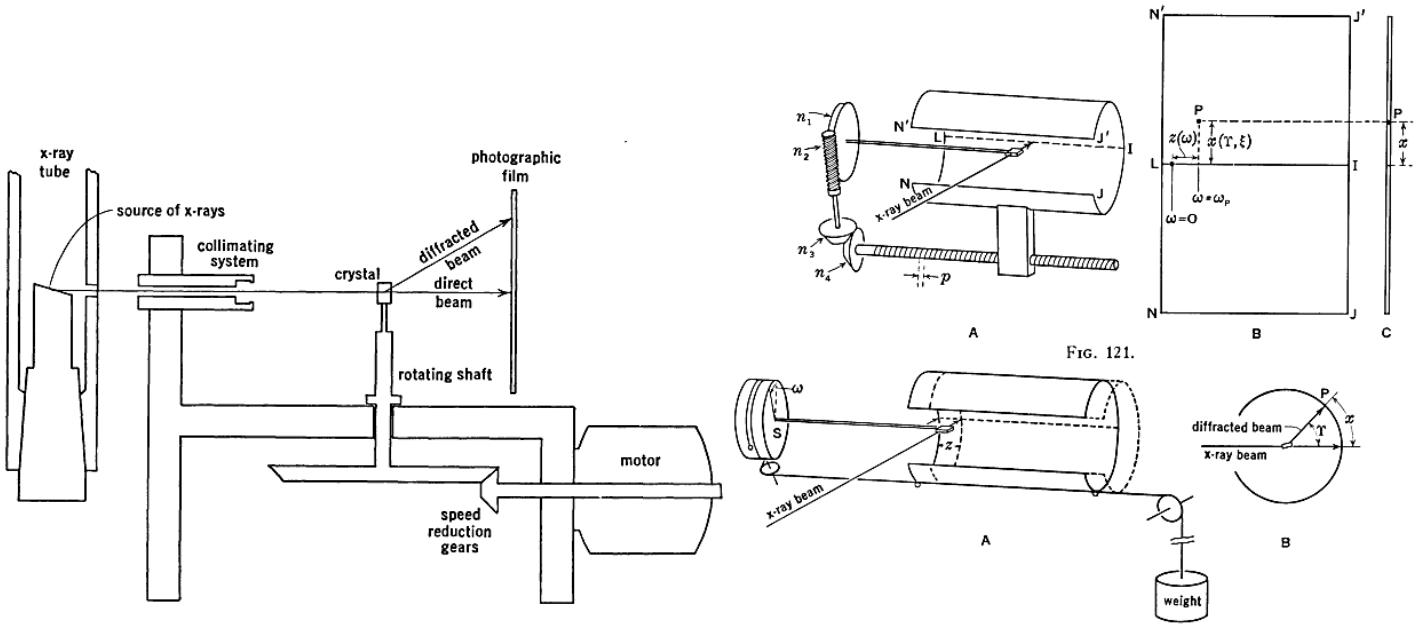


Figure 38: Schematics for common X-ray diffraction apparatuses (Buerger, 1942). Left: A device for rotating and imaging a mounted crystal. Right: the central element of the Weissenberg method, which coupled crystal rotation (e.g., by means of the gear labeled n_1 in the upper diagram) with translation of a cylindrical film (by means of the gear n_4).

A noteworthy achievement is mentioned in Crowfoot's report of 1936: "J. M. Robertson has now achieved in his work on the pthalocyanines the first absolutely direct analysis of an organic molecule" (Cox & Crowfoot, 1936, p. 214). Robertson, a former student of Bragg's, had achieved the first fully direct method for structure determination by a method called isomorphous replacement. This involved soaking a crystal in a solution of heavy atoms, some of which would bind to the crystal without altering its structure.

Using data from heavy atom derivatives, and subtracting the native crystal data, researchers were able to calculate their independent contribution to intensity readings. Using the Patterson function, they could locate where heavy atoms were positioned in the unit cell. By triangulating the intensity perturbations due to different heavy atoms, their contributions to the phase of the native crystal's structure factor—and so the original phase information—could be deduced. A potential solution to the phase problem was on hand and known to Bernal and Crowfoot,[123] but would wait over a decade for its application to proteins.

In 1938, the first high quality diffraction photographs of crystalline haemoglobin were published in Nature by Bernal, Isidor Fankuchen, and the Austrian researcher Max Perutz.[124] The latter's cousin Gina was married to the haemoglobin biochemist Felix Haurowitz. During a 1936 visit from Perutz, then en route to study under Bernal, Haurowitz shared his latest findings. He demonstrated to Perutz how the crystalline form of reduced (or deoxy-) haemoglobin underwent major changes upon exposure to oxygen (Putnam, 1994) and suggested he obtain haemoglobin crystals from Adair to study the molecule's structure at Cambridge (Finch, 2007). Once in England, Perutz was mentored in crystallography by Fankuchen and Bernal. Following Haurowitz's advice, he obtained haemoglobin crystals and learned to generate diffraction patterns from them. Perutz and his mentors contradict colloidal assumptions in their commentary on this data, noting that its clear spacings "proves the complete internal regularity of the protein molecules down to atomic dimensions." (Bernal, Fankuchen, & Perutz, 1938, p. 524). By the date of their joint publication, however, Fankuchen and Bernal had both left Cambridge. Seeking to remain, Perutz pursued a position in the Cavendish Laboratory, where Bragg had transferred all crystallographic research in 1931. After some hesitation Perutz presented his diffraction photographs to Bragg, who was immediately excited by the application of X-ray analysis to biological molecules.[125] Bragg had recently succeeded Rutherford as chair

[123] Bernal mentioned "the introduction of a heavy atom" as a method for solving protein structure in 1939 (Bernal, 1939, p. 665).
[124] The paper cites and confirms Reichert and Brown's description of the higher-scale structure of the crystals. This suggests that more extensive studies of haemoglobin crystals were not available nearly thirty years later.
[125] This was in contrast to the deep skepticism most other crystallographers had toward the project. As Kendrew later wrote, "As far as I could determine, Bragg was the *only* crystallographer in Cambridge—apart from Max Perutz—who did not believe we were wasting our time on a project much more complicated than had previously been attempted" (Kendrew, 1990, p. 89).

of the lab and used his influence to secure funding for Perutz from the Rockefeller Foundation (at that time a major investor in biophysics under the initiative of Warren Weaver).

Perutz and Bragg were not alone in their enthusiasm. Three cross-disciplinary meetings on the topic of proteins took place in 1938 (Law, 1973). This included a Royal Society meeting in November on "the protein molecule," published in its Proceedings the following year. In his opening address, Svedberg claimed this meeting "would have looked preposterous" earlier that decade (Svedberg & al., 1939, p. 1). His talk conveys a sense of the triumph of the molecular view of proteins. Above all else, he credited his sedimentation studies and X-ray analysis with cementing the view that each type of protein has an individuality corresponding to a definite atomic structure, that there is a general distinction between the fibrous and more heavily folded and symmetrical globular proteins,[126] and that there is likely to be "a similarity in the architecture of all proteins" (p. 6). He expounded on this last point in terms of "the finding of a rule of simple multiples for the molecular weights of the proteins. If we choose 17,600 as the unit the majority of the proteins may be divided into eleven classes with molecular weights which are multiples of this unit by factors containing powers of 2 and 3" (p. 6). Several theories had been advanced to account for this apparent regularity, including a law postulated by Bergmann and Niemann in 1937 according to which each amino acid residue occurs at a regular interval in a polypeptide chain. Svedberg notes with interest that these authors had been able to deduce his 'rule of simple multiples' from their theory. Their results were also compatible with the cyclol model of protein structure due to mathematician philosopher-cum-theoretical biologist Dorothy Wrinch. She proposed that proteins were comprised of ring-like complexes of peptides, linked together in "fabrics" that could form closed structures of a fixed number of amino acids: "The special geometrical nature of the fabric shows that closed, cage-like 'globular' proteins may be anticipated, containing not random numbers of residues, but specific numbers which include multiples of

[126] By this time researchers such as Bernal and Astbury were increasingly convinced that these were fundamentally the same. Astbury writes that "there is apparently a common factor in the synthesis of all proteins, whether fibrous or globular, and their shapes and sizes, as we know them now, seem to be somehow of a secondary consideration. The original distinction between fibrous and globular has almost disappeared […] The combined impression we receive from the different lines of approach is that all proteins are fundamentally of a fibrous structure, built up by the successive aggregation of relatively small units" (1939, p. 122).

72" (Svedberg & al., 1939, p. 24). Bergmann and Niemann's calculation that haemoglobin consisted of 576 (or 8 x 72) amino acids appeared to some, including Wrinch, as a striking confirmation of her theory. But many others regarded such hypothesized regularities with suspicion or as no more than "an inspired guess" (Bernal 1939, p. 663). Cavalier attempts to provide a mathematical theory of protein structure clashed with the conservative evidence-driven culture of many biochemists. They soon found ways to critique the cyclol model and related mathematical laws of regularity on their chemical merits.[127]

In the same year of this published symposium, both Bernal and Linus Pauling (with Niemann) published articles with the title "Structure of Proteins." They agreed that most chemical and X-ray evidence pointed to the amino acid buildings blocks of proteins being linked in peptide chains, that globular proteins had a spheroid form overall, and that they were likely built in the same manner as the fibers, since X-rays of denatured globules resembled those of the fibrous β-keratin. Both adopted Astbury's interpretation of his diffraction data, noting that the 10-angstrom amino acid unit formed a reflecting plane and so its atoms were co-planar. In both articles, the deepest puzzle concerned the relation between a protein's outer form and its internal structure; that is, in the three-dimensional folding of polypeptides and the bonds formed between their segments.[128] Bernal raised the question whether the constituent polypeptides were composed of a single unit or multiple sub-units and, if the latter, how these sub-units were linked to one another. Pauling and Niemann stressed the connection between protein folding and function:

> A native protein molecule with specific properties must possess a definite configuration, involving the coiling of the polypeptide chain or chains in a rather well-defined way […] We feel that the biological significance of proteins is the result in large part of their versatility, of the ability of the polypeptide chain to accept and retain that configuration which is suited to a special purpose from among the very great number of possible configurations accessible (Pauling & Niemann, 1939, p. 1867).

[127] Albert Neuberger, for instance, noted several structural assumptions in cyclols that violated known chemical behavior and observed that error rates in amino acid determination for proteins gave an 80% chance that a random distribution of residues might give values that accord with Bergmann and Niemann's frequency law. Pauling and Niemann (1939) rejected the cyclol model forcefully, arguing "there exists no evidence whatever in support of this hypothesis […] instead strong evidence can be advanced in support of the contention that bonds of the cyclol type do not occur at all in any protein" (p. 1860). Bernal and Fankuchen were similarly harsh: "no case whatever has been made out for the acceptance of the cyclol model on X-ray grounds; indeed so far as it goes, the X-ray evidence is definitely against the model" (Bernal, Fankuchen, & Perutz, 1939, p. 897).

[128] Both papers propose the bonds are due to hydrogen, but admit to no decisive evidence.

To Bernal this versatility was simultaneously a barrier to insight, permitting "a very large number of such possible models, and as yet nothing to choose between them" (Bernal 1939, p. 666). He routinely ended his papers of this period with a call for greater coordination among protein researchers.

Such efforts were stalled by the advent of World War II. In its aftermath there emerged a small but clearly defined community of researchers working on protein structure. At its core were two groups with common aims, though tending toward opposite strategies: Pauling and collaborators at Caltech, who focused on "bottom-up" model-building rooted in the geometry of chemical bonds; and the "top-down" X-ray crystallographers in Britain (Kay 1993, Ch. 4; de Chadarevian 2003, Ch. 4). Practically all prominent members of the second group were intellectual descendants of the Braggs (Fig. 39). This ancestry created a shared background and technical solidarity among practitioners, rooted in a common set of technical problems and skills, all of which formed the social basis for a distinct research front in crystallography (Law, 1973). Those working within this social sphere, or in areas that drew on parallel techniques, were taken more seriously in their lab results and communications, and so played a larger role in determining scientific consensus.[129]

[129] Some evidence for this solidarity is found in the policing of boundary lines of expertise, as in pronounced criticisms directed toward the models of an "interloper" like Wrinch. For example, authors often cited the chemical implausibility of Wrinch's model as a chief fault. Likewise, Pauling's criticisms of Wrinch and his correspondences with the crystallographers exhibits relatively greater solidarity between the Caltech and British groups, despite their scientific rivalry.

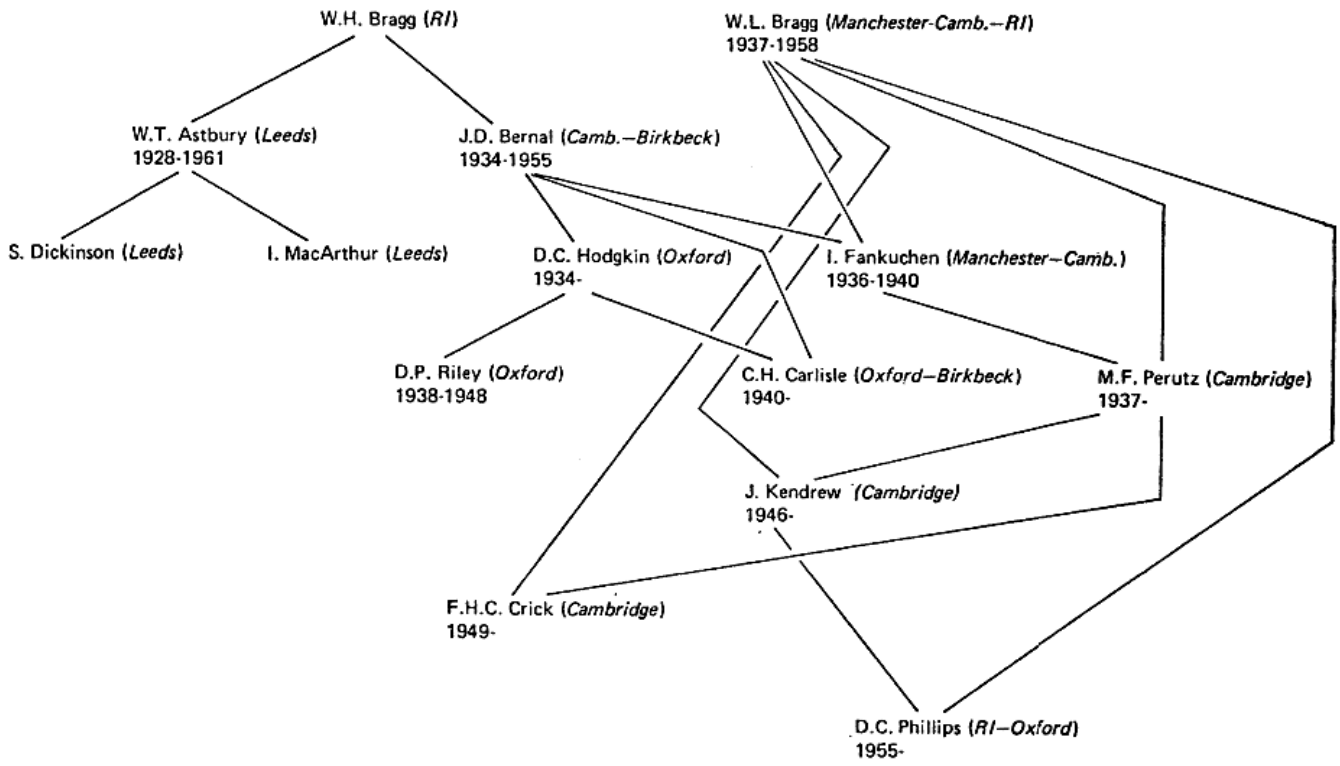


Figure 39: Major protein crystallographers in Britain prior to 1950 (from Law, 1973). The dates show approximate time of interest in proteins and lines indicate master-student relationships.

4.3.4. X-ray crystallography of haemoglobin

Perutz was granted refugee status upon the outbreak of war in Europe. After being detained as a possible traitor and subjected to miserable conditions in Canada, he was allowed to return to research at the Cavendish. He finished his PhD in 1940. In this work he determined the dimensions of the unit cell for haemoglobin crystals and, showing his advisor Bernal's influence, proposed the molecule was composed of regularly arranged sub-units. In 1942-43 he continued to diligently photograph haemoglobin crystals, collecting data during overnight shifts in a civil defense post. These totaled 7,000 diffraction spots (or "reflections"), with each photograph recorded on film for one to two hours. Perutz likely developed his renowned skill at naked eye readings of intensity while studying these. This involved comparing haemoglobin diffraction dots with a reference scale: a record on film of reflections from anthracene taken at different exposure lengths (Finch, 2007). The interpretation of so much data was exceedingly difficult.

To aid him, Perutz brought on John Kendrew at the Cavendish in 1945. Like Perutz, Kendrew had trained in chemistry. His interest in "the protein problem" was inspired by Bernal, with whom he served during the war. Further like Perutz, he only received sufficient training in crystallography once he had arrived at Cambridge. Two years later, both received funding to lead the Medical Research Council's Unit for the Study of the Molecular Structure of Biological Systems, housed at the Cavendish.[130]

Publications from the Unit began to arise during the latter half of the decade. Two clear influences on Perutz's work of this period can be tied to his 1938 Nature commentary on diffraction patterns with Bernal and Fankuchen. The first was theoretical: an assumption that there was a significant degree of internal regularity in the structure of haemoglobin, which had been a point of emphasis in Bernal's writing from the late 30s. The second influence was methodological: the 1938 paper had observed a directional bias in the shrinking and swelling of haemoglobin crystals upon de- and re-hydration. At a loss for solving the phase problem, Perutz and his co-authors in the 1940s attempted to control the rate of the shrinking process and take images of the crystals at different stages, inferring structural properties from there. In 1942, for example, Perutz argued that the length of molecule along one dimension must be equal to the length of the crystal's unit cell, since the crystals did not shrink along their b axis.

This work led to a series of co-authored publications in the later 1940s. Perutz's papers of this period have a common structure: their main object of study was haemoglobin from horse blood.[131] They first describe the method of crystallization,[132] the process of mounting the crystal, and of taking the X-ray

[130] On the details of the unit's establishment, see de Chadarevian 2003, Ch. 3.

[131] On Perutz's suggestion, Kendrew's PhD work focused on fetal and adult sheep haemoglobin. They coauthored a paper on this in 1948, in which the influence of Perutz's prior papers is evident. One noteworthy feature of these papers is that they routinely cite Reichert and Brown (1909) in characterizing the crystals but cast doubt on a core guiding assumption of *Crystallography of Hemoglobins*: that differences in crystal structure correspond to chemical differences. Though sympathetic to the idea, Kendrew and Perutz admit it is hard to square with the significant morphological variation in fetal haemoglobin crystals from the same source. By 1952, this assumption had been abandoned by the haemoglobin team, who framed the notion as counterintuitive: "[diffraction pattern] similarities support the conclusion that haemoglobins from different species of mammal are very similar in general structure, as might have been expected" (Bragg & Perutz, 1952, p. 426).

[132] Again, the process harkens back to 19th century techniques, with minor modification. Blood was washed in a saline solution to remove serum, frozen to break the cell membranes, filtered to separate out the stroma, oxidized with potassium ferricyanide, filtered again, and finally salted out with ammonium sulphate. The first three steps are practically the same as Hoppe's 1864 method. The salting out step was used by Funke by 1876. Only the oxidation step differs.

photographs. A considerable portion of the papers were spent characterizing the morphological features of the crystals, first their optical properties, then moving to unit cell dimensions and likely space groups (Fig. 40). Final remarks concerned protein structure.

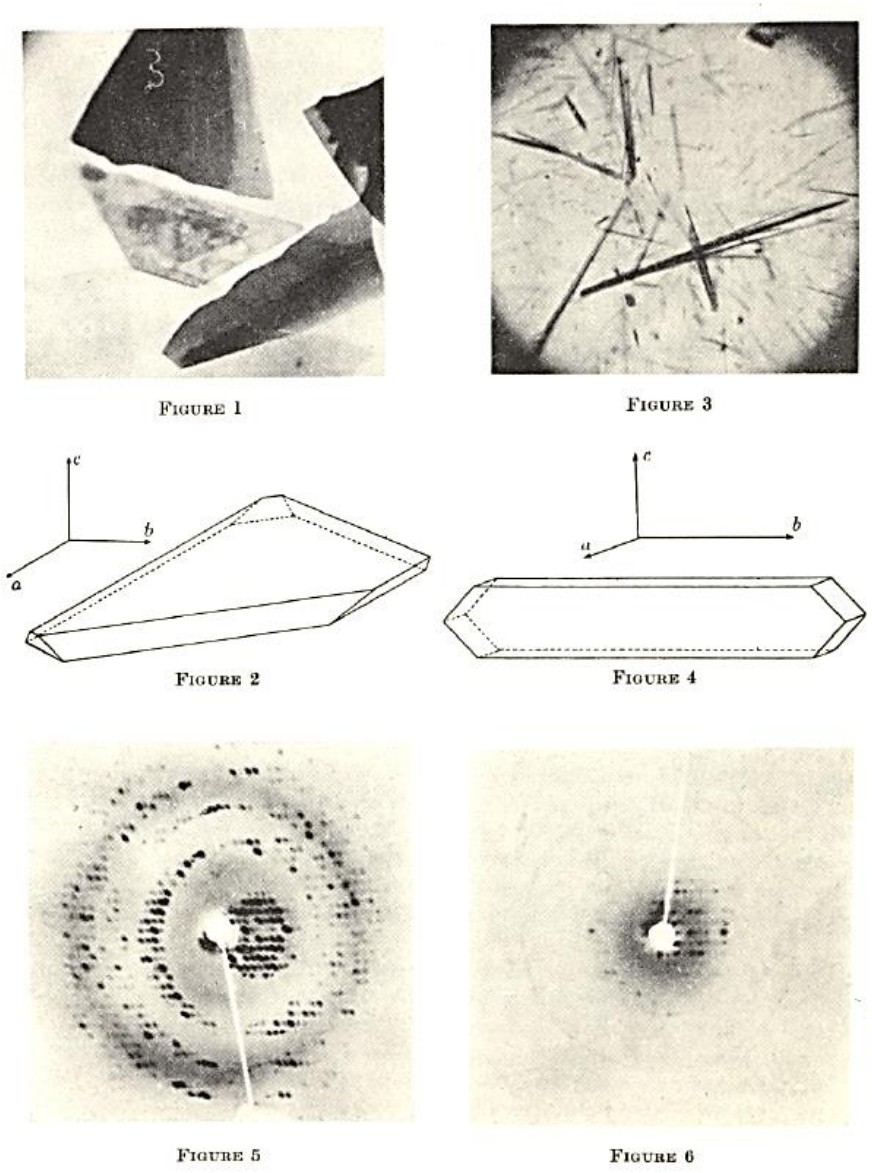


Figure 40: Images from Kendrew and Perutz (1948), showing crystal microphotographs, morphological features, and diffraction patterns for adult and fetal haemoglobin (left and right, respectively).

The reflection data offered limited insight into the internal structure of haemoglobin, as the authors acknowledged: "the present difficulty is that almost any model based on a folded polypeptide chain structure agrees with the main features of the X-ray data" (Boyes-Watson, Davidson, & Perutz, 1947, p.

125). They instead set a more modest goal: describe the relation of the molecule to its surrounding liquid of crystallization and build up a picture of its overall form from there.[133] Along with shrinking studies, they added heavy atoms to the liquor of crystallization and imaged the shift in intensities to study whether the proteins were tightly packed with 'bound water' within them, or whether the liquid formed sheets of 'free water' between the crystalline molecules.[134] By determining the boundaries of the liquid, they could also get a rough sense of the overall shape of the protein molecule and its packing within the crystal. The picture that emerged was of a cylindrical disk, wider than it was tall, packed in a hexagonal array of columns with sheets of liquid between each column and between successive column layers (Fig. 41). Bragg (1965) referred to this in retrospect as the "pill box" model.

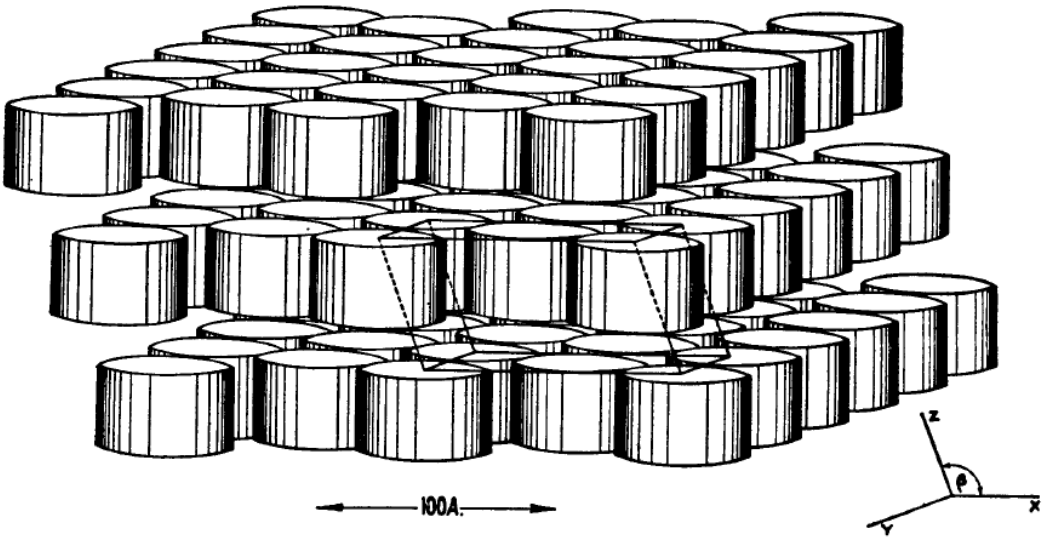


Figure 41: The "pill box" model of the haemoglobin molecule and its crystalline packing (from Boyes-Watson, Davidson, & Perutz, 1947). Each cylinder is an individual molecule separated by a layer of liquid.

The quantity of reflections from Perutz's wartime imaging speaks to the surge of data that began to confront protein crystallographers. This spurred two basic developments in method: the adoption of novel forms of technology and an increased division of labor. Already in the 1940s, Perutz had brought on two researchers, Joy Boyes-Watson and Edna Davidson, to aid with intensity readings and analysis. Frustrated

[133] It was accepted fact by this time that protein crystals were separated by large gaps of solvent molecules (Cf. (Granick, 1942)).
[134] These two possibilities had been raised by Bernal in his 1939 essay on protein structure.

with error-prone naked eye judgments, Kendrew came across a densitometer built by the cell biologist Peter Walker that could be easily repurposed for diffraction data. The process of analysis was increasingly refined as the mass of data increased, and by the 1950s teams of assistants (almost all women) were recruited to feed recorded film into a hand-operated Joyce-Loebl double-beam microdensitometer and measure the peaks it would output on a millimeter scale. Kendrew would review the work of each human computer and fire those with error rates higher than 5% (de Chadarevian, 2018). During the final steps toward the first haemoglobin model, this intensity measuring work had been further sub-divided, with densitometer outputs given to Tony North for processing with an Automatic Linear Diffractometer (Finch, 2007).

Intensity measurements served as the raw material for Fourier analysis: first, the calculation of a Patterson series (the Fourier series of the intensity measurements F^2) followed by inference of electron densities from these, given some assumptions about phase. In 1935, Crowfoot visited Manchester after reading a publication on Fourier analysis by the crystallographer C. Arnold Beevers and physicist Henry Lipson. She left with two £5 boxes of the so-called Beevers-Lipson strips (Hodgkin, 1990). These served crystallographers as the principal means of calculating Fourier series into the 1950s. Kendrew recalled the "diabolical invention" in 1987:

> It's a box containing a whole lot of strips of cardboard and on each strip is printed a sine wave, different strips have different amplitudes and different frequencies and different phases, so by selecting the strips and laying them out below one another, and then adding down, you are doing a Fourier series (Kendrew, 1987, p. 1).

This tedious process—requiring a new sum for every diffraction spot—was grudgingly accepted as the only way to process intensity data until the late 1940s. By that point, Bragg and Perutz were seeking to move from two- to three-dimensional Patterson analyses of haemoglobin. To do so, Perutz, Boyes-Watson, and Davidson produced over 58,000 reflections, "a task whose length and tediousness it will be better not to describe" (Perutz, 1949, p. 475). The 3D Patterson required summing around 7,000 Fourier terms for each reflection point. For reference, a calculation for several hundred reflections in only two dimensions could take weeks; at such rates a three-dimensional analysis was completely impractical. Instead, they turned to one "Miss E. Gittus" of the Scientific Computing Service in Britain, which provided calculation services

via Hollerith punched card machines.[135] In six months they had their results, published in 1949. Still committed to the pill box model, Perutz interpreted the electron density contour maps to show further details within the haemoglobin columns: "there is a structure, rod-like though rather tortuous, winding its way along the X direction […] All these rod-like vector peaks are roughly parallel to the X-axis," which he wrote "suggests that the haemoglobin molecule contains chains parallel to X" (Perutz, 1949, p. 488). In this article and a co-authored second piece titled "Arrangement of Polypeptide Chains in Horse Methaemoglobin," (Bragg, Howells, & Perutz, 1952), Perutz interpreted the density maps to suggest parallel chains that wound throughout the cylindrical envelope of the molecule, stacked in vertical and horizontal layers (Fig. 42). Yet his writing is replete with caveats. After stating that this type of chain configuration was compatible with the calculated density maps, he warned that "details in the two pictures [Fig. 42, right] are, of course, purely imaginative" (1949, p. 489).

[135] By some counts, the Cavendish was quite late to adopt the use of tabulating machines. Pauling had arranged for the use of IBM Hollerith machines at Caltech as early as January 1940, and the 1946 Conference on X-ray Analysis in London lists a number of computational techniques available to crystallographers (Cranswick, 2008, p. 72). On the other hand, publications outlining these techniques were uncommon (Pauling's group only described their methods in 1946) and the actual devices were extremely rare at the time.

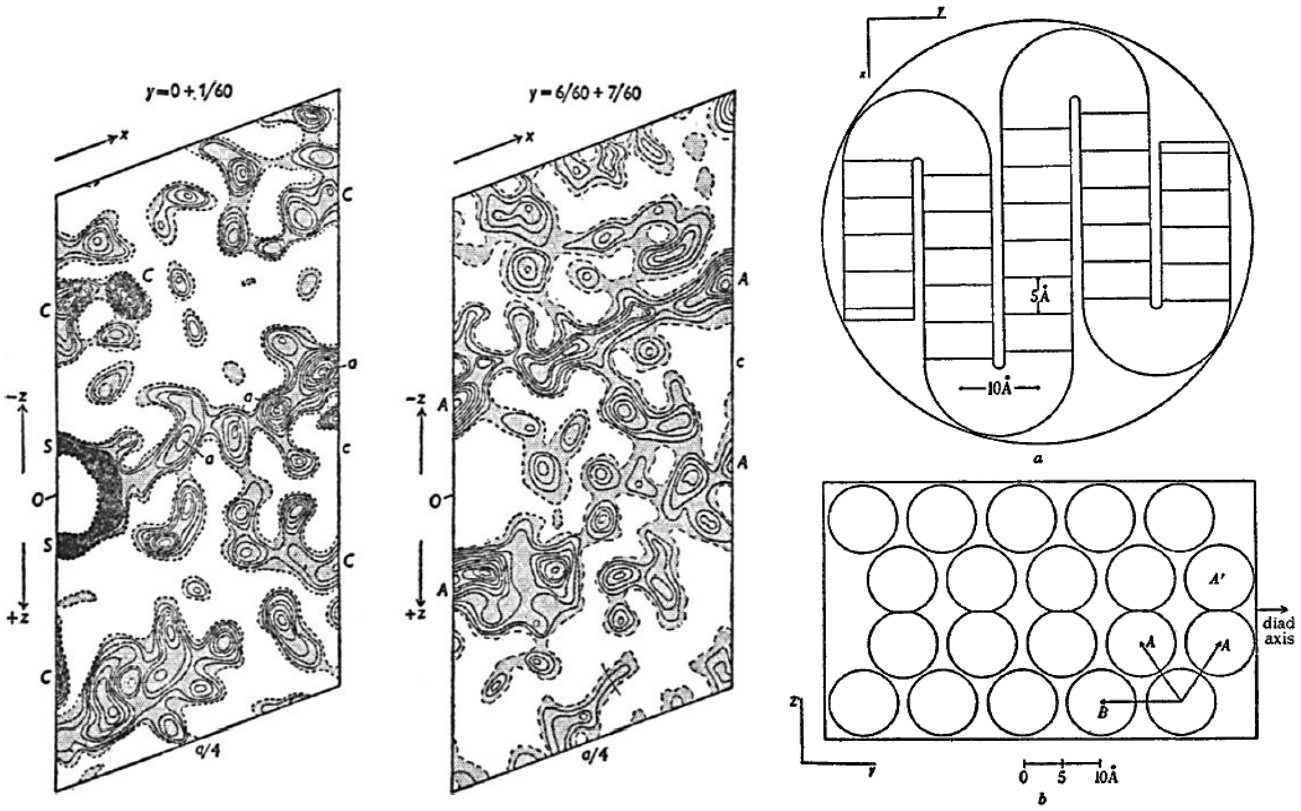


Figure 42: Images from Perutz's study of haemoglobin (1949). Left:a contour map depicting continuous ridges parallel to the x axis, which Perutz interpreted as polypeptide chains. Right: an idealized model showing the chains winding within the diameter of the cylindrical molecule (top) and the spacing between each chain, viewed down their length (bottom).

The contrast between the "top-down" and "bottom-up" groups attacking protein structure is clearly displayed in the differences between Perutz's uneasy interpretations of density contours and a publication from the same year by Pauling and Corey, also on configurations of polypeptide chains. Starting from an analysis of the geometry of the single bonds forming polypeptides, the authors built up an account of their stable configurations, the orientations of bonds in the sub-groups of amino acids, and from there a model of the larger-scale molecular structures that could be formed from them. Two basic types were described: a winding helix and a pleated sheet formed from bonds between helices (Fig. 43).

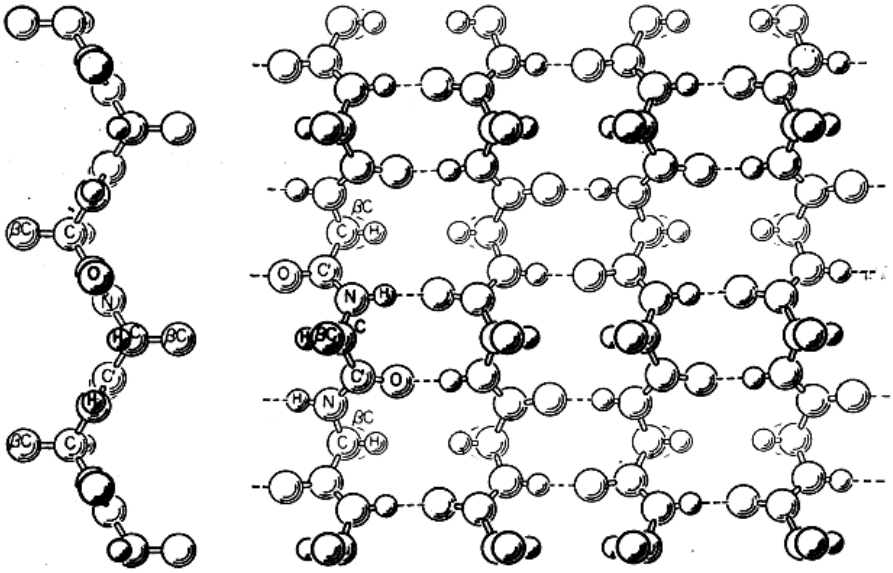


Figure 43: Structural elements of proteins from Pauling and Corey (1951). Left: a polypeptide helix. Right: a bound series of parallel helices forming a sheet.

Francis Crick, a recent arrival at the Cavendish, showed that the density in the Patterson map region containing Perutz's "rod-like peaks" was lower than required for parallel chains (Finch, 2007). Bragg, Howells, and Perutz (1952, p. 141) acknowledged that their pill box model predicted much higher intensities than were found in data, but struggled to attribute this to a definite structural discrepancy. Dorothy Hodgkin (Crowfoot) expressed uncertainty over the relation between Perutz's contour maps and the layered chain model (Hodgkin, 1950, p. 71), but still discussed it as the best going model. Pauling and Corey's paper finally drove these worries home, demonstrating structural oversights on the part of the Cavendish group. Their model showed the most stable configuration of polypeptide helices had 3.6 amino acids per turn, where the Cavendish group had only considered models with integral numbers. Bragg et al. had also assumed that each amino acid sub-unit could rotate freely around the peptide bond, where the Caltech group showed this rotation to be constrained. Combined with Crick's criticisms, Perutz's interpretations of the density maps looked hasty—too beholden to Bernal's vision of deep regularity within the protein molecule and too eager to confirm the parallel chain model. It was subsequently abandoned. As Bragg remembered the era leading to this moment in the early 50s, "the idea that the molecule contained some kind of regular structure of protein chains, which would give a strongly defined character to a

Patterson synthesis was a guiding star which encouraged the investigations. As events turned out, it was a false star" (1965, p. 3). While Bragg and Perutz returned to the drawing board, seeking to improve their methods of producing and interpreting Fourier transforms, a new development in data processing took place.

In 1949, Hugh Huxley, then a graduate student under Kendrew, enlisted his friend John Bennett in Cambridge's Mathematical Laboratory for help with a Patterson analysis. Bennett recognized the sum could be programmed on the Lab's newly operational Electronic Delay Storage Automatic Calculator, or EDSAC, one of the first general purpose electronic computers. The result was a program that took 30 minutes to calculate Huxley's two-week homework assignment. Though Huxley left to work on muscles, Kendrew continued the project with Bennett. EDSAC was a thermodynamic monster, its paper inputs were prone to jams, and machine time was fought over by different departments,[136] but when it worked, previously unfeasible calculations were obtained within hours. Working for multiple days at a time, Kendrew went through rounds of troubleshooting and evaluations of the errors and efficiency of the program until he arrived at a reliable set of sub-routines for calculating Fourier transforms from intensity data (de Chadarevian, 2018). In 1952 Bennett and Kendrew published a paper describing how to compute Fourier transforms and generate density maps electronically from these data (Fig. 44). The program could store up to 256 "words" storing pairs of amplitude entries and carried out 10 to 15 cosines calculations per second. The time to carry out a basic Patterson analysis could be reduced from two weeks to 30 minutes, with another 30 minutes for its considerable teleprinter output.

[136] See Kendrew 1987 and de Chadarevian 2003, Ch. 4.

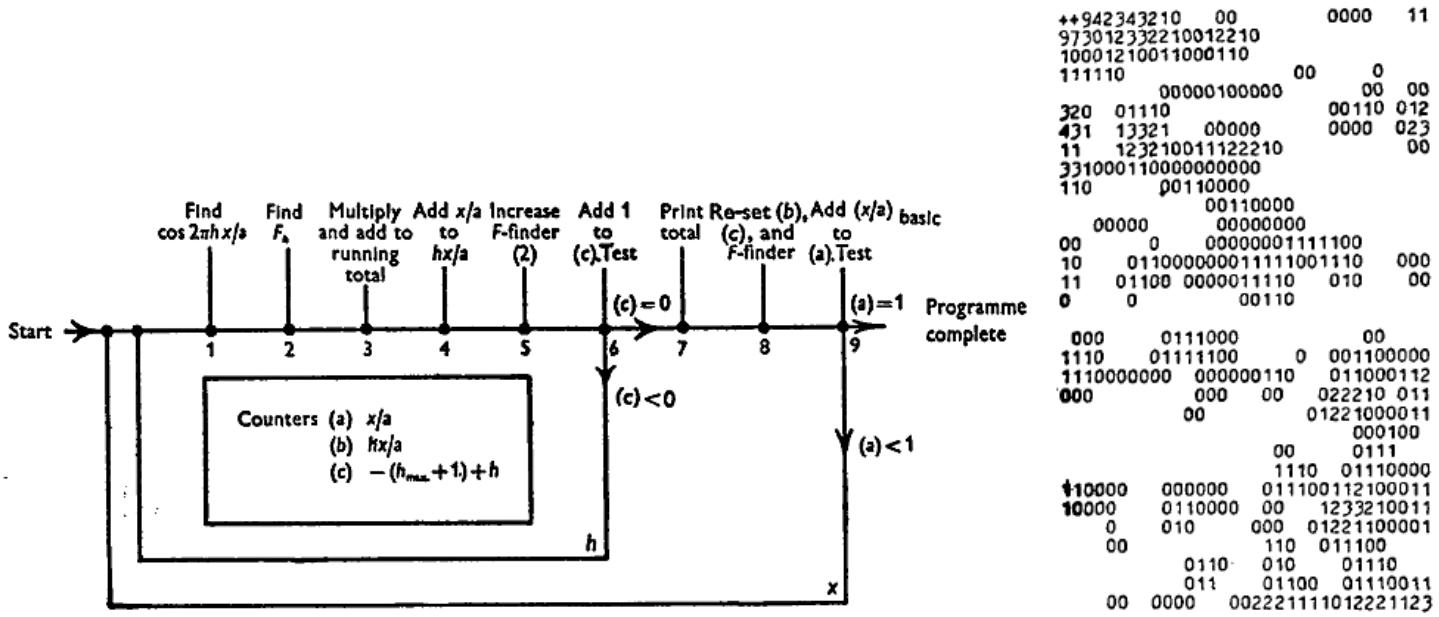


Figure 44: Images from Bennett and Kendrew (1952). Left: an algorithm for calculating a Fourier transform in one dimension. Right: a print-out giving information for drawing electron density contours (adjacent numbers of the same value belonging to the same contour plane). Reproduced with permission of the International Union of Crystallography.

Perutz, now working more closely with Bragg, made another attack on haemoglobin structure. This time, their strategy was to study changes in the sign of the structure factor under variation in the salinity of its surrounding liquid, deduce the breadth and packing of molecules from these, and use them to infer features of haemoglobin's external shape.[137] Chastened by critiques of prior models, they favored a more directly empirical approach, "If the transform is known, the crystal structure is deduced without making any assumptions about the nature of the molecule" (Bragg & Perutz, 1952, p. 425). Similarly, they considered it "premature to speculate on the detailed structure here. It appears certain that the molecule is a far more complex entity than a simple picture of parallel chains would suggest" (p. 434). With these

[137] The basic principle of the method are as follows: "When salt solution is substituted for water as the suspension medium of a protein crystal, the contrast in electron density between the molecule and the surrounding liquid is diminished. The value of the scattered amplitude *F* for any given diffraction may be regarded as the sum of two parts, one due to the intrinsic scattering by the structure in the molecule itself, and the other to that of the liquid in which the molecule is bathed. As has been explained in the previous paper, the changes in *F* when the density is changed from ρ_1 to ρ_2 are the *F* values for a unit which has the outer form of the molecule and a uniform density $\rho_2 - \rho_1$" (Bragg, Howells, & Perutz, 1954, p. 41). Because the crystals only shrink along a single axis, the diffraction spots due to the molecule shift in position in only one dimension in the X-ray photograph. By plotting the changes of *F* along each layer line of the photograph, a series of nodes and loops are formed. By estimating the sign of *F* (i.e., its phase) at each node, a picture of the electron density of the molecule is formed (as in Fig. 45).

precautions in mind, they concluded that the molecule had an ellipsoidal shape with dimensions of approximately 65 x 55 x 55 angstroms (Fig. 45).[138] Despite its low resolution, Bragg later upheld this as "the first definitive quantitative piece of knowledge to be won" (1965, p. 7).

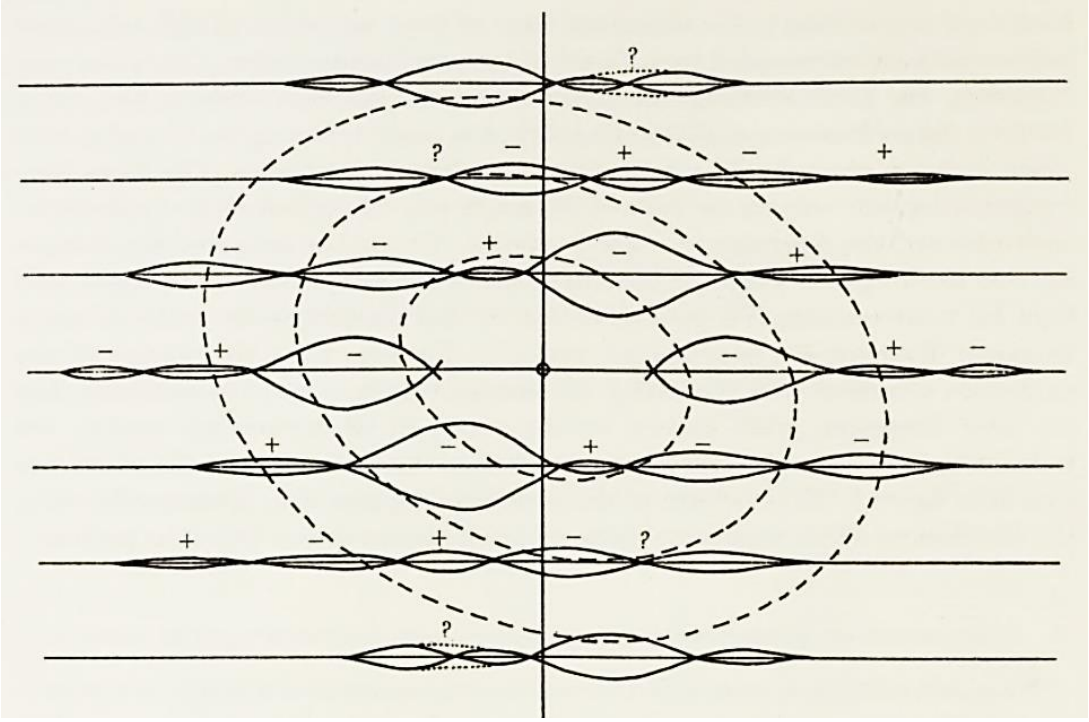


Figure 45: Diagram of the shape of haemoglobin corresponding to structure factor node and loop plots from Bragg, Howells, & Perutz (1954). See footnote 82 for an explanation of this method.

By this time, Perutz had successfully learned to apply Robertson's isomorphous replacement technique for haemoglobin. In 1953 he had obtained a reprint from Austin Riggs showing that mercury atoms could be attached to the sulphydryl group of haemoglobin without affecting its oxygen uptake. To Perutz, this suggested its configuration was not significantly altered by the heavy mercury atoms, meaning the protein structure was largely the same. Diffraction from the modified haemoglobin showed noticeable changes in reflection intensities. Perutz now shifted his focus to the internal structure of haemoglobin, using the same methods employed in his work with Bragg. However, this depended on accurate estimations of phase, the determination of which was still too difficult given the sheer density of atoms photographed and relatively low resolving power of the X-rays in use. Still, a path forward was in sight: Hodgkin advised

[138] By the 1960 structure, this would be updated to 64 x 55 x 50.

Perutz that difficulties with phase could be resolved by triangulating the present data with a second heavy atom derivative. To this end, Howard Dinitz, a chemist with knowledge of heavy atom complexes, was recruited to the MRC and put to the task. In May of 1957, Perutz reported on their progress at a US National Academy of Sciences conference on haemoglobin. Using two variations of mercury-bound haemoglobin, Perutz, his student Ann Cullis, and Dinitz were able to produce a high resolution map of a single two-dimensional reflection of haemoglobin. The image, Perutz admitted, was confounding: "clearly the projection contains too little information as yet to give any clue to the structure of the molecule […] It is evident that the riddle cannot be solved without a solution of the crystal structure in three dimension" (Cullis, Dinitz, & Perutz, 1958).

In the meantime, in 1958 Dinitz had found ways to produce heavy derivatives of myoglobin, yielding diffraction data that Kendrew fed into his computer program to produce a 6-angstrom three-dimensional Fourier synthesis of the complete molecule. This result convinced a skeptical Perutz of the value of the EDSAC and further galvanized the haemoglobin team, but Dinitz struggled to bind further atoms to the protein. He eventually departed for the US with the structure incomplete, leaving behind numerous untried crystals soaking in heavy atom solution. Cullis discovered these in a cupboard and, after some trial and error, found some produced the desired intensity changes (Finch, 2007). At last, everything was in place to properly derive structural information from blood crystals.

Haemoglobin had many more atoms than myoglobin, produced more reflections, and so called for more powerful computing. Luckily the EDSAC Mark II, armed with twice the memory of its predecessor, arrived in 1958. 40,000 reflections were collected on 9-degree precession photographs from six heavy atom derivatives. These were run through the complete chain of data processing—densitometers, diffractometers, and several days of EDSAC calculations. The results were further analyzed and run through additional computer programs designed to determine heavy atom parameters and, finally, calculate the most probable phases of the original structure factor. In 1960, the resulting map of haemoglobin was published at 5.5-angstrom resolution, constructed from 32 electron density maps in parallel layers at 2-angstrom intervals.

The team built an accompanying three-dimensional model from thermo-setting plastic to aid in its visualization (Fig. 46).

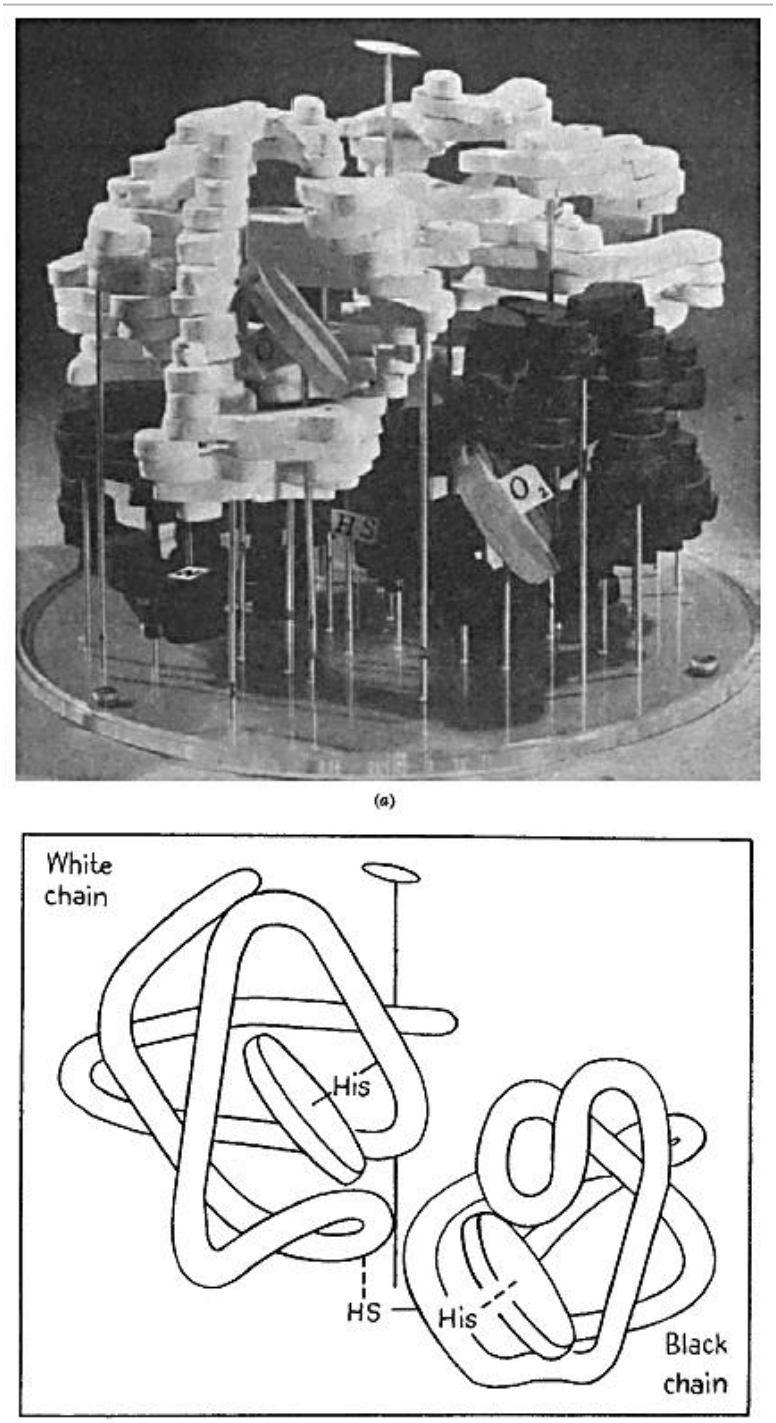


Figure 46: Three-dimensional model showing Perutz et al.'s (1960) Fourier synthesis of haemoglobin at 5.5 angstrom resolution. Top: a photo of the thermo-setting plastic model. Bottom: a schematic drawing isolating two of the four sub-units, clarifying the three-dimensional course of the polypeptide chain, the location of the heme group (disc shapes), and one of the sulphydryl groups (a preferred binding site in heavy atom derivatives).

The model was consistent with earlier claims of haemoglobin's ellipsoidal form (with dimensions now refined to 64 x 55 x 50), its containing four distinct polypeptide sub-units forming "more or less cylindrical clouds of high density," (Perutz & al., 1960, p. 417) and a nearly dyadic symmetry now ascribed to a pairwise identity between two "white chain" and "black chain" sub-units. They located the iron atoms at the highest peaks of electron density. These were the centerpiece of the four heme groups known to bind free oxygen molecules, distinguished in the form of large discs in the plastic model.

This model, along with Kendrew's myoglobin study, provided a definitive answer to a central question posed by Bernal (1939): whether proteins were comprised of a single chain polypeptide or sub-units. The answer was both: a single chain in the case of myoglobin, four sub-units in the case of haemoglobin. This helped do away with the notion, also entertained by Bernal, that all proteins shared a common form of arrangement. The haemoglobin model enabled inquiry into a further question Bernal raised regarding the nature of the linkage between sub-units, should they exist. A section of Perutz et al.'s paper was devoted to this matter, observing "comparatively little contact between the members of each [black-white] pair" while noting a "large area of contact" connoting a "striking" structural complementarity between the black and white chains in each pair (1960, p. 419). This suggested regions of flexibility and inflexibility, respectively, in the molecule's structure. But the team of authors was reticent to make claims about the relation between structure and function, save to remark that the heme groups were too far apart to directly influence oxygen uptake of their neighbors. The point of interest here was the phenomenon of allostery, in which binding to an oxygen molecule increases haemoglobin's affinity for oxygen. "Whatever interaction between the haem groups exists must be of a subtle and indirect kind that we cannot yet guess" (p. 421).

Further refinement and analysis of this model would preoccupy Perutz through the 1960s. By 1965, data from amino acid sequencing and higher-resolution sequencing of myoglobin could be paired with improved diffraction data to give "a tentative atomic model" of haemoglobin, building out the polypeptide chains from known atom positions. At this point, Perutz began to conjoin the word "function" with "structure" in the titles of his papers, as the structural model became a vehicle for drawing inferences about

haemoglobin behavior. This included comparisons between reduced and oxygenated horse haemoglobin. The most notorious result of this work was Perutz's explanation of allostery in the early 1970s in terms of a conformational shift in the molecule from a "tense" to a "relaxed" state, induced upon its initial binding with oxygen (reviewed against further evidence in (Perutz, 1990b)). Lehmann's complaints in 1851 over the "much-disputed question of the interchange of gases in the circulating blood," an interchange characterized as a redox reaction by Stokes in 1864, could now be approached in mechanistic detail using the molecular models rendered through X-ray crystallography.

4.4. Summary

These were the key developments in the technology and techniques of data production and processing on the road to a structural model of the haemoglobin protein. As I have shown, elements of the X-ray crystallography program go back to the mid-19th century; many were carried from there into the Cavendish lab. A theoretical analysis of the historical patterns revealed here will be presented in the following chapter. Here I will simply review and summarize the major moments in this developmental sequence described in this chapter.

The nineteenth study study of blood crystals occurred at the intersection of physiological chemistry and crystallography. The first three decades, spanning from Hünefeld's discovery in 1840 through Preyer's 1871 monograph, saw a proliferation of crystallization protocols, a reflexive understanding of the role of particular steps in a protocol, leading to a gradual convergence on procedure and corresponding vocabulary (such as 'laking' the blood). The crystals were studied in various ways: by goniometer, polariscope, elementary analysis, and somewhat later, by spectroscopic analysis. Using the first three techniques, researchers recorded a variety of differences in crystal type and chemical content for blood harvested from different animals. Spectroscopy led to the identification of the crystals' contents with the "coloring matter" of blood, the name for a substance combining the iron-containing haematin and the protein-like globulin,

for which Hoppe coined haemoglobin. It also allowed distinctions to be made between oxy-, deoxy-, and methaemoglobin, along with several other variants based on binding partners. Haemoglobin was thus identified as a carrier of oxygen within the blood, but these forms of data production offered limited insight into the structure of this substance. After compiling many tables of crystallographic data, Reichert and Brown could not say more of haemoglobin's structure at the end of their 1909 survey than was known at its beginning.

In the meantime, physiological chemists (some starting to adopt the title "biochemist") continued to probe the building blocks of proteins. Chemical formulae, now written in a form denoting structural information, were wielded by researchers like Fischer and Hofmeister to determine the chemical groups and bond types that served as the building blocks of proteins. As Fisher and his students zeroed in on the newly minted amino acids and polypeptides using synthetic methods, colloid chemists pushed back against their mechanistic view of organic matter, arguing that proteins were a form of aggregative protoplasm best studied within the living cell. It was only in the late 1920s, with the advent of improved membranes for diffusion studies of molecular weight, crystallization of novel enzymes, and the development of the ultracentrifuge, that the evidence began to accumulate against colloidists. By the mid-1930s, many were convinced that enzymes were proteins, that proteins had a determinate molecular structure, and that this could be studied outside of the cell without losing their vital characteristics. The "protein paradigm," as Kay (1993) calls it, increasingly gained traction in biologists' understanding of basic life processes.

The analytic and synthetic methods employed by chemists such as Fischer and Abderhalden supported the notion that proteins were built of a common set of building blocks, combined in a recurring pattern of bonds, yet they faltered with models of large-scale protein structure. Fischer would have been shocked at the molecular weight measurements produced by Svedberg. On the other hand, X-ray crystallographers had by this time developed a highly technical method, informed by a mathematical physical theory, for probing the atomic structure of crystals. Leveraging the transverse wave model of X-rays alongside a range of correction factors, they moved from trial-and-error model-building to a direct method for inferring electron distribution and atom positions from diffraction data, at the same time

building up a new evidential corpus and representational vocabulary for chemical structure. Just as consensus was forming around their molecular nature, Bernal and Crowfoot found a means for preparing globular proteins in solution for X-ray analysis. Researchers hoped this method would offer deeper insight into a number of hunches about proteins, in particular the idea that protein structure obeyed certain strict rules—with respect to the order and frequency of amino acids (Niemann and Bergmann); with respect to the geometry of bonds (Wrinch's cyclol model)—that gave them a common, regular form.

These might have inspired Perutz's first model of haemoglobin, based on interpretation of X-ray photographs from crystals in various states of shrinking of swelling in saline solution. Perutz and his assistants studied the distribution of liquid within the crystals, its points of boundary with the higher-density protein, and the internal structure to arrive at a crystal model suggesting a hexagonal pill box packing of proteins separated by liquid, with each cylindrical protein comprised of parallel layers in two-dimensions of polypeptide chains. An attempt by Bragg and Perutz to back this up with a three-dimensional Patterson analysis commissioned from a punched card machine was dismantled by Crick's criticisms and Pauling and Corey's model of the helical structure of polypeptide chains. Holding more closely to "direct" methods of inference from intensity data, the pair sketched an image of haemoglobin that was ellipsoidal in its external form but so electronically dense that its internal structure resisted easy interpretation under low-resolution diffractions.

Higher resolution images resulted in more diffraction spots, which called for more taxing analysis. Perutz's success with isomorphous replacement offered a solution to the phase problem for proteins, but also meant a further influx of reflection data. The Beevers-Lipson strips of old were deeply impractical in the face of Fourier calculations that ran to the order of 10^9 summands. Bennett and Kendrew's EDSAC program, however, was able to handle the required computations in about twelve hours. The resulting model depicted a protein comprised of four winding chain sub-units, forming two symmetrical pairs, with each chain hosting a ferric binding site for oxygen. At this point, nearly a century after variations in spectroscopic data led Hoppe to view haemoglobin as an oxygen carrier, variations that Reichert and Brown dimly probed by recording crystal angles and pleochroism, the phenomenon could be explained mechanistically in terms

of conformational changes in the protein molecule. In the process, researchers had moved through a series of techniques that allowed for the production of ever-finer data, to which protein models provided a three-dimensional interpretation.

References

Abraham, E., & Robinson, R. (1937). Crystallization of Lysozyme. *Nature, 140*, 24.

Allen, H. (1947). Charles Glover Barkla, 1877-1944. *Obituary Notices of Fellows of the Royal Society, 5*, 341-366.

Astbury, W. (1922). The Crystalline Structure and Properties of Tartaric Acid. *Proceedings of the Royal Society, A, 102*, 506-528.

Astbury, W. (1929). A New Integrating Photometer for X-ray Crystal Reflections, etc. *Proceedings of the Royal Society, A, 123*, 575-602.

Astbury, W. (1939). X-ray Studies of the Structure of Compounds of Biological Interest. *Annual Review of Biochemistry, 8*, 113-133.

Astbury, W., & Lomax, R. (1934). X-ray Photographs of Crystalline Pepsin. *Nature, 133*, 795.

Astbury, W., & Street, A. (1932). X-Ray Studies of the Structure of Hair, Wool, and Related Fibres. I. General. *Philosophical Transaction of the Royal Society of London, A, 230*, 75-101.

Astubry, W., & Bragg, W. H. (1925). Annual Reports on the Progress of Chemistry. In J. M. Bijvoet, W. G. Burgers, & G. Hägg (Eds.), *Early Papers on Diffraction of X-rays by Crystals, Vol. 2* (Vol. 20, p. 233). Utrecht: A. Oosthoek.

Barkla, C. (1905). Polarised Röntgen Radiation. *Philosophical Transactions of the Royal Society, A, 204*, 467-479.

Barkla, C. (1967). Characteristic Röntgen Radiation. In *Nobel Lectures, Physics 1901-1921* (pp. 393-399). Amsterdam: Elsevier.

Barkla, C., & Martyn, G. (1913). An X-ray Fringe System. *Nature, 90*, 647.

Barlow, W. (1883). Probable Nature of the Internal Symmetry of Crystals. *Nature, 29*, 186-188.

Bennett, J., & Kendrew, J. (1952). The Computation of Fourier Syntheses with a Digital Electronic Calculating Machine. *Acta Crystallographica, 5*, 109-116.

Bernal, J. D. (1927). A universal X-ray photogoniometer. *Journal of Scientific Instruments, 4*, 273-284.

Bernal, J. D. (1930). The place of X-ray crystallography in the development of modern science. *Radiology, 15*, 2-12.

Bernal, J. D. (1939). Structure of Proteins. *Nature, 143*, 663-667.

Bernal, J. D., & Crowfoot, D. (1934). X-Ray Photographs of Crystalline Pepsin. *Nature, 133*, 794-795.

Bernal, J. D., Crowfoot, D., Evans, R., & Wells, A. (1935). Crystallography (1934-1935). *Annual Reports on the Progress of Chemistry, 32*, 183-242.

Bernal, J. D., Crowfoot, D., Robinson, B., & Wooster, W. (1933). Crystallography. *Annual Reports on the Progress of Chemistry, 30*, 360-430.

Bernal, J. D., Fankuchen, I., & Perutz, M. (1938). An X-ray Study of Chymotrypsin and Haemoglobin. *Nature, 141*, 523-524.

Berzelius, M. (1813). Sur la composition des fluides animaux. *Annales de Chimie, 31*, 26-72.

Boyes-Watson, J., Davidson, E., & Perutz, M. (1947). An X-ray study of horse methaemoglobin. I. *Proceedings of the Royal Society, A, 191*, 83-132.

Bragg, W. H. (1912). X-rays and Crystals. *Nature, 90*, 219.

Bragg, W. H. (1912). X-rays and Crystals. In J. M. Bijvoet, W. G. Burgers, & G. Hägg (Eds.), *Early Papers on Diffraction of X-rays by Crystals, Vol. 1* (pp. 27-29). Utrecht: International Union of Crystallography.

Bragg, W. H. (1914). The Intensity of Reflexion of X-rays by Crystals. In J. M. Bijvoet, W. G. Burgers, & G. Hägg (Eds.), *Early Papers on Diffraction of X-rays by Crystals, Vol. 1* (pp. 103-110; 160). Utrecht: A. Oosthoek.

Bragg, W. H. (1915). X-rays and Crystal Structure. In J. M. Bijvoet, W. G. Burgers, & G. Hägg (Eds.), *Early Papers on Diffraction of X-rays by Crystals, Vol. 1* (pp. 111-114). Utrecht: A. Oosthoek.

Bragg, W. H. (1921). The Intensity of X-ray Reflexion by Diamond. In J. M. Bijvoet, W. G. Burgers, & G. Hägg (Eds.), *Early Papers on Diffraction of X-rays by Crystals* (p. 303). Utrecht: A. Oosthoek.

Bragg, W. H., & Bragg, W. L. (1913). The Reflection of X-rays by Crystals. *Proceedings of the Royal Society, A, 88*, 428-438.

Bragg, W. H., & Bragg, W. L. (1915). *X Rays and Crystal Structure.* London: G. Bell and Sons.

Bragg, W. H., & Glasson, J. L. (1907). On a Want of Symmetry shown by Secondary X-rays. *Proceedings of the Physical Society of London, 21*, 735-745.

Bragg, W. L. (1912). The Specular Reflection of X-rays. *Nature, 90*, 410.

Bragg, W. L. (1913). The Diffraction of Short Electromagnetic Waves by a Crystal. *Proceedings of the Cambridge Philosophical Society, 17*, 43-57.

Bragg, W. L. (1914a). The Structure of Some Crystals as Indicated by their Diffraction of X-rays. In J. M. Bijvoet, W. G. Burgers, & G. Hägg (Eds.), *Early Papers on Diffraction of X-rays by Crystals, Vol. 1* (pp. 59-71). Utrecht: A. Oosthoek.

Bragg, W. L. (1914b). The Analysis of Crystals by the X-ray Spectrometer. *Proceedings of the Royal Society, A, 89*, 468-489.

Bragg, W. L. (1920). The Arrangement of Atoms in Crystals. In J. M. Bijvoet, W. G. Burgers, & G. Hägg (Eds.), *Early Papers on Diffraction of X-rays by Crystals, Vol. 2* (pp. 34-39). Utrecht: A. Oosthoek.

Bragg, W. L. (1925a). A note on the paper 'The Atomic Structure Factor in the Intensity of Reflextion of X-rays by Crystals'. In J. M. Bijvoet, W. G. Burgers, & G. Hägg (Eds.), *Early Papers on Diffraction of X-rays by Crystals, Vol. 1* (pp. 298-299). Utrecht: A. Oosthoek.

Bragg, W. L. (1925b). The Interpretation of Intensity Measurements in X-ray Analysis of Crystal Structure. In J. M. Bijvoet, W. G. Burgers, & G. Hägg (Eds.), *Early Papers on Diffraction of X-rays by Crystals, Vol. 1* (pp. 304-306). Utrecht: A. Oosthoek.

Bragg, W. L. (1929). A Note on the Representation of Crystal Structure by Fourier Series. In J. M. Bijvoet, W. G. Burgers, & G. Hägg (Eds.), *Early Papers on Diffraction of X-rays by Crystals, Vol. 2* (pp. 391-403). Utrecht: A. Oosthoek.

Bragg, W. L. (1965). First stages in the x-ray analysis of proteins. *Reports on Progress in Phsyics, 28*, 1-14.

Bragg, W. L. (1975). *The Development of X-ray Analysis*. New York: Hafner Press.

Bragg, W. L., & Perutz, M. (1952). The structure of haemoglobin. *Proceedings of the Royal Society, A, 213*, 425-435.

Bragg, W. L., & West, J. (1928). A Technique for the X-ray Examination of Crystal Structures with many Parameters. In J. M. Bijvoet, W. G. Burgers, & G. Hägg (Eds.), *Early Papers on Diffraction of X-rays by Crystals, Vol. 2* (Vol. 69, pp. 383-385). Utrecht: A. Oosthoek.

Bragg, W. L., Darwin, C., & James, R. (1926). The Intensity of Reflexion of X-rays by Crystals. In J. M. Bijvoet, W. G. Burgers, & G. Hägg (Eds.), *Early Papers on Diffraction of X-rays by Crystals, Vol. 1* (Vol. 1, pp. 306-313). Utrecht: A. Oosthoek.

Bragg, W. L., Howells, E. R., & Perutz, M. (1952). Arrangement of Polypeptide Chains in Horse Methaemoglobin. *Acta Crystallographica, 5*, 136-141.

Bragg, W. L., Howells, E., & Perutz, M. (1954). The structure of haemoglobin. II. *Proceedings of the Royal Society, A, 222*, 33-44.

Bragg, W. L., James, R. W., & Bosanquet, C. H. (1921). The Intensity of Reflexion of X-rays by Rock-Salt. In J. M. Bijvoet, W. G. Burgers, & G. Hägg (Eds.), *Early Papers on Diffraction of X-rays by Crystals, Vol. 1* (Vol. 41, pp. 265-291). Utrecht: A. Oosthoek.

Bragg, W. L., James, R. W., & Bosanquet, C. H. (1922). The Distribution of Electrons around the Nucleus in the Sodium and Chlorine Atoms. *Philosophical Magazine, 44*, 433-449.

Bravais, A. (1866). *Etudes cristollagraphiques.* Paris: Gauthier-Villars.

Brewster, D. (1831). *Optics.* London: Longman, Rees, Orme, Brown, and Green.

Brown, A. (1864). On the Theory of Isomeric Compounds. *Journal of the Chemical Society of London, 18,* 230-245.

Buerger, M. (1942). *X-Ray Crystallography.* New York: Wiley & Sons.

Compton, A. (1917). The Intensity of X-ray Reflection, and the Distribution of the Electrons in Atoms. *Physical Review, 9,* 29-57.

Compton, A. (1923). The Quantum Integral and Diffraction by a Crystal. *Proceedings of the National Academy of Sciences, 9,* 359-362.

Cox, E., & Crowfoot, D. (1936). Crystallography. *Annual Reports on the Progress of Chemistry, 33,* 196-227.

Cranswick, L. (2008). Busting out of crystallograph's Sisyphean prison: from pencil and paper to structure solving at the press of a button. *Acta Crystallographica, A, 64,* 65-87.

Crowfoot, D. (1938). The crystal structure of insulin I. The investigation of air-dried insulin crystals. *Proceedings of the Royal Society, A, 164,* 580-602.

Crowfoot, D. (1972). Dorothy Crowfoot Hodgkin - Biographical. In *Nobel Lectures, Chemistry 1963-1970.* Amsterdam: Elsevier.

Cullis, A., Dinitz, H., & Perutz, M. (1958). X-ray analysis of haemoglobin. In *Conference on Hemoglobin: 2-3 May 1957.* Washington DC: National Academies Press.

Dalton, J. (1803). On the absorption of gases by water and other liquids. In *Foundations of the Atomic Theory* (pp. 15-26). Edinburgh: William F. Clay.

Darwin, C. (1914). The Theory of X-ray Reflexion. *Philosophical Magazine, 27,* 315-333.

Daston, L., & Galison, P. (2007). *Objectivity.* New York: Zone Books.

de Chadarevian, S. (2003). *Designs for Life.* New York: Cambridge University Press.

Debye, P. (1913). Über die Intensitätsverteilung in den mit Röntgenstrahlen erzeugten Inferenz-bildern. *Verhandlungen der Deutschen physikalischen Gesellschaft, 15,* 738-752.

Debye, P. (1914). Interferenz von Röngtenstrahlen und Wärmebewegung. *Annalen der Physik, 43,* 49-95.

Debye, P., & Scherrer, P. (1916). Interferenz an regellos orientierten Teilchen im Röntgenlicht I. *Physikalische Zeitschrift, 17,* 277-283.

Deichmann, U. (2007). “Molecular” versus “Colloidal”: Controversies in Biology and Biochemistry, 1900-1940. *Bulletin of the History of Chemistry, 32,* 105-118.

Donné, A. (1844). *Cours de Microscopie.* Paris: J.-B. Bailliere.

Dounce, A., & Allen, P. (1988). Fifty years later: recollections of the early days of protein crystallization. *TIBS, 13*, 317-320.

Duane, W. (1925). The Calculation of the X-ray Diffracting Power at Points in a Crystal. *Proceedings of the National Academy of Sciences, 11*, 489-493.

Edsall, J. (1972). Blood and Hemoglobin: The Evolution of Knowledge of Functional Adaptation in a Biochemical System. *Journal of the History of Biology, 5*, 205-257.

Epstein, P., & Ehrenfest, P. (1924). The Quantum Theory of the Fraunhoffer Diffraction. *Proceedings of the National Academy of Sciences, 10*, 133-139.

Ewald, P. (1913). *Physikalische Zeitschrift, 14*, 471.

Ewald, P. (1914). Die Berechnung der Kristallstruktur aus Interferenzaufnahmen mit X-strahlen. *Physikalische Zeitschrift, 15*, 399-401.

Ewald, P. (1925). Die Intensitäten der Röntgenreflexe und der Strukturfaktor. *Physikalische Zeitschrift, 24*, 29-32.

Ewald, P. (1962). *Fifty Years of X-ray Diffraction.* Utrecht: International Union of Crystallography.

Finch, J. (2007). *A Nobel Fellow on Every Floor*. Cambridge: The Medical Research Council Laboratory of Molecular Biology.

Finlay, M. (1992). *Science, practice and politics: German agricultural experiment stations in the nineteenth century.* Dissertation: Iowa State University.

Fischer, E. (1907). Die Chemie der Proteine und ihre Beziehungen zur Biologie. In E. Fischer, & M. Bergmann (Ed.), *Untersuchungen über Aminosäuren, Polypeptide, und Proteine II* (pp. 1-21). Berlin: Springer.

Forman, P. (1969). The Discovery of the Diffraction of X-rays by Crystals; A Critique of the Myths. *Archive for History of Exact Sciences, 6*, 38-71.

Foster, M. (1880). *A Textbook of Physiology.* New York: MacMillan and Co.

Frank, R. (1987). American Physiologists in German Laboratories, 1865-1914. In *Physiology in the American Context, 1850-1940* (pp. 11-46). New York: Springer.

Friedrich, W. (1913). Röntgenstrahlinterferenzen. *Physikalische Zeitschrift, 14*, 1079-1087.

Friedrich, W., Knipping, P., & Laue, M. (1913). Interferenzerscheinungen bei Röntgenstrahlen. *Annalen der Physik, 41*, 971-988.

Fruton, J. (1979). Early Theories of Protein Structure. *Annals of the New York Academy of Sciences, 325*, 1-20.

Fruton, J. (1999). *Proteins, Enzymes, Genes.* New Haven: Yale University Press.

Funke, O. (1851). Ueber das Milzvenenblut. *Zeitschrift für Rationelle Medicin, 9*, 172-218.

Funke, O. (1853). *Atlas of physiological chemistry.* London: Cavendish Society.

Funke, O. (1876). *Lehrbuch der Physiologie.* Leipzig: Leopold Voss.

Galison, P. (1987). *How Experiments End.* Chicago: University of Chicago Press.

Gamgee, A. (1898). Haemoglobin: Its compounds and the principal products of its decomposition. In E. A. Schäfer (Ed.), *Text-book of Physiology* (pp. 185-260). New York: MacMillan.

Giegé, R. (2013). A historical perspective on protein crystallization from 1840 to the present day. *FEBS Journal, 280,* 6456-6497.

Glasser, O. (1936). What Kind of Tube did Röntgen Use When He Discovered the X-ray? *Radiology, 27,* 138-140.

Graham, T. (1861). Liquid Diffusion applied to Analysis. *Philosophical Transactions of the Royal Society, 151,* 183-224.

Granick, S. (1942). Some properties of crystalline guinea pig haemoglobin. *The Journal of General Physiology, 25,* 571-578.

Haga, H., & Wind, C. H. (1899). Die Beugung der Röntgenstrahlen. *Annalen der Physik und Chemie, 68,* 884-916.

Hartree, D. (1925). The Atomic Structure Factor in the Intensity of Reflexion of X-rays by Crystals. In J. M. Bijvoet, W. G. Burgers, & G. Hägg (Eds.), *Early Papers on Diffraction of X-rays by Crystals, Vol. 1* (pp. 297-298). Utrecht: A. Oosthoek.

Haurowitz, F. (1949). The Internal Structure of Protein Molecules. *Experientia, 5,* 347-354.

Haüy, R. (1784). *Essai d'une Théorie sur la Structure des Crystaux.* Paris: Gogué & de la Rochelle.

Haüy, R. (1822). *Traité de Cristallographie.* Paris: Bachelier & Huzard.

Havighurst, R. (1925). The Distribution of Diffracting Power in Sodium Chloride. *Proceedings of the National Academy of Sciences, 11,* 502-507.

Hodgkin, D. (1950). X-ray Analysis and Protein Structure. *Cold Spring Harbor Symposium on Quantitative Biology, 14,* 65-78.

Hodgkin, D. (1990). Taking Penicillin to the R.I. In J. Thomas, & D. Phillips (Eds.), *The Legacy of Sir Lawrence Bragg* (pp. 91-95). Northwood: Science Reviews.

Holmes, F. (1993). Justus Liebig and the Construction of Organic Chemistry. In S. Mauskopf (Ed.), *Chemical Sciences in the Modern World* (pp. 119-136). Philadelphia: University of Pennsylvania Press.

Holmes, F. (1995). Crystals and Carriers: The Chemical and Physiological Identification of Hemoglobin. In *No Truth Except in the Details* (pp. 191-244). Dordrecht: Kluwer.

Hoppe, F. (1862). Ueber das Verhalten des Blutfarbstoffs im Spectrum des Sonnenlichtes. *Archiv für pathologischen Anatomie, 23*.

Hünefeld, F. L. (1840). *Der Chemismus in der thierischen Organisation.* Leipzig: F. A. Brockhaus.

Kahr, B., & Shtukenberg, A. (2012). Histories of Crystallography by Shafranovskii and Schuh. In J. Benedict (Ed.), *Recent Advances in Crystallography* (pp. 3-36). London: IntechOpen.

Kassabian, M. (1910). *Röntgen Rays and Electro-Therapeutics.* Philadelphia: Lippincott Company.

Kay, L. (1993). *The Molecular Vision of Life.* New York: Oxford University Press.

Keiser, E. (1901). *An Outline of the Progress of Chemistry in the Nineteenth Century*. Retrieved from Scientific American: https://www.scientificamerican.com/article/an-outline-of-the-progress-of-chemi/

Kendrew, J. (1987). *Talk by Sir John Kendrew on fifty years of Computing in Cambridge.* Cambridge Computer Laboratory.

Kendrew, J. (1990). Bragg's Broomstick and the Structure of Protein. In J. Thomas, & D. Phillips (Eds.), *The Legacy of Sir Lawence Bragg* (pp. 88-91). Northwood: Science Reviews.

Kendrew, J., & Perutz, M. (1948). A comparative X-ray study of foetal and adult sheep haemoglobins. *Proceedings of the Royal Society, A, 194*, 375-398.

Kidd, J. (1809). *Outlines of Mineralogy.* London: Oxford.

Klein, U. (2001). The Creative Power of Paper Tools in Early Nineteenth-Century Chemistry. In *Tools and Modes of Representation in the Laboratory Sciences* (pp. 13-34). New York: Springer.

Klein, U. (2003). *Experiments, Models, Paper Tools.* Stanford: Stanford University Press.

Kohler, R. (1994). *Lords of the Fly.* Chicago: University of Chicago Press.

Kohler, R. (2008). *From medical chemistry to biochemistry.* New York: Cambridge University Press.

Landecker, H. (2007). *Culturing Life.* Cambridge, MA: Harvard University Press.

Laszlo, P. (2001). Conventionalities in Formula Writing. In *Tools and Modes of Representation in the Laboratory Sciences* (pp. 47-60). New York: Springer.

Law, J. (1973). The Development of Specialties in Science: The Case of X-ray Protein Crystallography. *Science Studies, 3*, 275-303.

Lehmann, C. G. (1855). *Physiological Chemistry.* (G. Day, Trans.) Philadelphia: Blanchard & Lea.

Liebig, J. (1839). *Instructions for the Chemical Analysis of Organic Bodies.* Glasgow: Griffin & Company.

Loeb, J. (1912). *The Mechanistic Conception of Life.* Chicago: University of Chicago Press.

Lonsdale, K. (1929). X-ray Evidence on the Structure of the Benzene Nucleus. *Transactions of the Faraday Society, 25*, 352-366.

Mitchell, S. W. (1858). Observations on the Blood Crystals of the Sturgeon. *Proceedings of the Academy of Natural Sciences of Philadelphia, 10*, 2-4.

Mitchell, S., & Gronenborn, A. (2017). After Fifty Years, Why Are Protein X-ray Crystallographers Still in Business? *Philosophy of Science, 68*, 703-723.

Mitscherlich, E. (1843). *Über ein Goniometer.* Berlin.

Mohs, F. (1825). *Treatise on Mineralogy.* (W. Haidinger, Trans.) Edinburgh: Constable and Co.

Mould, R. F. (1995). The early history of X-ray diagnosis with emphasis on the contributions of physics 1895-1915. *Physics in Medicine & Biology*, 1741-1787.

Mulder, G. J. (1838). Sur la Composition de Quelques Substances Animáles. *Bulletin des Sciences Physiques et Naturelles*, 104-119.

Northrop, J. (1930). Crystalline Pepsin. *The Journal of General Physiology, 13*, 739-766.

Northrop, J. (1964). The preparation of pure enzymes and virus proteins. In *Nobel Lectures, Chemistry 1942-1962* (pp. 124-134). Amsterdam: Elsevier.

Olby, R. (2012). *The Path to the Double Helix.* New York: Dover.

Osborne, T. (1902). The basic character of the protein molecule and the reactions of edestin with definite quantities of acids and alkalies. *Journal of the American Chemical Society, 24*, 39-78.

Ostwald, W. (1917). *An Introduction to Theoretical and Applied Colloid Chemistry.* (M. Fischer, Trans.) New York: Wiley & Sons.

Patterson, A. (1935). A Direct Method for the Determination of the Components of Interatomic Distances in Crystals. *Zeitschrift für Kristallographie, 90*, 517-542.

Pauling, L., & Corey, R. (1951). Configurations of polypeptide chains with favored orientations around single bonds: two new pleated sheets. *Proceedings of the National Academy of Sciences, 37*, 729-740.

Pauling, L., & Niemann, C. (1939). The Structure of Proteins. *Journal of the American Chemical Society, 61*, 1860-1867.

Perutz, M. (1949). A X-ray study of horse methaemoglobin. II. *Proceedings of the Royal Society, A, 195*, 474-499.

Perutz, M. (1990a). How Lawrence Bragg Invented X-ray Analysis. In J. Thomas, & D. Phillips (Eds.), *The Legacy of Sir Lawrence Bragg* (pp. 71-86). Northwood: Science Reviews.

Perutz, M. (1990b). Mechanisms regulating the reactions of human hemoglobin with oxygen and carbon monoxide. *Annual Review of Physiology, 52*, 1-25.

Perutz, M., Rossmann, M., Cullis, A., Muirhead, H., & Will, G. (1960). Structure of Haemoglobin. *Nature, 185*, 416-422.

Polanyi, M. (1921). Faserstruktur im Röntgenlichte. *Die Naturwissenschaften, 9*, 337-340.

Preyer, W. (1871). *Die Blutkrystalle.* Jena: Mauke's.

Putnam, F. (1994). Felix Haurowitz. *Biographical Memoirs, 64*, 134-163.

Rasmussen, N. (1999). *Picture Control.* Stanford: Stanford University Press.

Reichert, E., & Brown, A. (1909). *The Crystallography of Hemoglobins.* Washington, DC: Carnegie Institution of Washington.

Rheinberger, H. (1997). *Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube.* Palo Alto: Stanford University Press.

Rheinberger, H.-G. (2009). Recent science and its exploration: the case of molecular biology. *Studies in History and Philosophy of Biological and Biomedical Sciences, 40*, 6-12.

Rich, A. (2003). The double helix: a tale of two puckers. *Nature Structural and Molecular Biology, 10*, 247-249.

Ritter, C. (2001). An Early History of Alexander Crum Brown's Graphical Formulas. In *Tools and Modes of Representation in the Laboratory Sciences* (pp. 35-46). New York: Springer.

Rocke, A. (2000). Organic Analysis in Comparative Perspective: Liebig, Dumas, and Berzelius, 1811-1840. In F. Holmes, & T. Levere (Eds.), *Instruments and Experimentation in the History of Chemistry* (pp. 273-310). Cambridge MA: MIT Press.

Rocke, A. (2001). Chemical Atomism and the Evolution of Chemical Theory in the Nineteenth Century. In *Tools and Modes of Representation in the Laboratory Sciences* (pp. 1-12). New York: Springer.

Röntgen, W. (1897). Weitere Beobachtungen über die Eigenschaften der X-Strahlen. *Annalen der Physik und Chemie*, 18-37.

Röntgen, W. (1995). On a New Kind of Rays. *Veterinary Radiology & Ultrasound, 36*, 371-374.

Sagnac, G. (1899). Transformations des rayons X par la matiere. *Éclairage électrique, 18*, 41-48.

Salkowski, E. (1903). *A Laboratory Manual of Physiological and Pathological Chemistry.* New York: Wiley & Sons.

Schickore, J. (2007). *The Microscope and the Eye.* Chicago: University of Chicago Press.

Schickore, J. (2017). *About Method.* Chicago: University of Chicago Press.

Schuh, C. (2007). *Mineralogy & Crystallography.* Tuscon: manuscript.

Séguy, G. (1897). Tubes for the production of Roentgen rays. *Nature, 55*, 296-297.

Seliger, H. (1995). Wilhelm Conrad Röntgen and the Glimmer of Light. *Physic Today, 48*, 25-31.

Shearer, G. (1923). The Relation between Molecular and Crystal Symmetry as Shown by X-ray Crystal Analysis. *Proceedings of the Physical Society of London, 35*, 81-100.

Sohncke, L. (1879). *Theorie der Krystallstruktur.* Leipzig: Teubner.

Sommerfeld, A. (1901). Theoretisches über die Beugung der Röntgenstrahlen. *Zeitschrift für Mathematik und Physik*, 11-97.

Stark, J. (1913). Zur Diskussion des Verhaltens der Röntgenstrahlen in Kristallen. *Physikalische Zeitschrift, 14*, 319-321.

Steensen, N. (1669). The Prodromus to a Dissertation on a Solid Naturally Contained Within a Solid. In N. Steno, T. Kardel, & P. Maquet (Eds.), *Biography and Original Papers* (pp. 621-660). New York: Springer.

Steuwer, R. (1971). Bragg's Corpuscular Theory of X-rays and y-rays. *The British Journal for the History of Science, 5*, 258-281.

Stokes, G. G. (1864). On the reduction and oxidation of the colouring matter of the blood. *Proceedings of the Royal Society of London, 13*, 355-384.

Stokes, G. G. (1898). On the Nature of the Röntgen Rays. In *Röntgen Rays* (pp. 43-65). New York: American Book Company.

Sumner, J. (1926). The isolation and crystallization of the enzyme urease. *Journal of Biological Chemistry, 69*, 435-441.

Sumner, J. (1937). The Story of Urease. *Journal of Chemical Education, 14*, 255-259.

Sumner, J. (1964). The Chemical Nature of Enzymes. In *Nobel Lectures, Chemistry 1942-1962* (pp. 114-121). Amsterdam: Elsevier.

Svedberg, T., & al. (1939). A Discussion on the Protein Molecule. *Proceedings of the Royal Society, B, 127*, 1-40.

Szabadváry, F. (1986). The History of Chemical Laboratory Equipment. *Periodica Polytechnica Chemical Engineering, 30*, 77-95.

Tal, E. (2017). Calibration: Modelling the Measurement Process. *Studies in the History and Philosophy of Science Part A, 65*, 33-45.

Taylor, W., & West, J. (1928). The Crystal Structure of the Chondrodite Series. *Proceedings of the Royal Society of London, A, 117*, 517-532.

Thompson, E. (1896). *Roentgen Rays and Phenomena of the Anode and Cathode*. New York: D. Van Nostrand Company.

Thomson, J. J. (1898). A Theory of the Connection between Cathode and Röntgen Rays. In G. Barker (Ed.), *Röntgen Rays* (pp. 69-72). New York: American Book Company.

Walter, B., & Pohl, R. (1909). Weitere Versuche über die Beugung der Röntgenstrahlen. *Annalen der Physik, 29*, 356.

Warren, B., & Bragg, W. L. (1929). The Structure of Diopside. In J. M. Bijvoet, W. G. Burgers, & G. Hägg (Eds.), *Early Papers on Diffraction of X-rays by Crystals, Vol. 2* (Vol. 69, pp. 320-334). Utrecht: A. Oosthoek.

Watson, G. (2001). *The Double Helix.* New York: Touchstone.

Weissenberg, K. (1924). Ein neues Röntgengoniometer. *Zeitschrift für Physik, 23*, 229-238.

Wheaton, B. (1983). *The tiger and the shark: Empirical roots of wave-particle dualism.* New York: Cambridge University Press.

Willstätter, R. (1927). Problems and Methods in Enzyme Research. *Journal of the Chemical Society, 0*, 1359-1381.

Wollaston, W. (1809). Description of a reflective Goniometer. *Philosophical Transactions of the Royal Society of London, 99*, 253-258.

Wollaston, W. (1813). On the elementary particles of certain crystals. *Philosophical Transactions of the Royal Society, 103*, 51-63.

Wyckoff, R. (1930). The Atomic Scattering Powers of Nickel, Copper and Iron for Various Wavelengths. *Physical Review, 36*, 1116-1120.

Zinoffsky, O. (1886). Ueber die Grösse des Hämoglobinmolecüls. *Zeitschrift für Physiologische Chemie, 10*, 16-34.

Zsigmondy, R. (1966). Properties of colloids. In *Nobel Lectures, Chemistry 1922-1941* (pp. 45-57). Amsterdam: Elsevier.

5.0. Historical Patterns and the Growth of Knowledge in Empirical Programs

What does the preceding history tell us, with respect to the fundamental questions opening chapter 4? How does appeal to the contents of an empirical program justify the actions taken in data gathering and analysis? And what about the nature of such programs prevents a representational regress, insofar as these programs employ models? How can we account for the accuracy and content of models without assuming a mapping relation between it and a target phenomenon? In this chapter and the next, I will draw on the prior case study to address these questions. Different stages in the development of protein X-ray crystallography illuminate aspects of the practices that provide justificatory bulwark for empirical programs and inform the use of scientific representations.

The present chapter is devoted to a philosophical distilling of the historical developments covered in Chapter 4. It recasts various episodes within this history in terms of general patterns and uses these to characterize features of empirical programs qua complex developmental systems. These are: the search for invariances within the "parameter space" of experimental conditions; the stabilization of procedures around such invariances; the incorporation of resulting "points of stability" into extended procedures and hence the scaffolding of novel techniques on prior results; the reciprocal influence between model-based target characterization and experimental techniques; and higher-scale developmental tendencies that emerge from these. I discuss how these patterns contribute to the increased reliability of data production and interpretation procedures and outline a corresponding theory of historical justification for bodies of knowledge that undergo regular alteration, as empirical programs do. Chapter 6 will once again consider the subject of scientific representation in closer detail.[139]

[139] While I discuss the evaluative judgments that justify an alteration to the methodological views within a program, I do not claim that these are the sole or primary mechanisms responsible for the actual evolution of scientific knowledge, nor do I intend to model that actual process (as with the models discussed by (Wu, O'Connor, & Smaldino, forthcoming 2023)).

5.1. Points of stability in parameter space

First, consider the development of blood protein crystallization. This work occurred at the intersection of two fields, mineralogy and physiological chemistry, and drew from their instrumentation and methodologies. Such methodologies were heavily empirical, consisting primarily of the classification of substances by their observed qualities, alongside elementary and stoichiometric analyses. Corresponding terminology and notation was initially developed to summarize, systematize, and retrodict these results. In the case of blood crystals, the primary concern of practitioners was understanding the chemical nature and physiological role of the crystals' constituents. Understanding these properties depended on a consistent means of preparing the crystals. This spawned a multitude of procedures. I have described this process (Ch. 4, Section 2.3) as an exploration of the crystallization parameter space. Researchers tracked and varied the conditions that gave rise to crystals and recorded the results. The study of these variances grounded causal generalizations regarding such procedures and, over time, a select number of steps persisted throughout operational shifts and tweaks.

We can gain a sense of how this took place by analyzing the progression of blood crystallization protocols over time. Figure 1 shows a select series of protocols employed from Hunefeld's discovery in 1840 to Perutz and Kendrew's joint paper in 1948. Each step that is common to multiple methods has been color coded.

1840 Hunefeld	1851 Funke	1855 Lehmann	1858 Mitchell	1858 Mitchell2	1863 Funke	1864 Hoppe	1871 Rollet	1871 Bottcher	1871 Kuhne	1871 Anon	1886 Zinoffsky	1904 Salkowski	1909 R&B	1935 K&H	1942 Granick	1948 Kendrew
Place between gl	Place between glass	Chop clots	Place on slide	Place in open vess	Dry cells	Defibrinate	Defibrinate	Inject water into ar	Cool in ice	Defibrinate	Defibrinate	Mix with ether	Defibrinate	Defibrinate	Wash in saline	Wash in saline
Dry (in desiccator	Mix with water/alco	Wash	Evaporate by half	Expose to light and	Mix with water/alc	Mix with salts	Mix with salts	Chloroform animal	Mix with salt solut	Mix with water + al	Mix with water + al	Cool to 0C	Mix with ether	Wash in saline	Dilute with water	Cool to -20C
	Evaporate	Defibrinate	Mix with water	Let decompose	Cool	Cool to 0C	Cool to 0C	Gather blood	Wash	Cool to 0C	Let cells sink	Mix with ether + al	Centrifuge	Cool	Centrifuge	Filter
		Dilute with water	Cover with glass	Evaporate	Evaporate	Draw off supernate	Thaw	Mix with water + al	Filter	Collect crystals	Decant	Cool to 0C		Mix with water + et	Discard supernate	Add pot. ferricy.
		Expose to sunlight	and air	Remove drop	Add alcohol or salt	Wash	Repeat previous 2	Evaporate	Repeat previous 2	Dissolve in warm v	Add salts	Filter		Centrifuge	Dissolve in water	Dialyze w/ water
		Conduct carbonic acid		Evaporate		Add water + ether				Cool to 0C	Dilute with water	Press out mother liquor		Extract fluid	Centrifuge	Mix with salt
						Filter				Wait 24 hours	Heat to 35C	Dissolve in water		Filter	Cool to 1C	Evaporate
						Cool to 0C				Collect crystals	Add ammonia	Filter		Cool	Wait overnight	
						Mix with alcohol					Mix with salt + sulf	Mix with alcohol		Add alcohol		
						Cool to -5C to -10C					Cool to 0C	Cool to 0C		Cool		
						Collect crystals					Add alcohol			Wait overnight		
						Wash with water +	alcohol				Wait 24 hours			Collect crystals		
						Dry					Wash with water +	alcohol		Centrifuge		
											Repeat previous 7 steps			Wash in water +	alcohol	
											Dry (in vacuum)			Mix with water		
														Heat to 39C		
														Add sodium hydroxide to dissolve		
														Neutralize with hydrochloric acid		
														Centrifuge		
														Cool		
														Repeat previous 4 steps		
														Dialyze w/ water		

Figure 1: A selection of crystallization protocols from 1840 to 1948.

Protocols clearly lengthened and varied over time, sometimes dramatically. They also began to show common features in the types of steps and their ordering. Figure 2 is a way to illustrate this process. Each line follows the position of a common step over time.

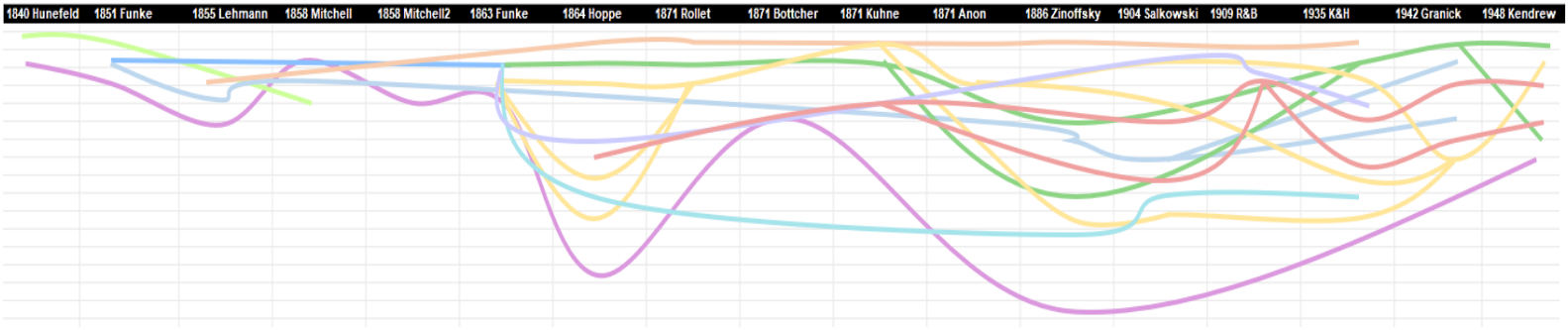


Figure 2: Trajectories of crystallization protocol steps, 1840-1948. Some repetitions of steps removed for clarity.

The extent that the relative position of a given step persists over time is a rough indication of procedural stability. There are issues with this method of analysis that render it a coarse, qualitative basis for historical inference. For one, the protocols are based on authors' presentations, which vary in their levels of details. This can make two protocols look more or less alike than they actually are. For instance, some steps may have been regarded as so obvious to practitioners at a certain time that they were not listed. This might explain the omission of defibrination (light orange) after 1935 and drying/evaporation (purple) between 1886 and 1948 (in Fig. 1), despite their relative constancy as first and last steps. Similarly, Kendrew and Perutz (1948) claimed their protocol (the final column) was a modification of a 1935 protocol from Keilin and Hartree (K&H (1935), third from the right), but their description appears to eliminate a number of steps.[140] A second issue with this method is that the sample here is taken from a limited number of noteworthy sources and does not purport to cover anything close to the entire range of protocols employed over this period. Some of these were listed in source texts as less reliable or less popular methods. With those caveats in mind, a focus on more influential methods may still prove useful.

[140] In this case, it is possible that many of these differences are genuine, and that Kendrew and Perutz were referencing what they viewed as the most important aspect of Keilin and Hartree's technique: the use of potassium ferricyanide to oxidize hemoglobin.

Consider the following four instances (Fig. 3), three of which (due to Hoppe, Zinoffsky, and K&H) produced important results or were used by those who did.

1864 Hoppe	1886 Zinoffsky	1904 Salkowski	1935 K&H
Defibrinate	Defibrinate	Mix with ether	Defibrinate
Mix with salts	Mix with water + alco	Cool to 0C	Wash in saline
Cool to 0C	Let cells sink	Mix with ether + alcol	Cool
Draw off supernatant	Decant	Cool to 0C	Mix water + ether
Wash	Add salts	Filter	Centrifuge
Add water + ether	Dilute with water	Press out mother liqu	Extract fluid
Filter	Heat to 35C	Dissolve in water	Filter
Cool to 0C	Add ammonia	Filter	Cool
Mix with alcohol	Mix with salt + sulfuri	Mix with alcohol	Add alcohol
Cool to -5C to -10C	Cool to 0C	Cool to 0C	Cool
Collect crystals	Add alcohol		Wait overnight
Wash with water + al	Wait 24 hours		Collect crystals
Dry	Wash with water + alcohol		Centrifuge
	Repeat previous 7 steps		Wash water + alc.
	Dry (in vacuum)		Mix with water
			Heat to 39C
			Add NaOH dissolve
			Neutralize with HCl
			Centrifuge
			Cool
			Repeat previous 4
			Dialyze w/ water

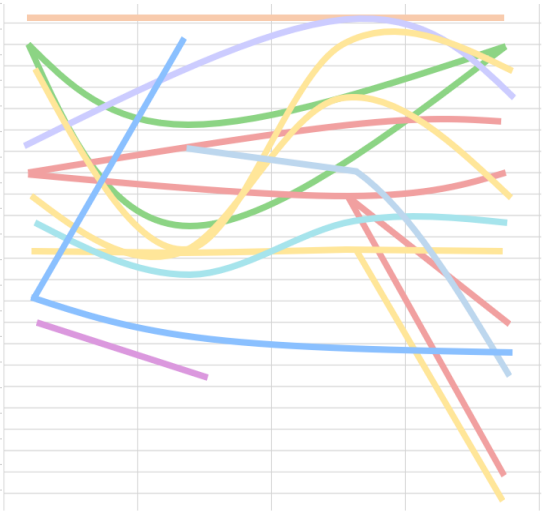


Figure 3: Four influential protocols and their step trajectories.

Certain trends are more visible here: an early use of salts by Hoppe, Zinoffsky, and K&H, likely to help burst cell membranes and separate stroma. Hoppe follows this with cooling, and later a water-ether solution; Salkowski skips the salts and moves directly to cooling and ether treatment. Hoppe, Salkowski, and K&H follow these initial steps with separation/filtration techniques. All four follow the first round of purification with treatment of alcohol and cooling, which Hoppe, Zinoffsky, and K&H follow with an alcohol-water wash once crystals have formed. Consecutive rounds of cooling and filtration also emerge as a prominent element.

Finally, more pronounced evidence of stability is shown by comparison of only Hoppe and K&H:

1864 Hoppe	1935 K&H
Defibrinate	Defibrinate
Mix with salts	Wash in saline
Cool to 0C	Cool
Draw off supernatant	Mix water + ether
Wash	Centrifuge
Add water + ether	Extract fluid
Filter	Filter
Cool to 0C	Cool
Mix with alcohol	Add alcohol
Cool to -5C to -10C	Cool
Collect crystals	Wait overnight
Wash with water + al	Collect crystals
Dry	Centrifuge
	Wash water + alc.
	Mix with water
	Heat to 39C
	Add NaOH dissolve
	Neutralize with HCl
	Centrifuge
	Cool
	Repeat previous 4
	Dialyze w/ water

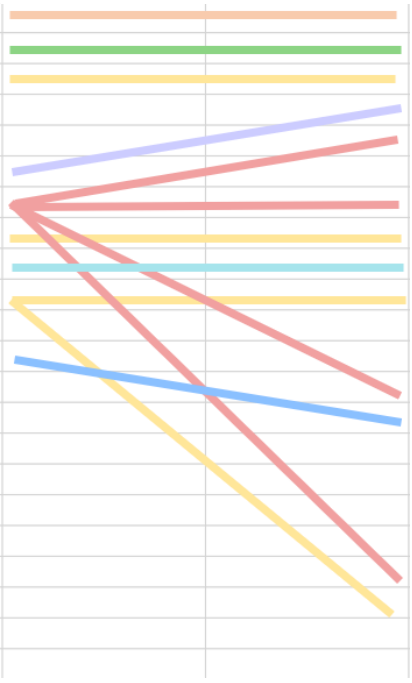


Figure 4: Comparison of protocols from Hoppe and Keilin & Hartree, showing procedural stability from 1864 to 1935.

This shows stark continuity between protocols spanning a 70 year gap. If we ignore the additional filtration/separation and cooling steps (aided by the invention of the centrifuge), all nine of the other steps common to Hoppe and K&H occur in the same relative order. Moreover, the authors selected here are distinctively influential within the history laid out in the previous chapter. Hoppe's work at this time established the identity of blood crystals with the "coloring matter" of blood, which he termed haemoglobin. Salkowski's instructional textbook (1903) cites it as the standard technique some three decades on. Reichert and Brown (1909) devote a disproportionate amount of their historical overview of haemoglobin crystallography to an account of Hoppe's methods, claiming them so successful that there had been little incentive to develop methods further. Keilin ran the Molteno Institute at Cambridge, which specialized in biochemistry. From 1938 to the early 1950s he provided Perutz his lab facilities to for protein chemistry, where he had a direct influence on the latter's crystallization methods. So there is reason to believe that the similarity between these protocols is no accident but rather shows a form of procedural stability running through the tangle of crystallization steps that rose and fell between the 19th and 20th centuries.

Such patterns reflect a tightening in the parameter space of relevant experimental conditions within which crystallization was carried out. This was justified not only by the increasing success scientists found with a particular ordering of steps, but also by the investigation of second-order concerns. Increased rates of success with certain orderings helped establish which of the conditions cited in various protocols were to figure directly in causal generalizations with respect to crystallization methods, and which could be consigned to background conditions.[141] Thus Preyer grew crystals in a range of conditions, from total darkness to direct sunlight, in response to Lehmann and Mitchell's insistence on the role of light, and relegated it to the background. Zinoffsky, in varying his methods, found that washing cell contents after laking was superfluous.[142] Such investigations allow researchers to determine the range of conditions under which generalizations about the interventions involved in crystal preparation remained invariant. Knowledge of these conditions is reflected in those aspects of protocols that are clearly defined (initial defibrination is important for obtaining a pure sample), those where some flexibility is allowed (a range of temperatures below 0°C can be used for cooling; saline and ether solution are equally adequate for laking the blood), and those that fall into irrelevance (light and atmospheric conditions receive little attention after the 1850s). The cumulative effect is to restrict and control variance in the conditions that have the most significant causal effect on crystallization outcomes, and to tolerate variance elsewhere, thereby establishing a particular causal niche in which satisfactory experimental results are produced.

As a result of this work, protein crystallization was stabilized into a procedure that yielded sufficiently consistent results to allow for successful execution by researchers across the European continent and over the Atlantic. Blood crystals were thereby able to serve as points of stability. The outputs of these procedures were studied with a growing set of instruments—the traditional analytical tools of organic chemists, the microgoniometer of the mineralogists, the novel methods of spectroscopy, and eventually X-ray crystallography (Fig. 5). This is an example of what I take to be a general feature in the development

[141] Cf. Woodward (2000b) on the role of this distinction with respect to the kinds of causal generalizations that figure in experiment.
[142] Perhaps it is no coincidence that this step is skipped by K&H.

of empirical programs: *an exploration of the parameter space of conditions accompanying a procedure generates a set of causal generalizations, i.e., statements of dependencies between empirically accessible factors such as background conditions, specific interventions, and measurable effects. These generalizations guide procedures toward consistent outcomes, or points of stability, typically in the form of a reproducible object, instrumental arrangement, or effect. These form the basis for an initial, superficial characterization of a target phenomenon in terms of the characteristic results predicted by these causal generalizations.*

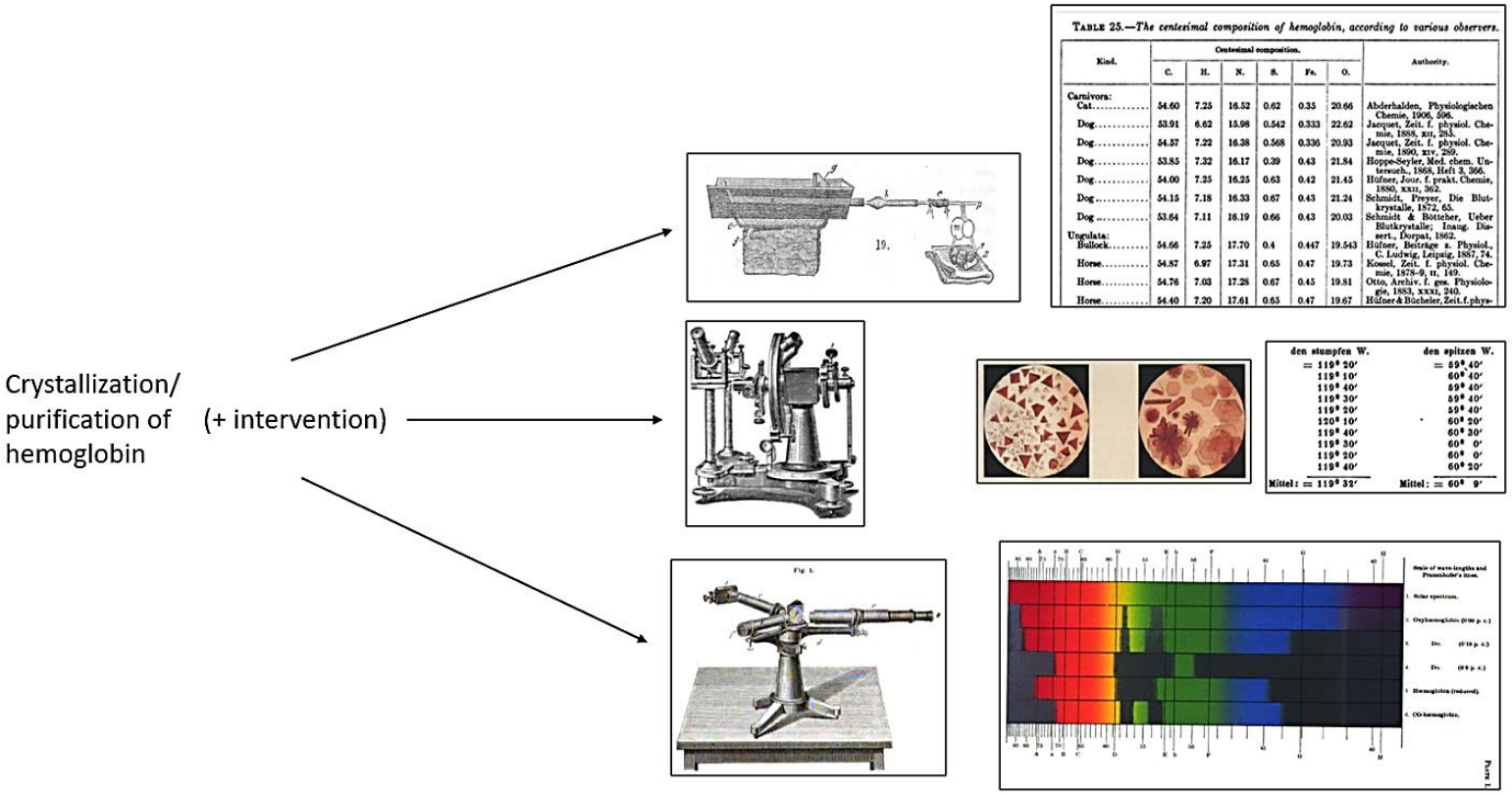


Figure 5: As a point of stability, crystals could be intervened on in different ways to yield different forms of data.

The same process can be seen in the case of X-rays. These were initially given a superficial characterization in terms of the set of empirical effects uncovered in Röntgen's experiments with shielded cathode ray tubes. Rays were subsequently explored in greater detail through a variation of experimental conditions, leading to a host of sub-classifications (soft vs. hard; primary vs. secondary; homogeneous vs. heterogeneous rays) and their associated effects. In this case, the most prominent point of stability was part

of the experimental apparatus itself: the X-ray tube, transported from one set-up to another in order to produce novel effects. As with crystallization, X-ray production was investigated with respect to second-order concerns: where the rays are emitted from; how their hardness varies over changes in time, vacuum quality, or anticathode material; what induction coil allows for the most consistent beam; and so on. In short, researchers developed a set of causal generalizations, an initial theory of technique, with respect to the niche of X-ray analysis. One sees a corresponding standardization in the diagrams of X-ray experiments over this time. As causal generalizations were affirmed, certain elements began to recur: the use of lead walls to shield sensitive parts of the apparatus, punctured lead sheets as beam collimators, dual anode tubes emitting rays at a predictable angle, and—in the UK—ionization chambers equipped with electrometers to measure the strength of an effect. Ackermann (1985) writes that "Instruments create an invariant relationship between their operations and the world, at least when we abstract from the expertise involved in their correct use" (p. 33) and this is indeed what allows the results of their use to serve as points of stability for other procedures. But his abstracting from "correct use" obscures that it is a hard-earned invariance, requiring procedural stability on the part of the experimenters and an awareness of the many conditions affecting the instruments' behavior.

What is the nature of the warrant possessed by causal generalizations of this sort? I propose to characterize these in terms of a process reliabilist theory of justification, as initially presented by Goldman (1979). According to this view, a belief is warranted when it results from a reliable belief forming process. This is a process that has a high probability of outputting correct discriminations among a set of competing claims and a low probability of outputting incorrect discriminations (in other words, it has low rates of type I and type II errors, however qualitatively expressed). Reliabilists typically assume that basic perceptual processes, occurring within some range of normal circumstances, are reliable for forming beliefs about the directly perceivable aspects of one's surroundings. While these certainly play a role in the exploratory work described above, many "observations" in science are not direct perceptions of this kind, but involve instruments that might be employed in a series of steps before producing materials that support discrimination between a set of claims. I will refrain from foundationalist questions about the most basic

belief-formation processes and instead assume that some background of warranted beliefs and reliable processes exists. I think of the relevant belief forming processes in this context along the lines of Woodward (2000a), as procedures for generating some range of data outcomes D_1, …, D_m and procedures for analyzing and interpreting these data outcomes in terms some set of competing claims P_1, …, P_n about a target phenomenon that plays a role in the data production process. A belief about a phenomenon is justified when the data production and interpretation processes that select for this claim are reliable. Note that this involves both practical and cognitive processes. The reliability of the data interpretation processes in particular will depend on background beliefs, including those that pertain to the data production process.

In the historical cases considered, reliable processes were not always on hand. They emerged from a body of empirical results reported by a research community. Researchers sought a process for examining the properties and contents of blood crystals. This first required a procedure for consistently producing samples of these crystals. Consistent procedures were not available from the start, but developed from observations of the conditions and interventions under which crystals were formed from animal blood. As scientists tracked the successful and unsuccessful outcomes, they produced a collection of statements about the coincidence of these outcomes with relevant experimental conditions. There are two basic requirements for a collection of such statements to provide material for a warranted causal generalization relating a condition C_1 to a condition C_2: (i) the observations or measurements of C_1, C_2, and all other relevant conditions are reliable, (ii) the variations in conditions have the appropriate counterfactual-supporting structure. Note that (i) requires that a practitioner have adequate background beliefs about those conditions that may be causally relevant to the relation between C_1 and C_2. As for (ii), this structure can be expressed in standard interventionist terms: there must be evidence for the counterfactual dependence of C_2 on C_1 conceived in terms of a surgical intervention capable of changing the state of C_1 alone, while cutting off the effects of all other relevant conditions on C_2.

This is a tall order in a new domain of inquiry. On the other hand, we should not underestimate a community of dogged researchers. One cannot look at a textbook on mineralogy or physiology from the era of early blood crystallization without seeing the clear priority given to recounting observations and

ordering empirical results according to classificatory schema. The drive among scientists in theory-poor terrain to collectively sift through variations, locate relevant conditions, accumulate observations, and uncover order was well-established by the time blood crystallizers came on the scene. Of course, not every set of observations reported by blood crystallizers was truly regular and not every interpretation of results was accurate; revisions to seemingly warranted beliefs were routine. Yet over time, as many actors attempted different interventions and recorded their results, trial and error led to a recurring set of experimental conditions, interventions, and outcomes—as reflected in the procedural stability that blood crystallization developed over time. To paraphrase Woodward (2000a), researchers were learning about the qualitative error characteristics of blood crystallization preparation methods. Gradually, a body of reliable knowledge began to accumulate within the exploratory empirical practices seen in early blood crystallization.

Where causal generalizations of this form can be secured, these may enable researchers to specify a reliable process for further belief formation. This is because such generalizations state dependencies between the conditions accompanying a procedure that produces data $D_1, \ldots, D_m$. The causal generalizations that describe a procedure in this way define a set of processes *Q* belonging to a general type—i.e., all such procedures carried out according to the steps enumerated, within the range of conditions for which these generalizations hold. If a sufficient number of conditions are accounted for in these generalizations, they enable a practitioner to predict the effects of interventions that occurs within this restricted range of conditions. Where this leads to consistent outcomes understood in terms of these causal generalizations, they specify a reliable process. This is the general process by which I conceive of researchers initially carving out a "causal niche" for reliable data-gathering.

Knowledge of these causal relations was the basis for ascribing properties to blood crystals. The initial characterization of a target phenomenon is given in terms of its role in the causal structure of a data-gathering process. I call this a *superficial* characterization, in the sense that it does not depend on ascriptions of substantial internal structure to a target. Blood crystallizers could adjust and evaluate their techniques without any knowledge of the structure of haemoglobin. Early X-ray scientists could carry on with their

studies without proposing one view or another about the "nature" of X-rays. A superficial characterization is based on the empirical qualities of the target and its causal relations to other factors in the data-gathering setup. For blood crystals, this initially included their solubility, their coagulability, and their responses to reagents. For X-rays, this initially included the study of how far they penetrated, how much gas they ionized how quickly, and what materials they would scatter from and in which directions. Here competing claims about the phenomenon were not model-laden; they concerned these empirically characterized behaviors. As more results were tied together using these points of stability, they served as constraints and incentives for the development of more fine-grained characterizations of their targets, including representational models.

5.2. Scaffolding and generative entrenchment

Blood crystals functioned as a point of stability, by which crystallization methods were incorporated into more extended investigative procedures (Fig. 6).

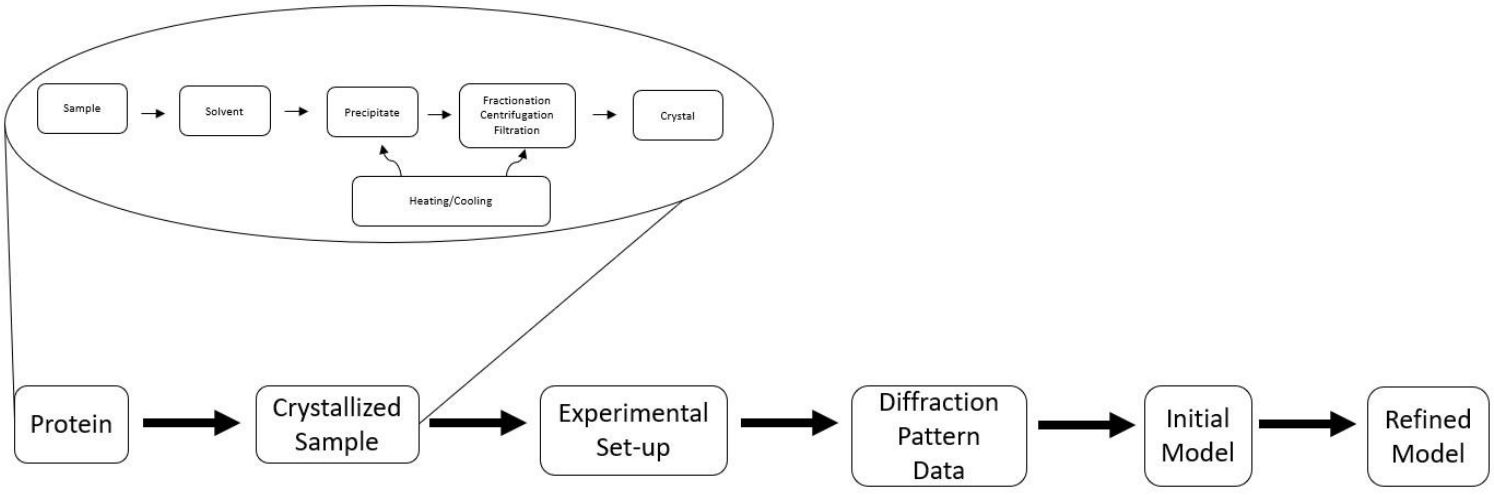


Figure 6: Incorporation of one procedure into another.

A variety of instruments were used to intervene on blood crystals. The findings that resulted from these investigations served two major purposes: they expanded the range of empirical results used to characterize

the properties and contents of blood crystals, and they provided further material for second-order knowledge of the procedures themselves. To take one example, spectroscopic analysis provided a new way to characterize differences in haemoglobin structure by means of absorption bands *and* provided evidence that certain methods of preparation—such as Zinoffsky's method of repeated dissolution and recrystallization—altered haemoglobin structure in ways that were previously undetectable. Characterizations of the target of inquiry were expanded, producing further distinctions in its properties and behaviors (Fig. 7). These findings then fed back into the theory of technique, revealing new causal factors that could provide refined guidance for practitioners and expand their technical palette.

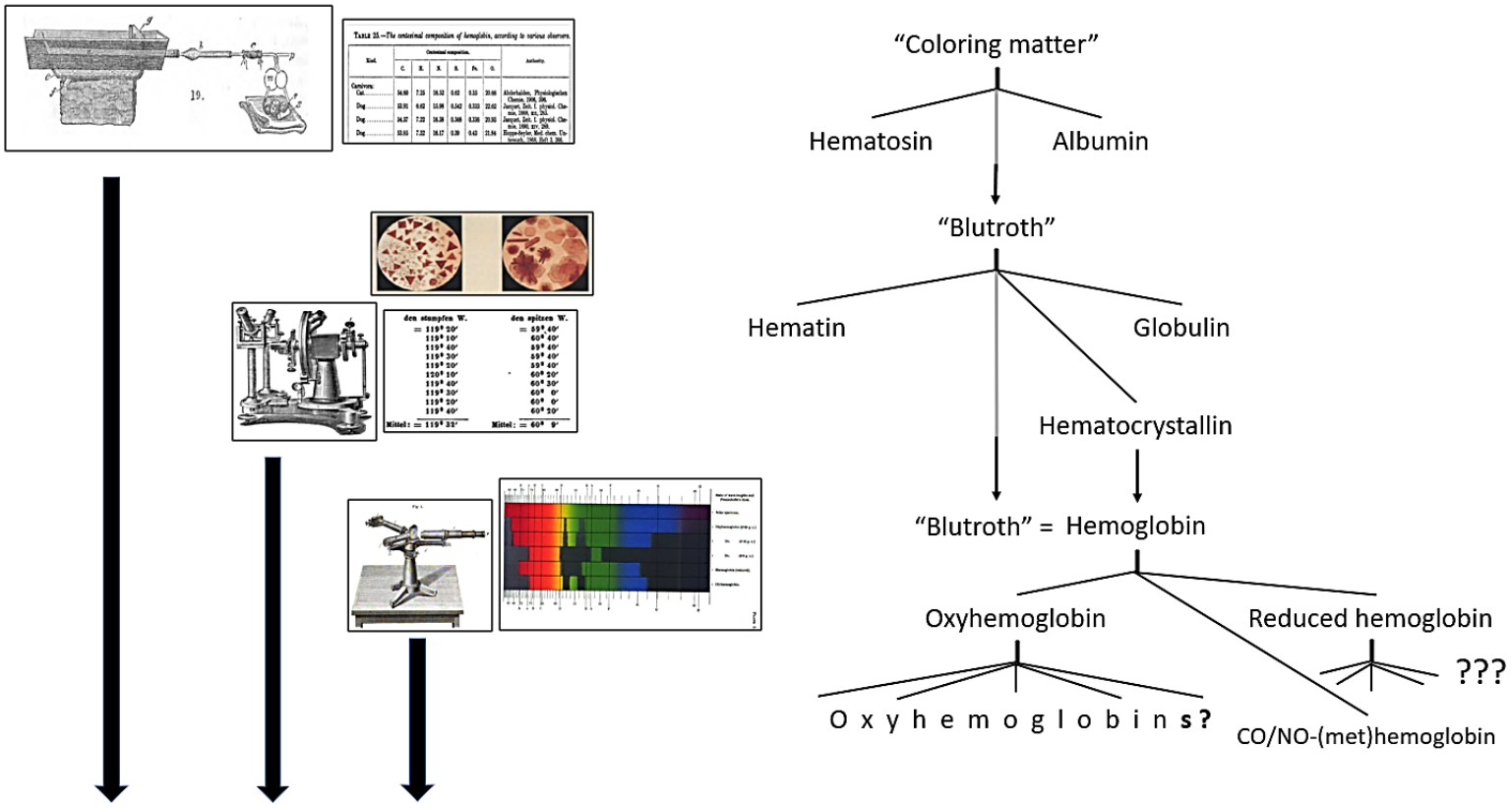


Figure 7: Refinement of the characterization of the coloring matter of blood alongside technical developments.

A similar process is seen in the case of X-ray research. Take, for example, two early characteristic behaviors of X-rays: that they scatter off some materials and that they ionize gases. These were initially empirical results demonstrated by Röntgen's laboratory experiments. Yet, by the time of Barkla and WHB's work one decade later, these results were being directly employed in the instrumentation and design of

further experiments. For example, WHB used ionization chambers to measure the forward and backward scattering of X-rays. This scaffolding process is commonplace in the developments discussed in Chapter 4: as one set of causal generalizations becomes stabilized and a corresponding characteristic effect becomes reliably producible, it serves as a point from which further effects may be investigated. A particularly dramatic instance is seen in the Braggs' reversal from using crystals to probe the nature of X-rays (thus establishing the transverse wave model) to using X-rays (now accepted as transverse waves) to investigate the finer structure of crystals.

In some cases, reports of novel empirical results prompt a revised interpretation of old data, instead of (or in addition to) a modification in data production procedures. Barkla's study on the characteristic X-rays emitted from radiated materials called into question a range of prior results from researchers that had not taken this into account as a causal factor in their experimental setup. Neglect of this factor could explain the poor quality of many X-ray photographs in the early 20^{th} century. This revelation prompted the use of additional shielding measures to improve the quality of data. Sometimes interpretive revisions are inspired by novel characterization of a target produced through other techniques. Recall how Perutz scuttled his parallel chains model of haemoglobin upon reading Pauling and Corey's paper on helix and sheet structures in proteins. The process here is both cumulative and reflexive, where points of stability serve as scaffolds both for further investigation of a target phenomenon *and* for probing methods and assumptions involved in data production and interpretation procedures.

Once a stable outcome is available, as with the routine production of blood crystals of sufficient size and purity, it can be subjected to alternative interventions. These interventions correspond to different processes p_1 and p_2 of general types Q_1 and Q_2 described by a different set of causal generalizations (say, chemical or spectroscopic analyses). Assuming these procedures and the interpretation of their results have been used as reliable belief formation processes well before their current application, and assuming there is nothing about the new target being intervened on that renders it causally distinct from prior subject matter, the token processes p_1 and p_2 are uncontroversial instances of Q_1 and Q_2. The data D_1, …, D_m produced by these interventions may be interpreted like other data and used to discriminate among competing claims

about crystal contents P_1, P_2, …, P_n. Hence, data from the chemical analysis of blood crystals was used to support the claim that blood crystal contents are protein-like, consisting of a similar proportion of C, H, N, O, P, and S. Likewise, data from the spectroscopic analysis of crystals supported the claim that they contain the coloring matter of blood, since they show the same spectrum as a liquid preparation of *Blutroth*. Researchers incorporated the crystals, as points of stability, into larger procedures whose reliability was already established. Comparing the results of these interventions with other reliable results produced by these methods allowed researchers to make new determinations about their target phenomenon.

Intervening with a reliable process on an outcome produced under the controlled variations of a prior procedure can also warrant new causal generalizations about this prior procedure. Consider, for example, repeated microscopic observation of crystals that are washed with alcohol before the drying step of crystallization and those that are not. These data may support the claim that washing crystals with alcohol before drying damages their quality. In this way, downstream intervention on the stable outcome of a (crystallization) procedure, coupled with an awareness of relevant causal factors of this procedure, may allow researchers to discriminate between causal generalizations about this procedure. These generalizations can then be applied in the future, further constraining the causal niche of a data-gathering practice in a way that contributes to its reliability.

These patterns in the development of empirical programs may be explained in terms of generative entrenchment. Generative entrenchment occurs when an element in a system functions as a building block or prerequisite for developmentally downstream elements, many of which thereby possess an asymmetrical reliance on this more foundational element (Schank & Wimsatt, 1986). Entrenchment in empirical programs is seen in the way that certain properties or effects, such as the ionizing power of X-rays, become reproducible through stable procedures, become canonical features in characterizations of target phenomena, and are then used as scaffolds to produce novel results. Entrenched features are distinguished not only by their temporal occurrence within a developmental system but also by the structural reliance of some elements of the system on others. This may be reflected, for one, in the resistance of some elements to change relative to others that they facilitate. It may also be seen in a tendency among practitioners to

compensate for the deleterious effects of more entrenched elements rather than simply doing away with them.

This helps explain the continuity of blood crystallization methods beneath significant changes in the techniques used to intervene on crystals. Crystallization was the first stage of any such procedure. In fact, comparisons of the efficacy of downstream techniques typically relied on crystals produced in the same way. It was paramount, then, that the outcomes of crystallization procedures were consistent over time. Similarly, the wave continuum model of X-rays was a foundational assumption in X-ray crystallography—the source of Bragg's Law of reflection—and was never seriously challenged from within this practice. Novel derivations of this law were presented throughout its history, but only to reaffirm the basic assumptions initiated by Bragg. A final example: once crystal rotation methods became the standard approach for producing diffraction patterns in protein X-ray crystallography, the program rarely shifted from this method of data production. Instrumentation was improved in minor ways—increasing the operating power of X-ray tubes was a great aid to data production—but on the whole practitioners shifted their concentration to downstream stages of the procedure having to do with data processing. Thus, a panoply of new techniques for quantifying intensities, calculating Fourier series, and interpreting contour maps arose while rotation methods remained relatively stable.

5.3. Reciprocity of target and technique

According to Wimsatt (2013), generative entrenchment is commonly found in systems with "recurrent iterative cycles of processes" (p. 309). Many such cycles were seen in prior examples, in which developments in target characterization and experimental technique inform one another. This is part of a basic dynamic in the development of empirical programs: *an ongoing reciprocal influence between refined*

characterizations of the target phenomenon enabled by the production of novel results, and refinements of experimental practice by inclusion of novel target characterizations into a theory of technique (Fig. 8).[143]

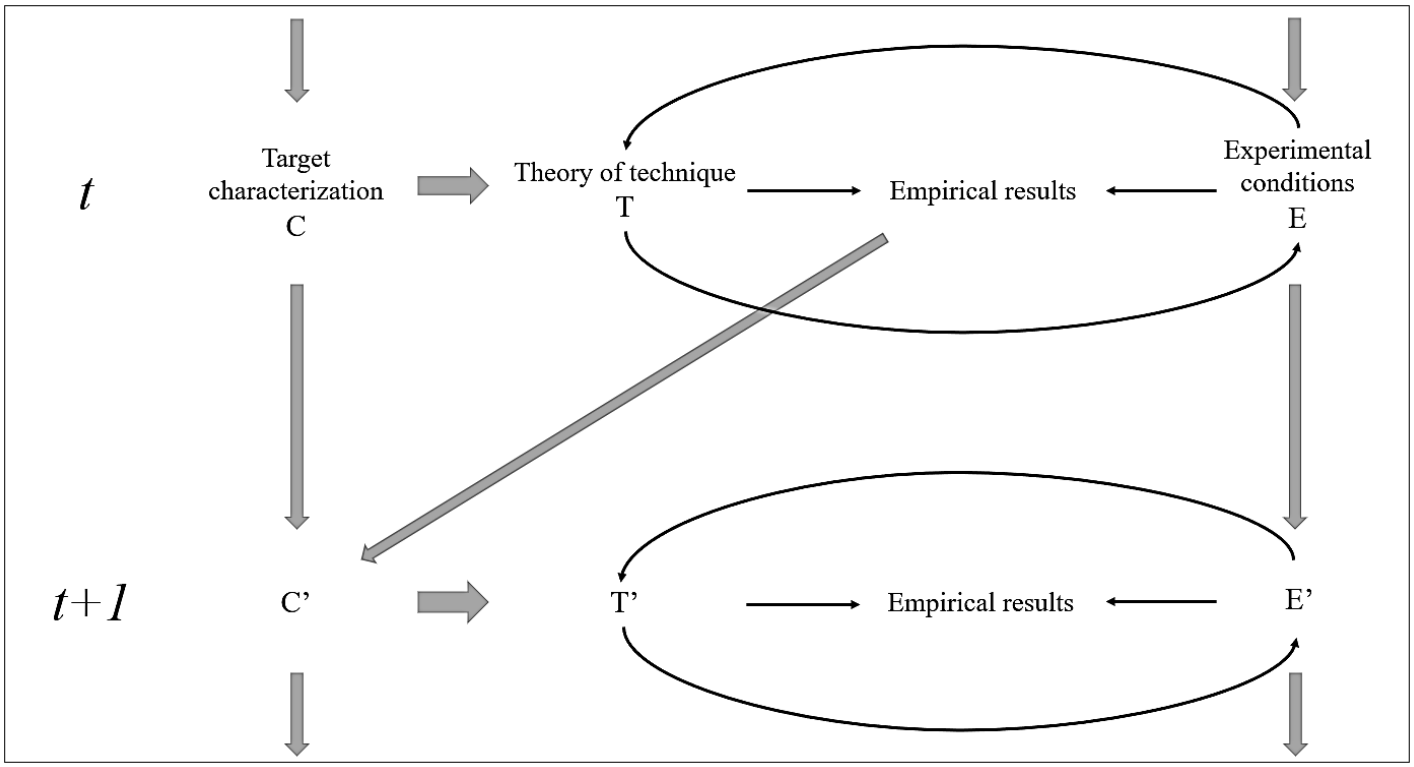


Figure 8: A graph depicting reciprocal influences in the development of experimental conditions and target characterization within an empirical program. The thick horizontal arrows stand for the inclusion of target characterizations into a theory of technique. The thick vertical arrows stand for the inheritance of features from prior target characterizations T and experimental set-ups E into later iterations. The diagonal arrow signifies the contribution of interpreted data on novel target characterizations. The arrows leading from T and E to empirical results for the claim that these are the product of a procedure that joins experimental conditions to a theory of technique. Finally, the arcing arrows from T to E stand for the guiding role that a theory of technique plays in the design and execution of a procedure, and the arcing arrows from E to T for the need to account for causal factors in an experimental setup within a theory of technique.

Investigations into crystal structure in the early 20th century exemplify the process of mutual refinement by reciprocal influence. Figure 9 presents a summary overview of some key developments discussed in Chapter 4, Sections 3.4 and 3.5. The left three columns present aspects of the theory of technique, with the leftmost concerned solely with target characterization. The second column to the right

[143] Rheinberger (2010) writes of this dynamic as "a broadly conceived epistemological question about the reciprocal relation between epistemic things and the technical conditions of their manipulation in experimental systems. The productivity of such systems depends on that relation. Technical things set the boundary conditions of experimental systems and in the process create the space in which an epistemic object can unfold. Scientific instruments as well—this is my general contention—acquire their epistemic meaning only against the backdrop of particular experimental systems or as a result of the fact that they have been tailored to determinate experimental systems" (p. 218).

lists a selection of significant empirical results and the rightmost column lists novel methods introduced over this time. Among other things, this shows a sharpening of the claim that diffraction intensity is based on atom positions to the claim that intensity is a Fourier sum of the electron density distribution surrounding Bohr-like atoms. This sharpening occurs alongside growth in the formula for total reflected intensity, presented in terms of an increasing number of correction factors, and a series of empirical results that confirm or motivate further characterizations. At the same time, further elements—such as atomic radii and bond directions—were introduced to representations of the crystal lattice (Fig. 9, far left).

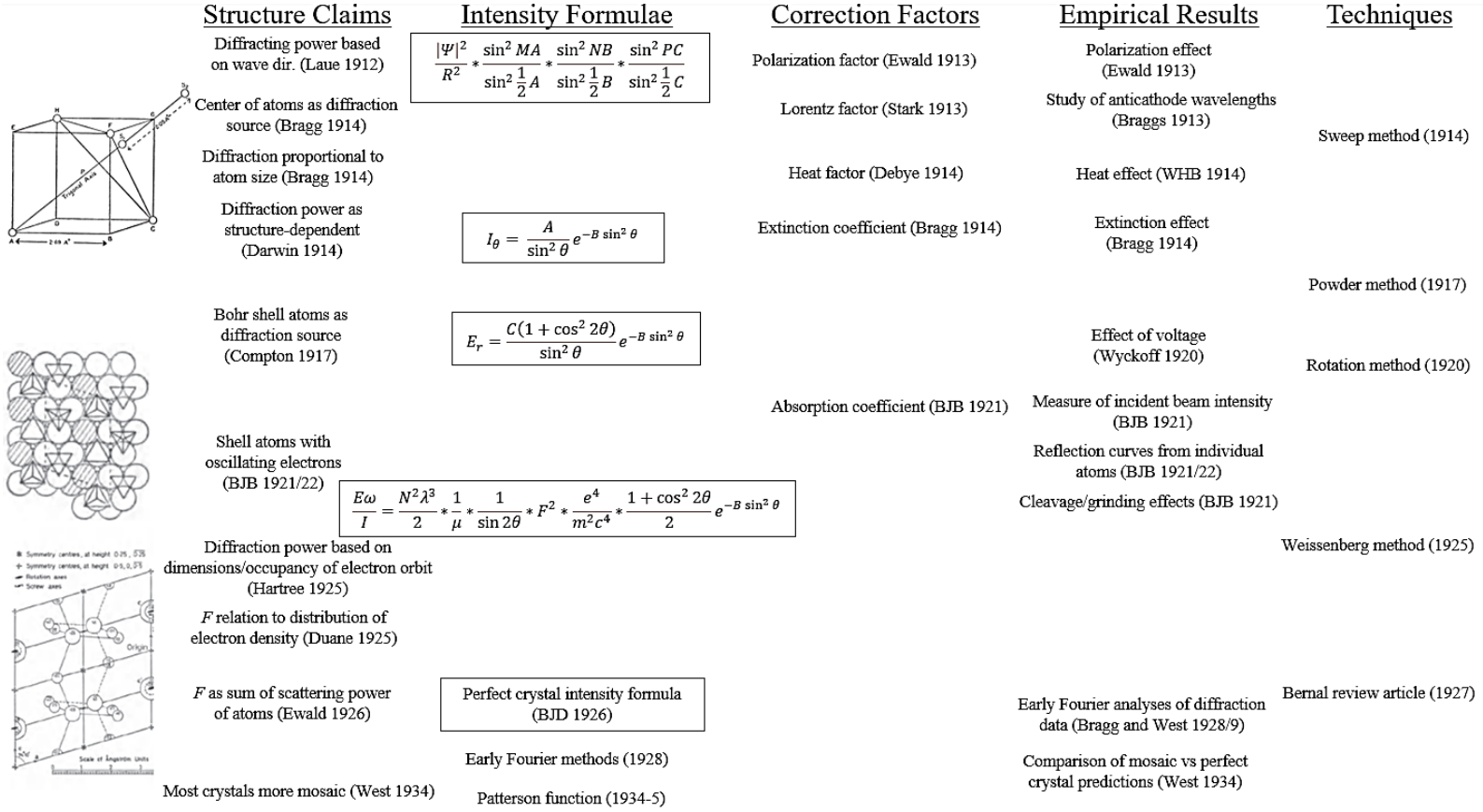


Figure 9: Aspects of the development of the X-ray crystallography program, 1912-1935.

While a complete accounting of the relations of influence between these different aspects of the program would produce a dense web, the interplay between target characterization and the production of empirical results can be traced. For example, Bragg claimed the correspondence between his cubic model of zincblende and Laue's diffraction results was evidenced for the notion that atom centers were the source

from which X-ray waves were diffracted. This became the basis for Debye's calculation of the effect of heat on intensity readings, derived from a model of independently vibrating atoms. In response to Debye, WHB designed an experiment to measure this effect and found agreement with his heat factor. WHB added this factor to his empirical formula for the reflected intensity (second formula box). Compton compared this formula to one predicted from electromagnetic theory to argue that the diffraction source in crystals was likely not atomic centers, but electrons arranged in a shell around them. This claim motivated further empirical investigation by Bragg, James, and Bosanquet, whose measurements of the scattering curves from rocksalt atoms was taken as further evidence for a kind of Bohr atom as the diffracting source within crystals (Fig. 10).

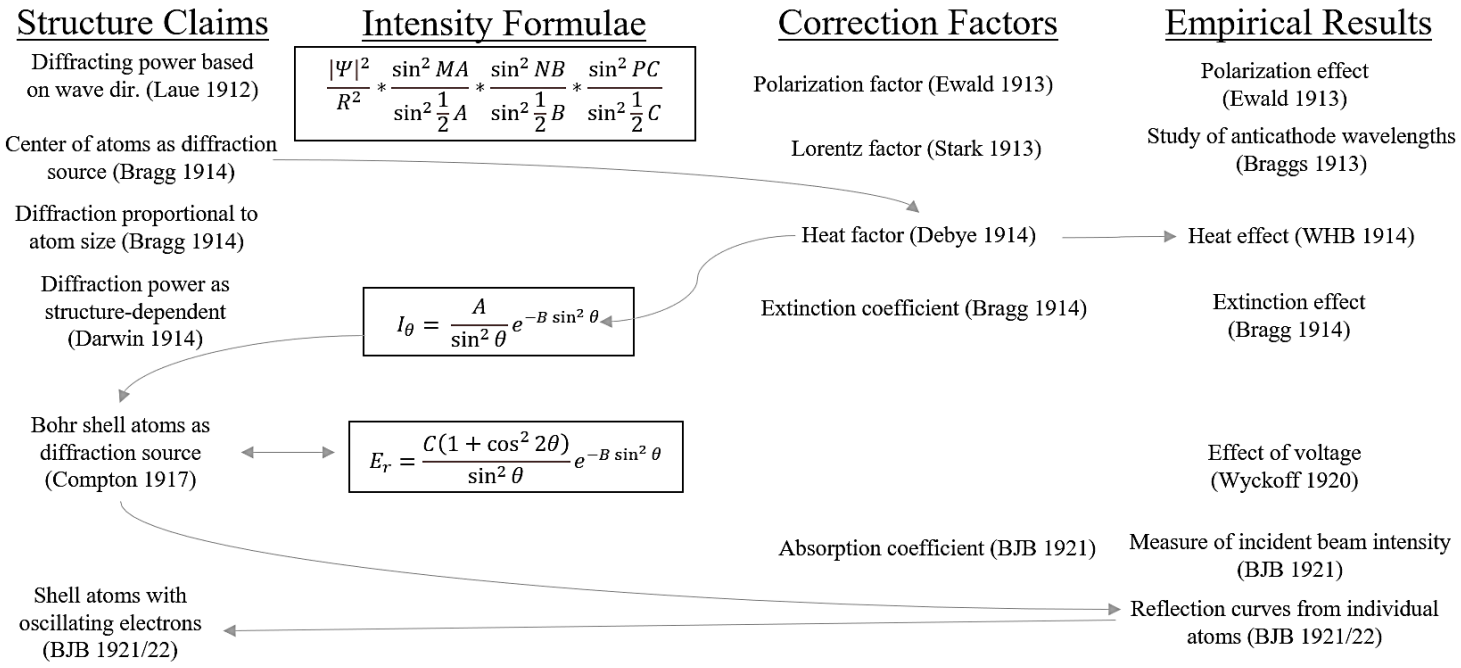


Figure 10: Lines of reciprocal influence between characterization of crystal structure and empirical results.

As crystallographers sought to understand the structure of more complex molecules, they arrived at a standard protocol that could be pursued through a range of investigative strategies (as reviewed by Bernal, Cf. Ch. 4, Sec. 3.4) and subjected recorded data to increasing layers of interpretation. One of the clearest progressions in this period is seen in the addition of correction factors to the formula for total reflected intensity. Changes in the formula track the emergence of a detailed understanding of the crystal-

ray apparatus interaction, each new term accounting for a further factor impacting intensity readings (Fig. 11).

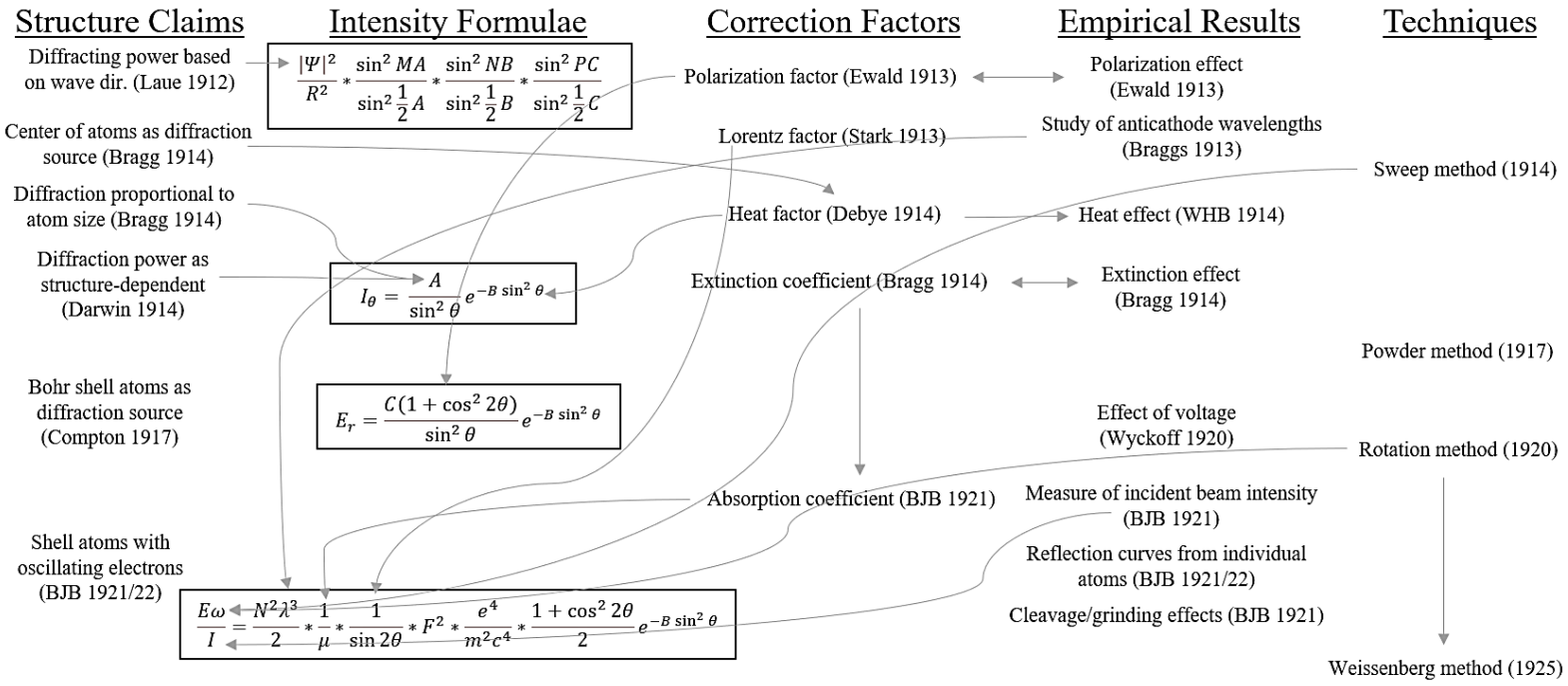


Figure 11: Elements contributing to development of the intensity formula.

By the 1920s, inferring a structure-dependent quantity (variously *A*, *C*, and *F*) from intensity readings called for a series of parallel measurements and estimates, in which initial recordings were subsequently modified with respect to measures of incident beam intensity, unit cell volume, anticathode wavelength, crystal angular velocity, the estimated number of atoms per unit cell, and so on. In this process, the initial theoretical principles of X-ray crystallography—the wave model and Bragg's Law—were gradually elaborated into a rich mathematical theory of technique combining elements of wave kinematics, crystal geometry, and atomic hypotheses inspired by the new quantum mechanics. With each correction factor, researchers eliminated from intensity readings additional sources of error in the determination of *F*. Conversely, as models of crystal structure became more sophisticated (most notably, as the role of atomic form factors was recognized), theoretical predictions of *F* from crystal structure gained in accuracy. Using intensity formulae and auxiliary measurements, crystallographers derived values of *F*. Using the

mathematics of waves and crystal lattice theory, they predicted values of *F* from possible molecular models (Fig. 12).

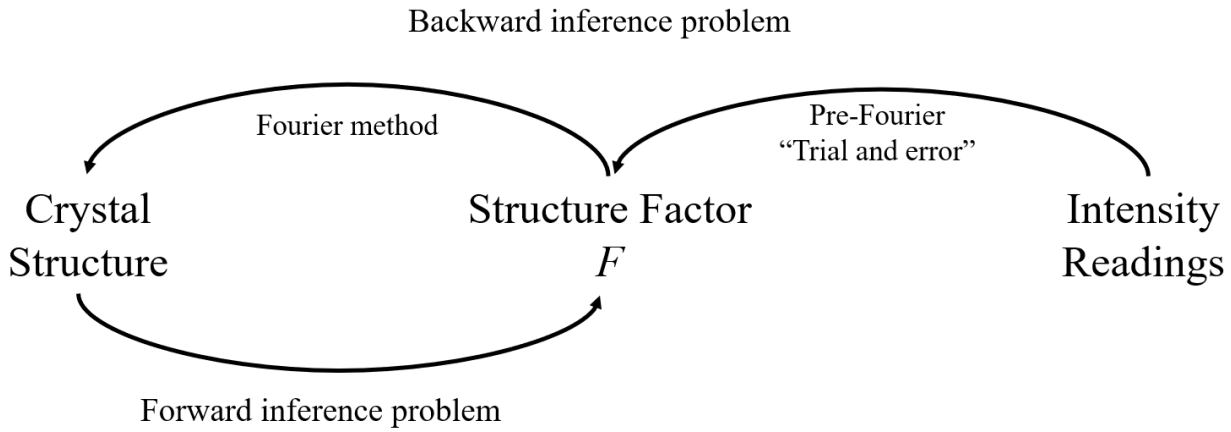


Figure 12: Forward and backward inference, one way that reciprocal influence was formalized in X-ray crystallography.

The reliability of X-ray data production and interpretation—the ability to discriminate between crystal structures using these data—was thus improved from two sides: that of modeling and in the handling of data. Each side was used to inform the other by means of this forward and backward inferential strategy, with the structure factor *F* serving as a formal linchpin. The degree of experimental precision and mathematical sophistication allowed for a corresponding level of detail in the abstract models of crystal structure.

The production and interpretation of results may occur against the background of markedly different, competing characterizations of a target. This was the case during the rivalry between the neutral pair and ether impulse models of X-rays, the colloidal and large molecule conceptions of proteins, and later in the cyclol and polypeptide chain models of protein structure.[144] At other times, or within a narrow line of inquiry, the reciprocity between target and technique may lead to piecewise, progressive developments, as with the investigation of crystal structure. This contrast reflects, on the one hand, a difference in the

[144] Each of these rivalries ended with one characterization conquering another, or both subsumed by a third, under further scrutiny. But it is equally possible that a "splitting" of the phenomenon occurs, in which researchers come see themselves as interacting with two different targets corresponding to alternate experimental conditions. Longino (2013), for example, gives an account of how this occurs with different operationalizations of human behavioral traits, such as aggression or sexuality.

grain of claims about phenomena P_1, …, P_n that are relevant to researchers. For a researcher who believes there is insufficient evidence to discriminate between the claims that proteins are colloids or large molecules, claims about whether proteins have molecular weights on the order of 10^4 or 10^5 atoms are too fine-grained to be relevant. On the other hand, improvements in techniques, such as the introduction of densitometers for quantitative readings of intensity data, can increase the precision of data production and interpretation. Novel means of generating data from a point of stability, as with the introduction of spectroscopic analysis, make new sets of data D_1, …, D_m available to researchers. In both cases, this enables them to eliminate a wider range of coarse-grained claims about a target and expand the range of fine-grained claims that are relevant to inquiry. As novel forms of intervention are developed, or more factors affecting data production incorporated into a theory of technique, data thus become more reliable in the sense that they allow for sharper discrimination between claims about phenomena.[145] Chapter 6, Section 2.1 and Appendix II discuss additional effects of this mutual refinement of target and technique.

5.4. Historical justification and alterations to empirical programs

5.4.1. Inter-practice transitions

The development of an empirical program is a non-linear process, marked by self-scrutiny and procedural revision. As a result, it is not purely cumulative: Prior methods may be called into doubt, old results may be stricken from the record, and once-common nomenclature buried and forgotten. On what grounds, then, is a practitioner warranted in adopting an empirical program, especially if judgments of its reliability may change over time? What kind of historical process might justify the adoption of a program that has gone through a sequence of these alterations?

[145] As show in Figure 7, for example, spectroscopy data allowed physiological chemists to discard the claim that *Blutroth* was distinct from haematocrystallin and to reliably distinguish between its different forms, such as oxyhaemoglobin and reduced haemoglobin. Recognition of this difference then informed methods for collecting further data, such as attempts to crystallize pure samples of one or the other form of haemoglobin.

Part of the answer requires that we simply acknowledge that scientific warrant functions *in media res* and is necessarily defeasible. There is no ultimate guarantee that the knowledge guiding a contemporary method will not be thrown away in the future. Even the processes leading to low-level empirical claims, including basic perceptions, may have their reliability re-evaluated in light of newer findings. On the other hand, just as the elements of a procedure accumulate and become incorporated in a piecemeal fashion, they are often overturned in similarly gradual processes, reflecting the relative entrenchment of different techniques and theoretical principles. I remarked previously on the significant continuity of instrumentation and techniques among organic chemists and mineralogists throughout the nineteenth century. Though the goals and theoretical orientations shifted, sometimes in ways that made previous research look quixotic,[146] it is hard to find momentous overhauls at the level of data-gathering practice. As organic chemistry shifted from radical theory to the theory of substitutions to valence theory, bedrock categories, such as those referring to the principal constituents of organic molecules, were never abandoned. That such elements persist through theoretical change provides reason to treat typical alterations to an empirical program as a sequence of alterations to an enduring system of practice.

While some elements persist through significant changes, those alterations that do take place occur against the backdrop of a specific set of scientific goals. Some alterations may be understood as instances of what Kitcher (1985) calls rational inter-practice transitions. These are transitions between sets of recognized problems or questions, accepted statements, forms of reasoning, and methodological standards stated in a common scientific vocabulary.[147] Two simple forms of inter-practice transitions are question-generation, in which members are added to the set of relevant questions, and question-answering, in which some members of this set are removed and new accepted statements or forms of reasoning are added.

[146] Fernet's (1858) treatise on blood finds him injecting milk directly into animal veins and observing the effects.
[147] Kitcher's focus is on the growth of mathematical knowledge and terminology differs from my account of empirical programs, but can be assimilated to it. That is, the vocabulary could be thought to consist of the content of partial, local theories employed in a theory of technique. The statements would be divided into those pertaining to the experimental target, the apparatus, and their interaction. The reasonings would pertain to the licensed inferences within the theory, along with the in situ strategies it makes available. Together, in conjunction with overarching research goals, these would inform the methodological views adopted by practitioners.

Question-generation may be induced by changes to the set of accepted statements or by extensions and shifts to scientific vocabulary. For example, acceptance of the claim that haemoglobin from different animals consistently crystallized into different forms gave rise (via accepted reasoning at the time connecting crystal form to chemical form) to new questions about the chemical differences between the haemoglobin of different animals—an instance of question-generation. Being able to pose this question depended on 'haemoglobin' having entered chemists' vocabulary. This was due to a prior instance of question-answering by Hoppe with respect to the relation between the constituents of blood crystals and the coloring matter *Blutroth*. Sequences such as these, involving the back-and-forth raising and answering of questions, are readily reconstructed from the history of an empirical program.

For Kitcher, a belief is ultimately justified if it issues from a practice that results from a grounded sequence of rational interpractice transitions. To be properly grounded, this sequence must originate from a directly warranted set of beliefs (such as those that are perceptually warranted). However, Kitcher does not explain in any detail how more complex elements of a practice, such as methodological standards and forms of reasoning, can "result from" perceptual beliefs through rational transitions. For my purposes, it is enough to show that, historically, the reliability of an empirical program tends to improve over time through its alterations. This is because alterations tend to lead to increases in the resolution of data and improvements in researchers' ability to distinguish among more fine-grained claims about phenomena, as seen in the processes of mutual refinement presented in Section 4. If a practice PR_1 warrants a certain set of beliefs, then this warrant extends to beliefs formed from its descendent PR_2, provided the sequence between PR_1 and PR_2 consists of rational alterations that improve the reliability of the practice as a whole. Note that, while this process transfers warrant to the beliefs that issue from each practice, this does not mean the same beliefs are warranted. A transition from one practice to another may involve significant changes to beliefs about a target phenomenon, to accepted lines of reasoning, or to assessments of a technique's usefulness for answering pertinent questions.

What motivates the significant modification of a method or a choice between techniques, from the standpoint of practitioners? Scientists evaluate methods in terms of their ability to satisfy specific scientific

goals. The claim that one technique is preferable to another typically involves appeal to procedural or epistemic norms that reflect these goals. Certain norms are stated at a very general level and are found across far-flung programs. Liebig's (1839) *Instructions* and Bernal's (1927) review of X-ray crystallography methods appeal to a common set of standards for data production. Their preferred techniques generally produce more data in less time, are applicable to a wider range of targets, involve instruments that are simpler to understand, and are easier to execute.[148] Both authors also speak of a more fundamental standard: the accuracy of their data. Judging by the texts, this is a reference to how well the data supports claims about a target of inquiry. In other words, they want a reliable method.

A demand for reliability underlies the investigation of second-order concerns and is a significant driver in the development of empirical programs. But the appeal to reliable or "good data" over time raises questions about the historical character of such judgments. A technique's reliability is determined by how well one can discriminate between competing claims about phenomena on its basis. Thus, judgments of reliability are made with implicit reference to a competing set of claims about a target phenomenon. From a practitioner's standpoint, these are indexed to the prevailing characterizations of a target phenomenon that are available at a point in time. Attending to the development of empirical programs from this perspective can shed light on how judgments of reliability change over time in ways that motivate rational alterations to empirical programs.

As an empirical program develops, it undergoes routine alterations in its motivating questions or problems. New questions are posed atop the old; target characterizations are solidified and spawn more refined problems; the domain of targets under a program's purview is expanded or contracted. As new questions arise, the reliability of previously trusted techniques may become confined to an increasingly limited range of claims. This occurred in the first decades of X-ray crystallography, where the earliest methodological advances were achieved by the Braggs, Moseley, and Darwin's use of ionization spectrometry. This technique enabled investigation into the characteristics of atoms and the detailed

[148] As Bernal lays out (Ch. 4, Fig. 37) and as was discussed with reference to in situ strategies (Ch. 3, Sec. 4.2), one often has to accept trade-offs between these qualities depending on the target of interest and contingencies at hand.

structure of simple, highly regular minerals. Yet as crystallographic ambitions grew to larger, more complex molecules, it became apparent that spectrometry was more time-consuming and more liable to basic errors than photographic techniques. Accordingly, photographic techniques became the preferred means of studying most new crystals, and ionization spectrometry was increasingly limited to a small sub-set of problems.

This is an instance where the methodological judgment takes place in response to developments that are *internal* to an empirical program. By this I mean the motivating questions are based on prior target characterizations developed within that program. Such questions are often quite detailed, concerning further refinements of a target; the program that spawned these questions is a strong candidate for answering them. Questions like "what empirical qualities distinguish oxyhaemoglobin from reduced haemoglobin?" or "are X-rays diffracted directly from the centers of atoms or from points distant from their centers?" became salient in the course of development of specific techniques. In the case of ionization spectrometry used by Bragg, Mosely, and Darwin, the reliability of the technique was re-evaluated due to later alterations in the empirical program. Cycles of scientific problem-solving had given rise to newly relevant claims about target phenomena and new means for investigating them. The shift in preferred methods was in direct response to these alterations, and so was rationally consistent with them. Indeed, it would have been irrational for a practitioner to hold onto the same technique despite its increasingly recognizable shortcomings.

In some cases, judgments of reliability shift when a technique is transferred from one program to another. As Bernal and Crowfoot observed (Ch. 4, Sec. 4.2), crystallization techniques that were perfectly reliable for biochemical study were inadequate for good X-ray data. Here the shift in judgment resulted not from refinements in the questions of target characterizations internal to a given program, but from a technical procedure being incorporated into a novel program. On incorporation into the X-ray diffraction procedure, methods of crystal preparation were being employed to answer an entirely different set of questions than those for which they were initially developed. This called for revision to previously accepted methods, resulting in Crowfoot's "wet crystal" preparation.

In still other cases, scientists or other thinkers may consider *external* questions, raised independently from central lines of inquiry within a program and not clearly answerable by its means. Haüy had speculated about the "structure" of the constituents of crystals in the late 19th century. Subsequent crystallographers from Lehmann (1851) to Reichert and Brown (1909), gestured at the notion of protein structure or "internal constitution," but acknowledged there was no crystallographic data to answer such questions at any grain. The assumed connection between crystal form and chemical structure kept hopes aloft for decades, yet it ultimately became clear that answering any questions on this matter required a markedly different way of producing data. Protein crystal morphology became an increasingly recondite area of study in the early 20th century, until it was finally subsumed under X-ray crystallography.

These cases illustrate the relationship between reliability, target characterization, and methodological change within or between empirical programs. A technique's reliability, that is, its ability to produce "good data" depends on the questions about phenomena that its data and accompanying interpretations allow one to discriminate. These questions are often posed in terms that reflect internal developments of a specific program, drawing on more fine-grained target characterizations prompted by the production of empirical results. Evaluations of a technique's reliability will vary depending on its aptitude for answering a given question within a given program at a given stage of development. Judgments of reliability made from the standpoint of practitioners—judgments whether a technique yields "good data"—are thus contextual and variable over time. This does not make them less rational or well-formed; they are simply sensitive to the facts on the ground. Methodological alterations to an empirical program, such as the decisions to switch between techniques, revise a previously accepted technique, or give up on a program entirely are justified by reliability judgments of this sort.

Finally, how do shifts in reliability judgments cohere with the overall increase in reliability that I attribute to empirical programs as they develop over time? This can be resolved by distinguishing between judgments of reliability from practitioners' standpoints and from the historian's standpoint. The practitioner judges the reliability of techniques with respect to their ability to discriminate between claims about a target phenomenon, where these are based on presently available means of characterizing that target. The historian

judges the change in reliability of an empirical program as a whole by comparing the ability to discriminate between claims about a target phenomenon at two (or more) points in time. The relevant metric here is not how many competing claims can be eliminated at a given point in time. Instead the focus is on how many claims become eliminable over time as reflected, e.g., in the resolution of data or the degree of refinement in target characterizations.[149]

There is a distinct risk that the historian's judgment takes on a Whiggish quality. This may happen when current practices are treated as the standard for judging the past, when all development in an empirical program is viewed as a march toward the highly reliable practices of today. I have largely focused on episodes that support my claim that reliability improved over time and have identified patterns exhibiting this trend. But there is also evidence of uneven development within these programs, marked by blind alleys and mistaken confidence. With that in mind, I offer a few examples that lack this triumphal tenor:

- Despite various alterations to protocol, it is not clear that blood crystallization procedures were markedly improved after the wider reception of Hoppe's methods c. 1870s. In general, the understanding of crystallization advanced very slowly—if at all—until the 1930s, when novel techniques such as vapor diffusion began to be developed with more theoretical rigor.
- While many elements of X-ray experimentation flourished from 1895-1912, only a minor strain of this work could help answer fundamental questions about the physical nature of X-rays. Thus, the review of X-ray nature from Kassabian's (1910) textbook largely consists of quotes from Röntgen's early papers. This issue was especially pronounced in the photographic tradition in continental Europe (as opposed to the ionization chambers popular with Barkla and WHB), which were of poor quality and difficult to interpret. Hence the debate over Haga and Wind's purported photographic evidence of diffraction from 1899 carried on for over a decade until Laue, Friedrich, and Knipping's results.

[149] These are terms of art. A practitioner can take up the historian's perspective as, for example, when they review the state of the field as it has developed up to their present day (though here they risk a Whig interpretation).

- Despite various developments in protein biochemistry, cleavage and synthetic techniques did not advance much between Fischer's searching efforts in the early 20th century and the development of the Bergmann-Zervas carbobenzoxy method in 1932. Protein bonds and sequencing were poorly understood throughout this period. Even with diffusion techniques and the introduction of X-ray analysis, many fundamental questions about protein structure of interest to biochemists (e.g., whether they were large molecules or aggregates) remained undecided until the turn of the decade.
- Reliable protein X-ray crystallography was more fantasy than reality in the 1930s and 40s, as many of Perutz's colleagues thought at the time. By Bragg's later assessment, the first solid quantitative result was not achieved until 1952. This was preceded by a period of overly optimistic, hunch-driven interpretations of X-ray data, exemplified by Perutz's parallel chain model.

Stagnation may persist in the face of problems that gained primacy within a research community. As mentioned several times in this chapter, this was the case with early efforts to infer protein structure using 19th century crystallographic techniques. Here practitioners became increasingly aware that the question of structure was external to this program. In such moments, practitioners may lose interest in the focused technical problems within a program and move toward judging its reliability as a whole. The slump in work on the crystallization of proteins in the early 20th century thus resulted from the conclusion that these methods were generally unreliable for solving the reigning problems at this time.

5.5. A note on the identity of empirical programs

Given these methodological alterations, on what grounds can I claim that these changes are being made to the "same" empirical program? In Chapter 3, I claimed empirical programs were individuated by a central data-generating object-apparatus interaction. I distinguished empirical programs from specific

protocols, appealing to a higher level of generality in discussion of their underlying procedures. An example of a procedure described at this level of generality is provided by Bernal's late-20s review of X-ray crystallography:

> (6) Measurement of spacings of planes of known indices leading to the determination of the size and shape of the unit cell.
> (7) Measurement of crystal density from which with (1) the number of molecules per cell can be found.
> (8) Determination of the symmetry class by the methods of ordinary crystallography.
> (9) Determination of the indices of absent reflections leading with (3) to the determination of the space group; and this combined with (2) giving the molecular symmetry.
> (10) Measurement of the intensities of the reflections from planes of known indices, and of the correcting factors necessary to obtain the structure factors, from which, together with (4), the complete structure may be derived (Bernal, 1927, pp. 275-276).

Bernal goes on to discuss a variety of different techniques, all of which are compatible with this general line of attack. Over the preceding fifteen years, practitioners had developed many ways for producing diffraction data via crystal-ray interactions. All of them maintain a family resemblance to the extent that they are instances of this general procedure and can achieve the central aims outlined by Bernal. Each individual technique also possesses a phylogenetic identity, so to speak, insofar as it originated from modifications to an earlier technique based in the same general approach to data production and interpretation. When techniques splinter off and are re-purposed for other ends, like Kendrew's program for digital computation of Fourier transforms, they become parts of other programs. Conversely, techniques developed elsewhere, like the use of densitometers, may become part of the X-ray crystallography program. Ultimately it is employment of techniques in the service of this foundational interaction—diffracting X-rays from crystals and inferring crystal properties therefrom—that maintains the identity of the program.[150]

[150] One might think this claim is further complicated by the fact that practitioners' understanding of the central interaction and what it can be used to measure also evolve in time. There are likely cases where these changed in such dramatic ways that it led to a break in the identity of an experimental practice or, likely more often, its abandonment. But I do not see this in the cases that I focus on, where development is more gradual and the program maintains its basic procedural character (due, in part, to the generative entrenchment of points of stability). But at the end of the day, I cannot provide a strict set of criteria for identifying two instances as part of the same "empirical program." It is an analytical category, meant to serve as a tool for identifying patterns of historical development and justification in the sciences, not as part of a joint-carving ontology.

5.6. Conclusions

Of course, whether a program thrives, survives, or dies is subject to all kinds of historical contingencies. Perutz was only able to pursue haemoglobin with such abandon due to factors like Bragg's connections to the Rockefeller Foundation, his refugee status during the war, and the post-war boom in government funding of "basic research." Kendrew joined the Cavendish in 1945 after a military post alongside Bernal in South Asia stoked his interest (de Chadarevian, 2018). The formation of the Laboratory of Molecular Biology would have been impossible were David Keilin, the biochemist who loaned desk space for Perutz and Kendrew's crystallizations, not on friendly terms with Sir Edward Mellanby, the head of the UK's grantmaking Medical Research Council (Perutz, 1997). And yet, through the thicket of chance encounters are woven paths of justified knowledge, forging reliable production and interpretations of data.

This chapter was meant to provide an overview of the processes by which an empirical program develops from its exploratory stages into a highly integrated, sub-divided body of procedural knowledge. I have focused on the developmental patterns that emerged from the historical case presented in Chapter 4. These were meant to highlight the interplay between data-gathering techniques and target characterizations in terms of their initial stabilization and subsequent interactions through mutual refinement. I have sketched an account of the justification of the beliefs guiding these techniques from reliable belief-forming processes. These contribute to the growth of knowledge through processes of local justification and rational alteration, spurred by mechanisms such as question-asking and -answering and practitioners' assessments of reliability. In Chapter 6, I return to matters of scientific representation, using the prior chapters to argue for an account of scientific representation as the warranted used of models as inferential tools in relation to an empirical program.

References

Ackermann, R. (1985). *Data, Instruments, and Theory.* Princeton: Princeton University Press.

Bernal, J. D. (1927). A universal X-ray photogoniometer. *Journal of Scientific Instruments, 4*, 273-284.

de Chadarevian, S. (2018). John Kendrew and myoglobin: Protein structure determination in the 1950s. *Protein Science, 27*, 1136-1143.

Fernet, E. (1858). *Du role des principaux éléments du sang dans l'absorption ou le dégangement des gaz de la respiration.* Paris: Martinet.

Goldman, A. (1979). What is Justified Belief? In G. Pappas (Ed.), *Justification and Knowledge* (pp. 1-23). Dordrecht: Reidel.

Kassabian, M. (1910). *Röntgen Rays and Electro-Therapeutics.* Philadelphia: Lippincott Company.

Keilin, D., & Hartree, E. F. (1935). The Combination between Methaemoglobin and Peroxides: Hydrogen Peroxide and Ethyl Hydroperoxide. *Procedings of the Royal Society of London, B, 117*, 1-15.

Kendrew, J., & Perutz, M. (1948). A comparative X-ray study of foetal and adult sheep haemoglobins. *Proceedings of the Royal Society of London, A, 194*, 375-398.

Kitcher, P. (1985). *The Nature of Mathematical Knowledge.* New York: Oxford University Press.

Lehmann, C. G. (1851). *Physiological Chemistry, Vol. 1.* (G. Day, Trans.) London: Harrison and Son.

Liebig, J. (1839). *Instructions for the Chemical Analysis of Organic Bodies.* Glasgow: Griffin & Company.

Longino, H. (2013). *Studying Human Behavior*. Chicago: Chicago University Press.

Perutz, M. (1997). *The Medical Research Council Laboratory of molecular biology*. Retrieved from NobelPrize.org: https://www.nobelprize.org/prizes/uncategorized/the-medical-research-council-laboratory-of-molecular-biology/

Reichert, E., & Brown, A. (1909). *The Crystallography of Hemoglobins.* Washington, DC: Carnegie Institution of Washington.

Rheinberger, H.-G. (2010). *An Epistemology of the Concrete.* Durham: Duke University Press.

Salkowski, E. (1903). *A Laboratory Manual of Physiological and Pathological Chemistry.* New York: Wiley & Sons.

Schank, J., & Wimsatt, W. (1986). Generative Entrenchment and Evolution. *Philosophy of Science, 1986*, 33-60.

Wimsatt, W. (2013). The role of generative entrenchment and robustness in the evolution of complexity. In C. Lineweaver, P. Davies, & M. Ruse (Eds.), *Complexity and the Arrow of Time* (pp. 308-331). New York: Cambridge University Press.

Woodward, J. (2000a). Data, Phenomena, and Reliability. *Philosophy of Science, 67*, S163-179.

Woodward, J. (2000b). Explanation and Invariance in the Special Sciences. *British Journal for the Philosophy of Science, 51*, 197-254.

Wu, J., O'Connor, C., & Smaldino, P. (forthcoming 2023). The Cultural Evolution of Science. In J. Kendal, R. Kendal, & J. Tehrani (Eds.), *The Oxford Handbook of Cultural Evolution.* New York: Oxford University Press.

6.0. The Accuracy and Content of Scientific Representations: A Robust Pragmatic Account

Chapter 5 distilled the historical details of the preceding chapter into a developmental epistemology. The patterns identified there show how an understanding of data-gathering practices becomes reliable over time. I also argued that the revisions to these practices can be understood as rational processes that transfer warrant from earlier to later practitioners. Protein X-ray crystallography became an empirical program through the creation, integration, and alteration of reliable procedures of data production and interpretation spanning a century of scientific research. In the process, the domain-specific causal claims guiding these procedures were fused with an expanding range of mathematical tools. What started as a qualitative process of checking spacing and size of diffraction patterns against Bragg's Law-based predictions from hypothetical models became a rigorous method of transforming one set of quantities into another.

In this chapter, I will incorporate these results into an account of scientific representation. I use the example of Laue and Bragg's modeling strategies to advance an account of representational accuracy that depends on a model's integration with an empirical program. From an account of these integrative practices, I derive a traditionally pragmatist notion of representational content for scientific models. I argue, in short, that representational content should be understood interactively, in terms of the action-guiding inferences that are licensed in relation to a given empirical program. I show how this elaborates on elements of the use-based account of scientific representation presented in Chapter 2 and highlight several consequences of the view. Throughout, my view is contrasted with the assumptions that underlie Ontic Priority and Semantic Imperialism. I argue that this view offers an explanatory account of surrogative reasoning that avoids these assumptions, and therefore avoids the Paradox of Scientific Representation.

Several decades ago, Pickering (1995) wrote, "*how the material world is* leaks into and infects our representations of it in a nontrivial and consequential fashion" and proposed analyses that show "an intimate and responsive engagement between scientific knowledge and the material world that is integral to scientific practice" (p. 24). This perspective is in marked contrast to an account of models as free-floating abstracta

to which scientists apply a set of denotational labels—what Weisberg (2013) places under the umbrella of "interpreted structures." I believe Pickering's perspective is worth further consideration. If right, then philosophers of science miss out on important aspects of scientific models when they adopt a rarefied view that abstracts from the history and epistemic practices that coincide with the development of model-based representations. With this in mind, I will begin by presenting some ways that empirical programs contribute to the development and adoption of scientific models. In doing so, they have a direct influence on the understanding of a model's domain of application.

6.1. Empirical programs and domains of application

6.1.1. Reciprocal influence at larger time-scales

In Chapter 5, I discussed how techniques for producing empirical results and characterizations of the targets they interact with exert a reciprocal influence on one another. A further, related tendency that results from the reciprocal influence of target and technique is the following: as certain empirical methods stabilize and are adopted by a community of researchers, they become reliable with respect to a limited range of claims about target phenomena. Relevant claims are delimited by the forms of target-apparatus interaction and resolution of data that distinguish each technique. Where one technique predominates, the kinds of questions that can be answered by it gain salience, which may bias researchers toward pursuing some problems and ignoring others. This can give rise to entrenched ways of modeling a target that coincide with specific empirical programs. Mutual refinement of target and technique can tend toward equilibria, where modeling features and problems are significantly informed by the means of data production that are available to researchers. This can be seen in three phases of inquiry into the nature of proteins.

Physiological chemists' methods centered on the classification of organic substances through chemical analyses and properties observed in response to various treatments. Berzelius in 1813 classified "albuminous bodies" based on their coagulability, water-solubility, and yellowing in response to nitric acid.

Agreement or difference in these properties was the first method for identifying or distinguishing proteins. Thus Lehmann (1851) noted that proteins have soluble and insoluble forms, where the former are principally obtained from animal fluids by boiling. Coagulability would remain a signature characteristic of proteins throughout the century, prominently cited by Fischer (1907) and Osborne (1908) in their review articles. In general, protein characterization during the reign of physiological chemistry was largely a matter of (i) clustering of empirical properties, (ii) chemical analysis into a common set of elements. Hence Osborne wrote, "While it is not possible to establish the individuality of any protein, it is possible to show differences between the various forms which can be isolated, and to establish a constancy of properties and ultimate composition" (1908, p. 419). Likewise, Lehmann's complaint that "it is impossible [by this means] to form any decisive conclusion regarding their internal constitution" (1851, p. 290) remained true of Reichert and Brown's 1909 study of haemoglobin crystals.

This was beginning to change at the turn of the century. Structural models had become powerful tools in the hands of chemists. Those working on biological substances in Germany had developed analytical methods for reliably breaking complex substances into sub-components and inferring valence structure from these results. In the 1900s Hofmeister identified CO-NH-CH as a recurring unit in proteins, Cohnheim and Loew showed that proteins could be broken down into amino acids, and Fischer used these as "building blocks" linked by anhydride bonds to synthesize polypeptides. In addition to their bulk behavior under chemical treatment and other empirical properties, proteins started to be characterized in terms of molecular structure composed of recurring units linked by characteristic bonds. These were investigated through cleavage and synthesis techniques. While these methods were effective for identifying constituent bonds and components, researchers were slowly realizing that proteins were massive molecules. Their structure simply could not be deduced from biochemical techniques. This served as a fundamental limit to the range of claims that could be addressed by such techniques. Due to these difficulties, colloidal models of proteins remained viable competitors.

In the late 1930s it became possible to extract useful X-ray data from protein crystals. The methods of modeling and interpreting X-ray data that had developed in the prior decades made higher-scale

molecular structure of proteins accessible, in principle. Proteins were characterized as comprised of one or several polypeptide backbones with recurring side units, bearing "complete internal regularity […] down to atomic dimensions" (Bernal, Fankuchen, & Perutz, 1938, p. 524), either extended in a long fiber or folded in upon themselves. With diffraction data on hand, Perutz and co-authors first modeled the external shape of haemoglobin crystals, their packing within liquid, and eventually (with the help of novel technologies) its tortuous internal structure.

In each of these cases, the predominance of a particular empirical program enabled distinct ways of characterizing proteins. Some questions about the nature of proteins could be fruitfully investigated by these means, other questions were unanswerable—if raised at all. It would make little sense to wonder, in the 1800s, about the internal folds of protein chains, as opposed to any other of a potentially infinite number of three-dimensional models. The transitions between these phases of relative equilibrium was brought about by changes to the entrenched ways of characterizing protein structure, themselves spurred by the introduction of novel techniques. The relationship between target characterization and techniques illustrated here should lead us to reject the view that empirical programs only play a negative role with respect to scientific representation (i.e., as a constraint or reduction in the size of an abstract "model space"). Instead, this history provides evidence for the positive contribution of empirical programs to the representational status of scientific models.

6.1.2. Empirical constraints and incitements

I have described how the dynamic driving the development of an empirical program can reach a form of equilibrium where the questions pursued within a program are formulated in terms of prior refinements in target characterization enabled by that same program's techniques. This may turn into stasis or decline if the questions raised with respect to the target phenomenon outpace the resolving power of available techniques. Hence protein crystallography's standstill by the early 20th century, as researchers began to pose questions with respect to molecular structure that received techniques could not answer. This

supports a deceptively simple point about the relation between the conception of a target of inquiry and the means of investigating this target: *what can be said, with scientific plausibility, about the nature of a target is constrained by the techniques involved in its characterization.* Moreover, these techniques play a positive role—serving as incitements—in establishing regions in model space within which representations gain empirical traction and inform users about their targets.

A scientific concept like "coloring matter" (or *Blutroth*) is coined and acquires consistent use within a specific historical context—in this case, the mid-19th century, German-dominated field of physiological chemistry. Like the chemical formulae that were soon applied to it, it first served as a superficial characterization, i.e., to summarize the empirical results that had been established in physiological studies of blood composition. Above all, it was the fact that blood crystals showed the same absorption band as the coloring matter of blood—but not its products of decomposition—that led Hoppe to claim that the coloring matter was the primary constituent of blood crystals. These crystals were points of stability that could be incorporated into a range of empirical techniques, producing novel data and informing their interpretation. Like other animal substances, haemoglobin could be examined under the microscope, treated with reagents to produce novel qualities, analyzed into its elements and their combined molecular weight, and so on. The concept of haemoglobin was thereby inferentially linked with a network of ongoing scientific procedures that enabled predictions far beyond the immediate locus of Hoppe's findings. Similarly, techniques other than crystallization were already available for isolating the coloring matter and treating it chemically. It would thus be unnecessarily restrictive to claim that the scientific meaning of "haemoglobin" was limited just to those operations by which Hoppe confirmed its identity with blood crystals, or any set of operations of this sort. Still, there is a significant form of constraint provided by these interlocking sets of procedures and the empirical results they yield. These limit the range of informative, action-guiding inferences that can be drawn with respect to "haemoglobin."

Take inferences about haemoglobin's structure in the early twentieth century: to be informative, these had to be able to rule out some alternatives within a well-defined contrast class. Although any kind of structure imaginable could be posited for haemoglobin, it would not have been informative for scientists

to posit any conceivable arrangement that fell under the general concept of structure (whatever that means). It was plausible to speak of the structure of haemoglobin only because it was the sort of thing that was regarded as having structure in the sense that chemists were coming to regard this concept at the time—i.e., with respect to the valence relations between constituent atoms, typically construed as having a three-dimensional, geometrical form. To be minimally informative, claims about haemoglobin structure had to rule out some structure *of this sort*. The deeper trouble, from the perspective of crystallography, was that the available techniques were inadequate for saying anything definite about haemoglobin's structure, even in the accepted sense of three-dimensional form. Given the established methods for characterizing haemoglobin within protein crystallography, an extensive range of alternatives with respect to its structure could be neither ruled in nor out. None of the available techniques would allow researchers to infer structural features at the level of detail scientists sought. In this way, the scientific use of the concept of haemoglobin—the inferential role it played, the things that could be said of it with respect to structure, and the actions that could be taken to investigate this—was limited by the procedural context within which the contours of this concept could be explored.

A comparable situation can be found with respect to scientific models. Unlike concepts such as "haemoglobin," models are relatively self-contained, in the sense that they can be manipulated and used to derive results according to a set of internal inference rules. Yet, as Boesch (2017a) has argued, the representational function of a model is also conditioned by contextual elements shaping the purposes of model users. Though they are not reducible to summaries of empirical results, scientific models are developed in response to ongoing empirical inquiry into specific target domains, as they are understood by working scientists at the time, and often have features built into them based on presently available data. Modeling choices are empirically constrained from the start, in virtue of those facts with which a model is designed to be consistent. Forms of empirical constraint affecting models are found in the assumptions built into them and in the results one aims to derive from them.

The contribution of these forms of constraint is historically specific and varies between the more and less explicit. This makes them difficult to systematize with respect to the wealth of scientific models. I

will settle for some examples. The utility of a scientific model for drawing informative inferences often builds on its ability to unify familiar empirical effects within a single, self-consistent form (e.g., by enabling users to derive comparable results from a set of core assumptions). The debate over the ether impulse and neutral pair models of X-rays hinged on precisely this point: Barkla's pulses were better suited to explaining secondary radiation; Bragg's particles better at accounting for the ionization of gases. In both cases, however, there was a well-known set of empirical results that both models had to produce in order to be plausible: the rays should travel in straight lines, concentrated within a solid angle from their source of emission; they should have a penetrating power affected primarily by the character of incident material; they should be consistent with production by cathode rays striking a surface; and so on. These were not trivial demands; they drastically constrained the possible forms that candidate models could take by requiring that these effects be derivable from them.

In other cases, the decision to include or exclude some feature from a model is directly informed by empirical results. We saw this with Sommerfeld's attempted ether impulse model of diffraction, in response to Haga and Wind's controversial photograph (Ch. 4, Sec. 3.2). In his model, Sommerfeld assumed there was no additional contribution from pulses reflected from material on the thin edge of a diffraction slit. This was because, he claimed, there was no good evidence that X-rays could be reflected in this way. The choice of a particular boundary condition in his model was directly motivated by the empirical results available to him. Similarly, Perutz's interpretation of density maps, which led to the ill-fated parallel chain model of haemoglobin, was based on confidence among protein crystallographers at the time that globular proteins shared their basic structure with their fibrous kin. Astbury's X-ray analyses of fibers were among the empirical results that established a repeating chain structure for fibrous proteins and was likely to influence Perutz's search for analogues in haemoglobin.

This point applies beyond so-called empirical models. Both forms of constraint are found in Laue's model-building approach. Laue's theoretical model of X-ray diffraction assumed the validity of the ether impulse model for incident X-rays, but posited a specific form of interaction between these rays and the atoms in crystals. According to Laue's model, these atoms functioned as resonators, absorbing the energy

from pulses and then emitting them again in the form of new, wave-like X-rays with a discrete number of wavelengths. This assumption was directly inspired by the secondary radiation effects established by Barkla. Barkla had shown that different materials could absorb "primary" X-rays and then re-emit them as characteristic "secondary" radiation. Moreover, these secondary X-rays were homogeneous—i.e., had a fixed penetrating power—and were characteristic of the irradiated material. These empirical results rendered Laue's model a plausible mechanism for diffraction phenomena, providing important motivation for its construction.

The role often assigned to experimentation vis-à-vis scientific models summons a vision of a preordained model space viewed from without, which empirical results reduce in volume. A better picture of the process I have in mind is that of a space of variable size and dimension whose contours are explored from within. The first presents a God's eye view of hypotheses in the sky slowly cordoned off. The second envisages miners digging in different directions, sometimes opening up caverns, sometimes causing them to collapse. This is why I describe empirical results as incitements as well as constraints. As with Sommerfeld or Laue, empirical results may directly inform the choice of assumptions built into models. Other results may establish some basic requirements that models ought to satisfy, i.e., consistency with the target characterized in terms of these empirical results; ability to produce derivations comparable with these results; and so on. Models that do not meet any such demands are reasonably dismissed as implausible. But empirical results also have a positive role; they direct modelers to a particular set of assumptions to explore more thoroughly, as Darwin, Compton, and others did in modeling the energy of waves reflected from the electrons surrounding atoms under various configurations. Empirical results push researchers to develop modeling tools beyond established knowledge, as seen in the growth of mathematical interest in matrix algebra and probability theory brought about by quantum mechanics and the statistical sciences of the 19th century. And empirical results characterizing dynamics in one domain may encourage adoption of modeling techniques developed in another, as with Lotka's transfer of formulae for the consumption of a substrate by a chemical reagent to predator-prey relations, or Fisher's transfer of statistical modeling from the interactions of gas particles to those of gene carriers.

If we take seriously the claim that models are developed in response to historically-conditioned empirical constraints and incentives, then we would expect the features of a model to be developed with a preferred interpretation based on the empirical results characterizing (however superficially)[151] its target phenomenon. A model's ability to supply informative inferences functions against a background of plausible uses conditioned by available empirical results. Bragg's wave model of X-rays, for instance, would not have been accepted were it not for the patterns produced in Munich, along with the growing body of empirical results suggesting analogies between X-rays and light. Conversely, the application of a model plays an inferential role in characterizing a target from empirical results. Thus, the success of Bragg's model led to the acceptance of the Munich results as evidence of the phenomenon of X-ray diffraction. The historical interplay between empirical constraints on models and model-based characterizations of target phenomena leads to forms of equilibria between scientific modeling and empirical research, as I claim are found in the long history of protein science. The full implications of this interplay for the relationship between model and target depend on the following analyses of representational accuracy and content, and will be discussed in Section 6 of this chapter.

6.2. Representational accuracy

At this point I anticipate an objection of the following form: "Everything presented here describes how a community of scientists come to judge or believe a model to be adequate for representing some target phenomenon. Similarly, well-guided empirical techniques warrant scientists' beliefs that some empirical result is reliably connected to a phenomenon. It is a mistake, however, to think this accounts for the ability to draw inferences about a target from a model. This only explains why scientists believe they are justified in doing so. Historical influences on what scientists deem plausible restricts the range of targets that scientists relate to models; data collection and analysis provides evidence for the claim that these relations

[151] See the end of Chapter 5, Section 2 for a discussion of superficial characterizations.

hold. But there remains the question of the relation *itself*, separate from these beliefs, that ties a model to its target phenomenon. The existence of this representation relation is a separate matter, the ontological status of which can be considered independently from the justification for its being posited in any individual case. And only this posit, justified or not, is needed to explain the ability to draw informative inferences about a target. Whether there is supporting evidence for this relation is a secondary matter."

This objection is based on an appeal to a principle discussed in Chapter 1. There I referred to it as *Ontic Priority*: the claim that a theory of scientific representation must begin with a particular account of the *what*—the nature of scientific representation qua model-target relation—in order to explain the *how*—that is, the ability to use models to reason about target phenomena. On this view, no epistemic matters affect whether a relation between a model and its target obtains. This entails that the accuracy criteria of a model with respect to some target can be analyzed independently from criteria involved in evaluating the model, say, with respect to data generated from that target.[152] In Chapter 1, I identified two conjoined claims involved in a commitment to Ontic Priority: first, that a set of ontic conditions are responsible for, and so explain, the outcome of epistemic evaluations of modeling accuracy; second, that an investigation of these conditions is therefore methodologically prior in any effort to establish accuracy criteria. I called this methodology an ontic-first approach.

The defender of Ontic Priority seeks the *source* of these criteria—that which makes a model a *truly accurate* representation of a given target. They seek analyze these criteria in isolation from those involved in evaluating the model with respect to data generated from that target. Scientists may not actually compare model features to target features directly, they claim, but the reason a model is in fact accurate with respect to this data is due to the appropriate ontic conditions. On this view, the task of determining the success criteria for accurate representation is independent from consideration of practices like data collection and

[152] This view is not necessarily affected by the common claim that whether a model represents a target depends on the specific purposes of its user and the specific features that one is targeting. Thus, in Contessa's (2007) account a user provides an interpretation of a model in terms of a mapping from model to target features, which involves no reference to data gathering processes. Weisberg's (2013) weighted feature matching account similarly specifies a model-target relation where the relative importance of matching features is weighted according to users' goals as part of an account that is conceptually prior to considerations of how models are confirmed.

analysis. These are viewed as strictly epistemic concerns associated with the testing or confirmation of models. Facts about empirical techniques and instrumentation, about how one *comes to know* whether a model affords inferential success, are separate concerns. If the question of the representation relation between model and target were settled, one could argue that models are accurate if and only if the proper representation relation in fact obtains between them and their targets. How, after all, can we assess whether a model represents its target well without some fundamental, ontic presuppositions about its relation to a target?

The trouble with this view, as stated in Chapter 1, is that ontic-first approaches produce criteria for the accuracy of models that are not accessible to many cases of scientific practice. They claim that a model is accurate if and only if a specific relation obtains between model and target. It is suggested that this criterion is applied through a comparison of model and target features, where this is understood to verify whether the appropriate relation between model and target obtains. Prima facie, this fails to account for typical cases where a characterization of a target's features is only given by the model under consideration. In such cases, a model's ability to successfully predict an empirical result does not by itself warrant the conclusion that a specific relation obtains between the model and its target. This remains true even if the appropriate relation exists between a model and a data structure. One may still posit that the model successfully predicts results in virtue of the proper relation existing between model and target, though this is best viewed as a metaphysical thesis that does not capture the standpoint of scientific practice. If a theory of scientific representation is meant provide accuracy criteria that scientists *themselves* can apply, then we must look elsewhere.

I will propose an alternative account of representational accuracy based on the historical studies in this dissertation. Consider scientists' judgments of the accuracy of the wave continuum model, which evolved alongside the empirical techniques used to characterize X-ray phenomena. A similar strategy for judging this accuracy is found in both Sommerfeld's application of the ether impulse model and Laue and Bragg's use of wave models. These authors proceeded by integrating their models into an inferential framework that reflected the specificities of the data generating process. Additional structure was added to

the model of interest that reflected the interaction that experimentalists had reliably isolated: in Sommerfeld's case, the interaction between a traveling ether pulse and a geometrically modeled slit; in Laue and Bragg's case, that between traveling X-ray waves, both among themselves and with a crystal structure. By this means, they derived model results that could be compared with recorded data.

Let's see how this worked in the cases of Laue and Bragg. Guided by a theory of technique, Friedrich and Knipping were able to record an interaction between a crystal and X-ray apparatus. This allowed for an initial characterization of their data as reliably produced by this interaction (Fig. 1).

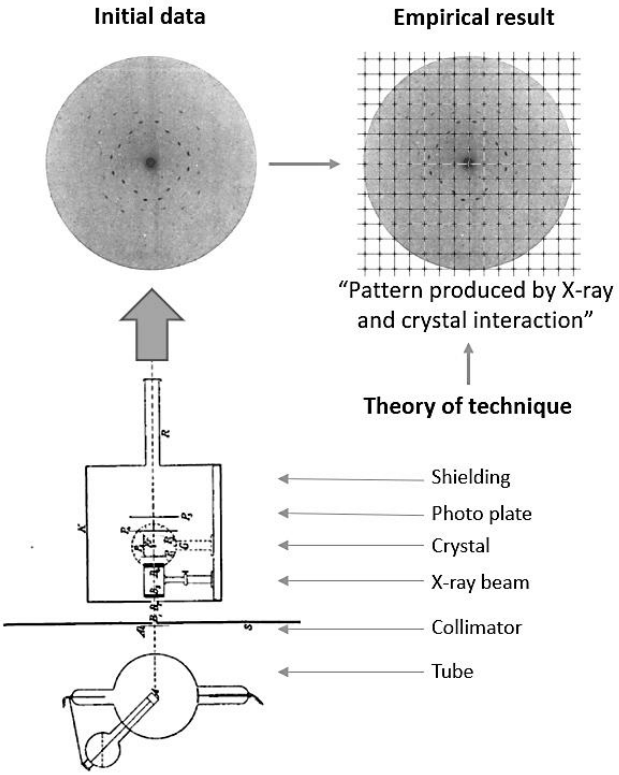


Figure 1: Producing an initial empirical result using a theory of technique.

Laue and Bragg developed two different models to account for this pattern as a diffraction effect. Each took a particular model of the X-rays—for Laue, a finite set of wavelengths emitted from an irradiated crystal; for Bragg, a continuum of waves reflected by the crystal layers—and incorporated this into the experimental setting as described by the theory of technique. Both Laue and Bragg developed a model of the process isolated within the experiment (Fig. 2).

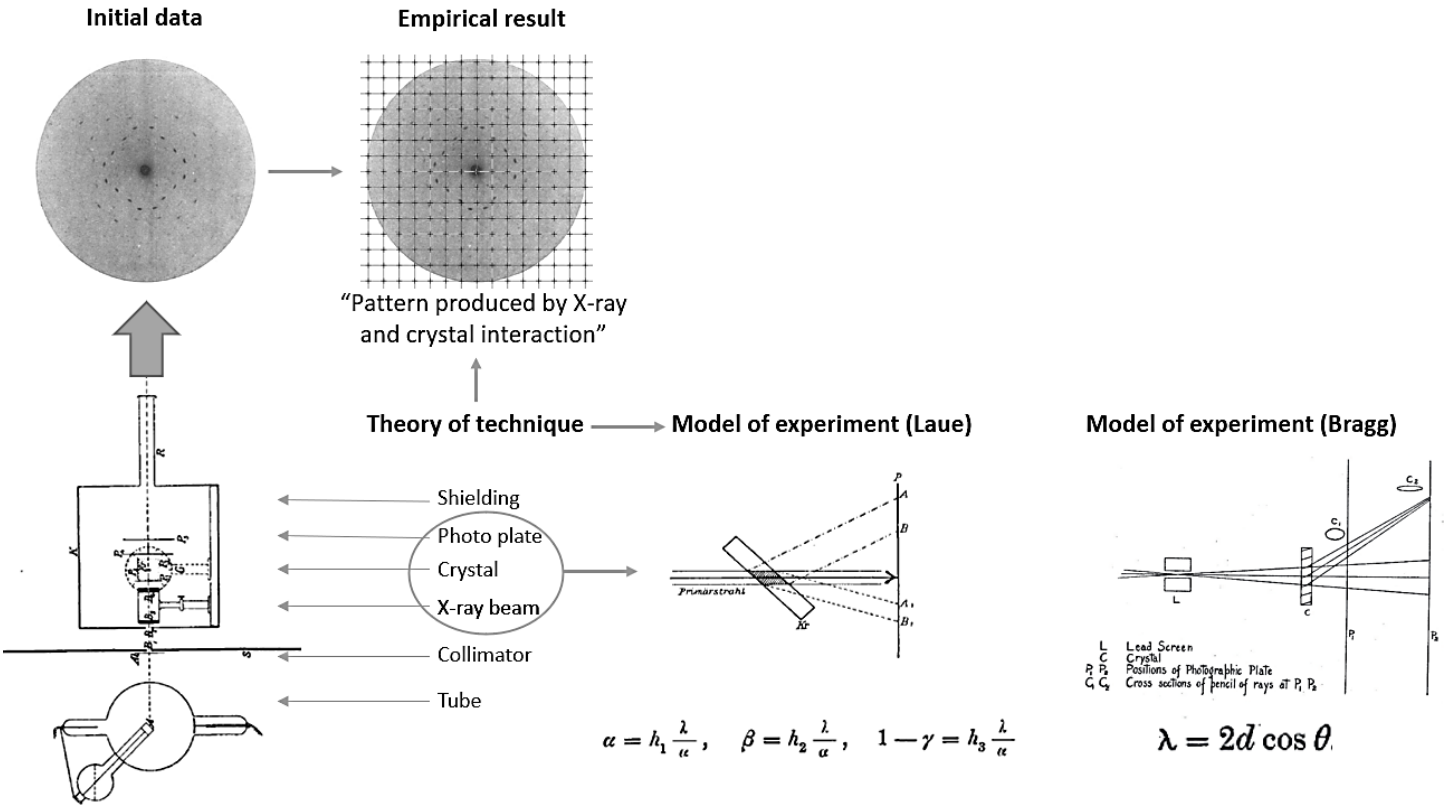


Figure 2: Integration of different wave models with the description of the experiment.

Once integrated with a theory of the technique, these models could be used to produce derivations that were comparable to the empirical result (Fig. 3).

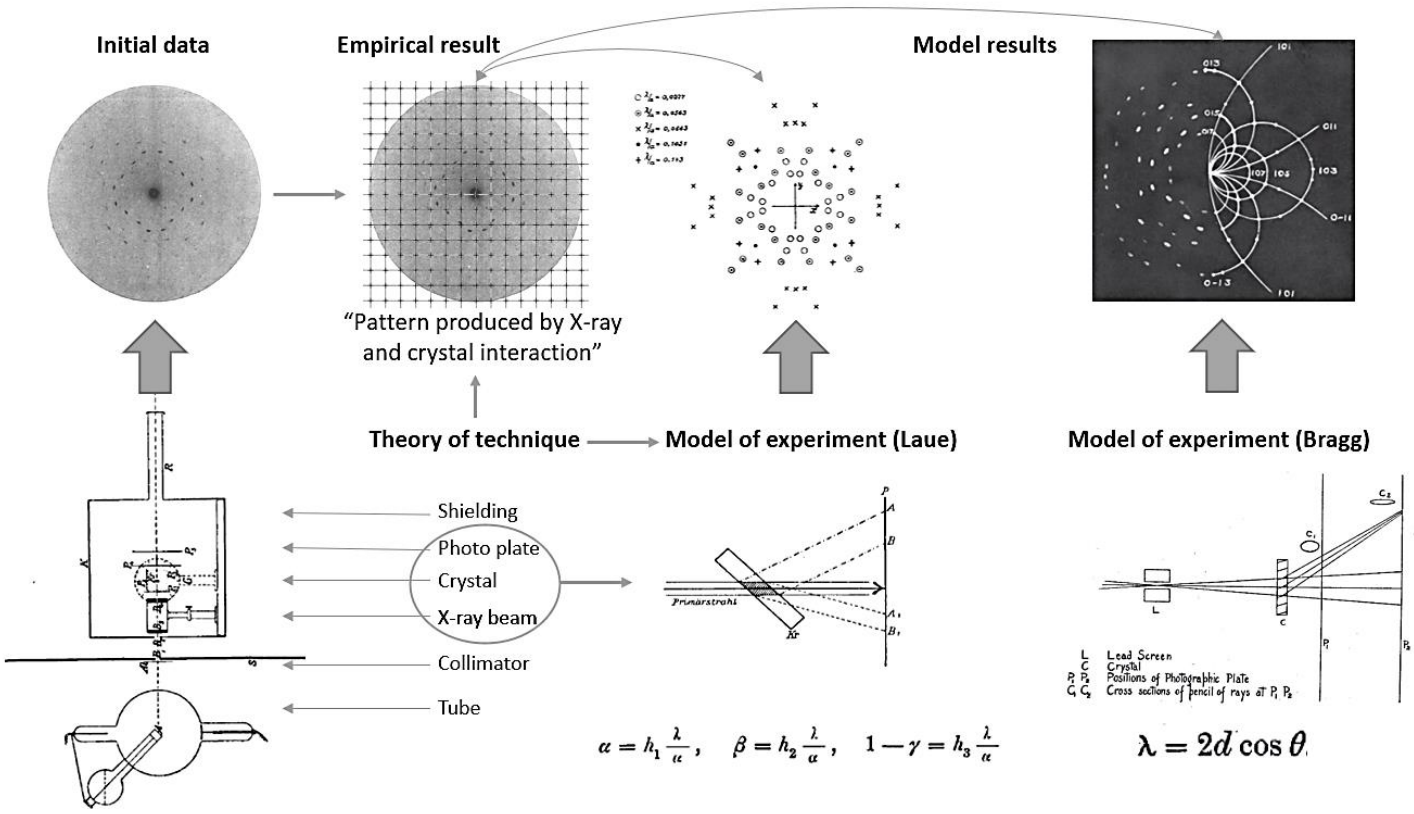


Figure 3: Derivation from model and comparison with empirical results.

This enabled the models to be compared to data *and* to further characterize their target phenomenon. The models provided a means to interpret Friedrich and Knipping's pattern as the product of X-ray wave interference. It wasn't the comparability to data alone that underwrote judgments of Laue and Bragg's models; it was the fact that these models could predict data patterns in a way that cohered with the theory of technique. Moreover, Bragg's model was accepted over Laue's not only because it more precisely fit the data pattern but also because no further, arbitrary decision was required on the part of the modeler to produce this fit once the model of the experiment was in place. In such a case, characterizing a target phenomenon from experiment is at the same time a model-based predictive process with respect to an empirical result.[153]

[153] There is a parallel here to Tal's (2017) account of measurement as the outcome of a predictive model of the measuring process. Similarly for my account of target characterization below (Section 6).

The coin toss and test response examples from Chapter 1 showed that a mere fit between a model and an empirical result is insufficient for judging a model to be accurate. In addition, the empirical result must be reliably connected to the phenomenon that the model user targets. Cases from the history of X-ray research demonstrate a general strategy for judging a model's accuracy on this basis, even when a prior model-independent characterization of the target phenomenon is absent: include the theoretical model as a component of an account of the data gathering process, further specifying or concretizing the model as necessary, and use *this* to derive results comparable to the empirical result. This is an instance of a general criterion according to which a model is accurate if and only if integrating the model into a theory of the data acquisition process yields well-fitting predictions of patterns in the data.

Mitchell (2003) distinguishes general theoretical models from their specific applications in concrete cases. The former are highly idealized in the sense that they do not reflect the majority of the causal factors that play a role in the occurrence of an actual phenomenon. In cases of application, these must be tailored to the causal complexity of the case at hand. Different models may help account for the contributions of different causal factors. Where these models overlap, this may require what she calls *piecemeal integration*. Certain assumptions or features of a model may be relaxed, ignored, or reworked in order to reduce conflict with others and to provide a coherent account of the case. This method of integration can be seen in the application of Laue's and Bragg's models to Friedrich and Knipping's experiment. Each was working with an abstract model of X-rays—ether impulse vs. transverse wave continuum—which was then tailored to the diffraction case. Laue sought to model these waves in relation to their absorption and emission by resonators in the crystal; Bragg in terms of their reflection from crystal planes. In each case, this involved downplaying some features of the abstract model. Laue largely ignored the properties of the incoming pulses, including those that Sommerfeld held would produce an indiscriminate blackening of the photographic plate, and focused on the wave-like byproduct of their absorption. Bragg only considered his waves in terms of how they would behave when reflected from regularly spaced planes at a given angle. Both focused on a limited set of features drawn from more general models of X-ray waves and crystal

structure, and used these to examine the interaction isolated in the experimental setup in greater causal detail.

The theory of technique functions as a linchpin in this account, serving two inferential roles. One is retrodictive, moving from recorded data to an interpretation of these in terms of causal factors within the data acquisition process. The other is predictive, working in conjunction with the model to simulate a data pattern. The first specifies those factors in the experimental set-up that are the primary contributors to the data. The second directs the inclusion of additional elements into a model and their interpretation in terms of these specified factors. Hence both Laue and Bragg incorporate the X-rays, the crystal, photographic plate, and their spatial relations into their accounts. By tying these terms to elements of their models, they embed it within the set of inferences that comprise an established theory of technique. In this way it provides the causal specificity lacking in the abstract models.

This provides scientists with a means for evaluating a model's accuracy: it must derive a result that can be compared to a data pattern *and* it must cohere with the technique that generates that data. This is not a post hoc reconstruction of the model's success or failure in terms of its relation to the target. Rather, it is a way of incorporating the model into an account of the apparatus, and so into the practical inferences that guide actions in this context. For example, integrating a model into a theory of technique licenses the move from a claim like, "my spectroscopic analyzer isn't showing a strong reading at this angle" to "the reflected X-rays are not interfering constructively at this point." With the Bragg model on hand, one can sensibly respond to the above inference with the suggestion to rotate the *crystal* rather than the detector.[154] A significant contributor to the model's accuracy is whether the practical inferences it licenses bear out in such contexts. On the other hand, the only direct comparison taking place, of the kind that resembles a similarity judgment, does not concern the model and its target but its derived result and a data pattern. When a model is properly integrated into this broader inferential network, and its simulated results satisfy standards of fit vis-à-vis recorded data, then the model is accurate with respect to its target phenomenon.

[154] I explore this claim in more detail below in Section 6.

This criterion of accuracy is broadly applicable by practicing scientists and is readily found in many areas of model-based research.[155]

6.3. Ontic Priority reconsidered

A few points of clarification are in order. The account I have presented does not entail that, in all cases, scientists or philosophers ought to cease thinking of models in abstraction from data-gathering practices. In Chapter 2, this matter was discussed in terms of the distinction between representational and non-representational uses of models. I do not deny that scientific models are employed in abstraction from specific practices for legitimate scientific purposes other than evaluations of accuracy, though these often retain implicit empirical content as a result of the process of their construction. Nor does my account entail that a model's accuracy can only be determined with respect to a single data gathering process. Precisely because of their capacity to be abstracted from specific contexts, theoretical models can be integrated into multiple lines of inquiry. Though the success criterion for model accuracy always depends on the way it is integrated into a theory of technique, the specific character of this criterion—including the standard of fit applied to its predictions—may vary with the form of target-apparatus interaction theorized in each case. Accordingly, the same model can be judged by different criteria based on different forms of interaction with the same target phenomenon and may even yield different judgments of accuracy in each case. I have proposed a general criterion for model accuracy, but this is compatible with pluralism regarding the specific form the model-data comparison can take in different instances and the corresponding judgment that results.

The account presented thus far only pertains to how criteria of model *accuracy* should be addressed by a theory of scientific representation. Still, I anticipate this view will not satisfy proponents of Ontic Priority. I have said nothing about the nature of the relation between a model and its target phenomenon

[155] To take one example, the methods of judging the accuracy of theoretical models of particle physics at the LHC accord with this criterion. There, models of collision events are first specified based on the operating energy of the accelerator using a suite of techniques referred to as particle phenomenology, their decay paths are simulated, and then integrated into a much larger model of the detector itself to generate a prediction of LHC data.

that, one might think, grounds the accuracy of a scientific model. On this view, it is a fact whether such a relation obtains, and thus it is a fact whether the model is accurate or not. I have only described how scientists can judge the accuracy of models, a form of judgment that might vary based on the theoretical background and technique in use. But, on an ontic view, whether a model is an accurate representation of a target is a fact that cannot vary due to these epistemic factors. Whatever the method that scientists use to judge a model's accuracy, its success is only explained by the appropriate relation obtaining between model and target.

I have responded to this line of reasoning by raising the issue of scientists' epistemic access to such facts. In Chapter 1, I argued that the relation between model and target theorized in abstraction from epistemic factors is a kind of fact that is often inaccessible from the standpoint of scientific practice. There is no general way to infer *from* the kind of information that scientists can access *to* the precise nature of the relation between a model and a target, beyond how the model itself coheres with the data acquisition process. The argument I take to favor this model-integration account is based on the further desideratum that an adequate theory of scientific representation ought to propose success criteria for model accuracy that scientists themselves can, and do, apply. But a defender of Ontic Priority can push back here by distinguishing, on the one hand, the ontic relation in virtue of which a model is successful and, on the other, the facts about modeling success that are epistemically accessible. They may concede that these latter facts might not include or even entail identifying information about the ontic relation itself, but still insist that it would be a mistake to draw conclusions about what accuracy *is* from facts about how one comes to judge a model to be accurate. One might modify my desideratum on this basis, proposing that a theory of representation ought to provide both a metaphysics of accuracy *and* an epistemology of those facts and procedures that determine accuracy judgments.

It's a fair proposal. The basic assumptions that comprise a metaphysics of accuracy—that representation is grounded in a strictly ontic model-target relation and that accuracy can be defined in terms of whether this relation obtains—are self-consistent and are not contradicted by my view. By these counts, ontically-minded philosophers are perfectly free to consider the possible relations between models and

targets, while others fuss over practice. The only issue concerns the explanatory connection between these two realms. Defenders of Ontic Priority claim that the successes of model-based reasoning observed in scientific practice cannot be explained without an account of this ontic relation. But if the facts relevant to evaluations of modeling success that are accessible to scientists neither include nor entail facts about this ontic relation, and if philosophers do not have privileged access to certain facts over the methods available to science, then it seems we are left with two choices: Either we accept that this success can be explained in terms that reflect scientists' evaluative resources (i.e., without reference to an ontic relation) or we are left to explain this success by appeal to facts that are epistemically inaccessible.

I have proposed a notion of accuracy that follows the first disjunct. Contra Ontic Priority, I affirm that the model-integration criterion *does* explain how models are used to accurately reason about target phenomena. It gives an account of models being integrated into a theory of technique, of their use to simulate data, and of a comparison being made between simulated and reliably recorded results. This is not the kind of explanation favored by an ontic-first approach, but the alternative appears deeply obscure from the standpoint of scientific practice. The second disjunct invokes a form of explanation that raises more questions than it appears able to answer, insofar as it depends on assertions of fact that cannot be confirmed through the methods of fact-finding available to us. What, after all, is the difference between an explanation based on the existence of an epistemically inaccessible fact and one based on an a priori postulate? Each explanans plays the same functional role, insofar as their truth cannot be evaluated a posteriori. This is a particularly strange role for factual claims, since these are just the things that we typically regard as evaluable a posteriori. Whatever these inaccessible facts and accompanying explanations are, they seem to belong to a realm that is unfamiliar to the naturalistic philosopher of science.

At issue here are the conjoined claims that the ontic explains the epistemic and that ontic questions have methodological priority over the epistemic. If a metaphysics of accuracy is a desirable complement to my epistemic account, and if it is to be compatible with naturalism, then one of these must be given up. First, a metaphysics of accuracy could be defended on grounds other than its explanatory merit. One could argue that there is conceptual value to considering hypothetical relations between models and targets in

isolation from the question of how these are empirically instantiated and affect scientists' determinations of accuracy. In this case, a notion of accuracy might be investigated as a purely theoretical construct with no direct bearing on practices of model-target comparison. Distinct notions of ontic and epistemic accuracy could be developed for distinct philosophical ends. On the other hand, if this separation is unsatisfactory, then the connection between scientists' evaluations of modeling success and the ontic conditions in virtue of which their models are successful needs to be articulated in further detail. One way to bring an ontic account into harmony with the epistemic view defended here would be to incorporate the latter's pluralism and contextualism. On this view, different ontic relations will play the role of grounding representational accuracy in different cases, in a manner that is sensitive to the practical and theoretical context at hand. Here a practice-first approach may be more fruitful for analysis. Instead of starting from a singular relation, inquiry can proceed from scientists' evaluations of accuracy in contexts of empirical application, working out the metaphysical implications on this basis.

6.4. Semantic Imperialism reconsidered

6.4.1. Pragmatic representational content

Let's consider a related objection: successful representation and in particular the epistemic processes that establish the accuracy of successful representation are different from representation *tout court*. A scientist can represent and reason about a target system regardless of whether there is any way of checking the accuracy of this reasoning. All that is required to explain this ability is that the existence of some denotative, component-wise mapping relation between aspects of a model and its target. There is nothing incoherent in claiming this in a circumstance where a target is not well-described without a model; the notion that the model is related to something other than itself is a necessary component for making sense of representation. That is, without this assumption it makes no sense to claim scientist use models to reason *about* target phenomena. Of course, a scientist could be wildly mistaken in treating their model as a *good*

representation of a given target, but this matter is secondary to their treating it as a representation at all. It is a conceptual necessity, then, that there be some relation between the model and its target. Ontic-first approaches simply aim to characterize what this relation underlying the concept of scientific representation might be. The best I have shown with the preceding is that the concept of accurate representation depends on epistemic factors, but it would be unsound to claim this of representation *tout court*.

We need to work more carefully through the assumptions about scientific representation presented in this objection. Recall, first, that scientific representation is distinct from stipulative fiat. Scientists cannot just claim anything represents anything else and, on this basis, reason effectively about their target. This is why enabling informative inferences is identified by Suárez (2004) as a necessary condition for model-based representation in science The requirement for informative inferences is supposed to reflect the objectivity of scientific representation, in contrast to other forms of representation. Second, we need to distinguish internal and external inferences with respect to a model. Deriving a result with a model does not by itself license informative inferences with respect to a target. We thus require some further conditions that define and constrain how model and target are related in a way that allows for informative inference. Appeal to mere similarity is insufficient, for reasons laid out in Chapter 1. Instead, most authors have claimed that some criterion of relevance, based in scientists' purposes or goals, is also needed to sharpen the relation between model components and a target's features of interest. Still, this relation between parts is taken to be denotative and to conform to a component-wise mapping between parts of the model and parts of the target, weighted according to a relevance function determined by modelers' purposes. This denotative mapping establishes the representational content of the model as those conditions that it alleges are true of the world. The ability to grasp such content does not depend on whether these conditions in fact hold, i.e., whether the relevant target's parts, properties, and behaviors are truly ordered according to the model, but such grasp is required to allow for informative inferences.

According to this objection, a denotative mapping is the minimum that must be presupposed by any account of scientific representation. Without it, the notion of representation just doesn't make sense. In Chapter 1, I described this view as a product of *Semantic Imperialism*. It leverages a particular tradition

within the philosophy of language to account for scientific representation. This entails a narrow account of representational content for scientific models rooted in this claim: that which determines what the model "says" of the world is restricted to the relation between model and target, without reference to matters of concrete application. In Chapter 1, I discussed problems facing Semantic Imperialism. In Chapter 2, I argued that a necessary condition for a model derivation being informative is its being comparable to the results of a data-gathering practice. Here, I will use the view that I advanced in the previous sections to put forward an alternative account of representational content. This account does not identify a model's representational content with the hypothetical conditions in the world that would make the model true. Instead, it content is comprised by the action-guiding inferences that the model licenses when integrated with a theory of technique. This account explicitly frames scientific representation in terms of the warranted use of models as inferential tools in relation to an empirical program.[156] I will argue that this notion of representational content can explain the use of models to draw informative inferences in a way that does not depend on a denotative relation between model and target as a foundational presupposition, nor does grasp of this content presuppose such a relation.

Let's go back to the example from Section 4. There I stated that integrating a model into a theory of technique licenses the move from a claim like,

(1) "My spectroscopic analyzer isn't showing a strong reading at this angle"

to

(2) "The reflected X-rays are not interfering constructively at this point."

Familiarity with the model in relation to this technique would then license a practical inference[157] to

[156] This account shares many inspirations and aspirations with work on scientific representation currently being developed by Khalifa, Millson, and Risjord (2022).

[157] By practical inference, I mean an inference whose consequence is a normative judgment with respect to a particular action. If this inference is licensed then the action is warranted. I also refer to licensed practical inferences as 'action-guiding inferences'.

(3) "I should rotate the crystal and try again."

This sequence involves an inference from a claim based on concepts within the theory of technique to a claim that is based on the assumptions of Bragg's experimental model and then back to a claim understood within the theory of technique. I call these inferences from (1) to (2) and from (2) to (3) "model-entry" and "model-exit" moves, respectively.[158] Importantly, an inference from (1) to (3) would *not* be licensed by a theory of technique that is not informed by a model such as Bragg's, since this depends on specific assumptions about the nature of the interaction between X-rays and crystal structure. Laue, for instance, did not grant any importance to the orientation of planes within the crystal. He thought the diffraction strength was solely a function of the direction of incoming radiation. On this assumption, it would have made more sense to rotate the position of the detector relative to the beam line instead of rotating the crystal. Thus, the specific model of the interaction proposed by Bragg plays a direct role in licensing this practical inference.

The consequence of a practical inference is a normative judgment with respect to a particular action. If this inference is licensed, it provides defeasible warrant for acting. I call such an inference action-guiding and claim that *the set of action-guiding inferences licensed by a model when integrated with a theory of technique is that model's representational content*. I call this *pragmatic representational content* to distinguish it from other notions. Integration of a model with a theory of technique depends on the model being understood in terms that cohere with the language of the experiment. In Section 4, this integration was presented as the combined work of several distinct aspects:

- The piecemeal integration of general theoretical models into a model of experiment (e.g., a wave model and a crystal model)
- Inclusion of additional elements into the experimental model based on the causally relevant factors identified by a theory of technique

[158] Thanks to Kareem Khalifa for suggesting this terminology.

- Interpretation of features of the model in terms of concepts used to characterize the target by the theory of technique: “X-ray,” “crystal,” “photographic plate” and so on.

These steps are depicted schematically in Figure 4:

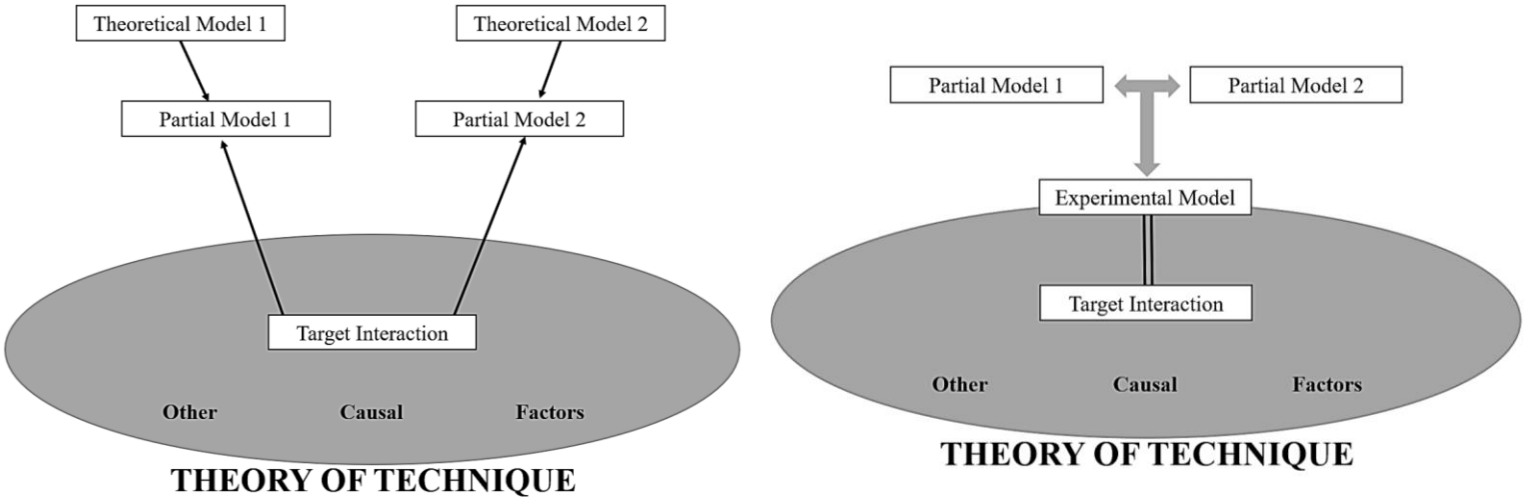


Figure 4: Schema showing the piecemeal integration of models and interpretation with a theory of technique. Left: the account of the target interaction drawn from this theory influences the choice of partial models. Right: the resulting models are adjusted (arrows) to one another and to other features of the target interaction to produce an experimental model. This model takes on (double lines) the inferential roles of the terms that comprise this interaction according to the theory of technique.

How does integration of this form permit model-entry and -exit moves? As elements of a theory of technique, the concepts used to interpret this experimental model are already enmeshed in a wider set of inferences guiding reliable activities in the lab. Thus, when scientists say some bit of modeling represents the X-rays reflecting off layers of a crystal, terms like “X-ray” and “crystal” are not restricted to the formal inference rules internal to the model. These terms also carry implications for causal factors that researchers know how to manipulate in well-defined ways to gather data. By this means the model can be used in two ways: *either* to license novel practical inferences, and so expand the range of warranted actions within the data acquisition process; *or* to provide further structural details underlying a dependency relation, thereby providing more robust warrant for a known intervention (Fig. 5).

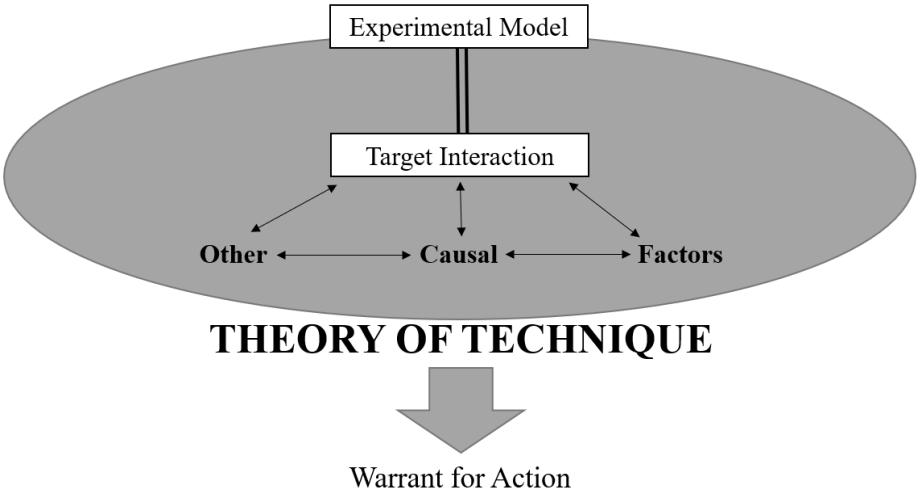


Figure 5: Schema showing an experimental model embedded in the broader inferential system of a theory of technique, which can license practical inferences on its behalf.

A theory of technique gives an account of the dependency relations between factors within a data gathering setup, along with an account of their contributions to the recordings produced by this setup. Integrating a model with this theory allows scientists to reason in finer detail about a subset of these causal factors and their effects. In many cases, including those of Laue and Bragg, the model allows researchers to simulate data patterns. Inferences from the model to interventions that alter causal factors within the setup provide a more precise way of explaining or anticipating data outcomes under some set of experimental conditions. As discussed in Section 4, scientists can then evaluate the quality of these model-based inferences by comparing data outcomes to model derivations.

6.4.2. Grasp of pragmatic representational content

"Grasp" of pragmatic representational content depends on a scientist's understanding two things: (i) the domains in which it makes sense to apply the model and (ii) the consequences of the application of the model, including its evaluation.[159] Empirical programs play a role in each of these, and I will discuss them in turn. Understanding the appropriate domains of application depends on warranted belief in two

[159] This framing is indebted to Robert Brandom's account of content. See, e.g., (Brandom, 2010).

claims: (a) that the model is plausible given existing target characterizations and (b) that the data generated within a particular domain is reliably linked to this target. The thought experiments from Chapter 1, Section 3.2 showing that similarity to data is insufficient evidence for model-target similarity are based on examples that fail to satisfy these requirements.

With respect to (ia), the first thought experiment in Chapter 1 was based on the use of a restricted model of quantum spin states measured by an inhomogeneous magnetic field to simulate patterns that fit coin toss data by a likelihood test. The coin example suggests two ways that plausibility standards constrain model applications: for one, the assumptions built into the model system ought to accord with background knowledge of its target, if available. The model used to simulate the pattern is based on the assumptions that the angular momentum of quantum systems is quantized and can exist in mixed states. Neither of these assumptions have clear analogues in a coin toss or appear otherwise justified. Second, the model states ought not be arbitrarily constrained, e.g., in a way that violates received principles or otherwise lacks theoretical justification. In this case, the quantum model has higher-dimensional states that are arbitrarily restricted to yield a desired pattern.

These norms do not provide much guidance for would-be modelers when stated at this level of generality. It is helpful to keep in mind that most scientists are working within an ongoing research context in which a body of empirical results has been accumulated and provides a preliminary characterization of potential modeling targets. Minimally, this takes the form of "that which is responsible for the effects observed under variations of a standard procedure." Chapter 5 described how prominent models in the history of protein crystallography were developed in response to specific empirical programs and were refined in tandem with methodological alterations to these programs. Section 2 of this chapter used this to claim that empirical techniques constrain and incite scientific modeling. The examples covered there show how the plausibility of a scientific model for a given target is conditioned by the facts produced by empirical programs.

With respect to (ib), the second thought experiment was based on the use of a procedure designed to generate an empirical result homomorphic to the structure predicted by a latent variable model. The

inadequacy of this method for claiming model-target similarity is based in the fact that it neglects the role of data's origination in grounding informative inference. That is, it is important that a data structure be there "for the right reasons." This dissertation has shown how empirical programs provide warrant for claiming that data acquired within a given domain is reliably linked to a candidate target. Chapter 3 described the synchronic warrant for reliability judgments in terms of the consistency and coherence of data acquisition with theories of technique. Chapter 5 presented a diachronic account of the reliability of an empirical program in terms of processes of local justification from points of stability and rational alterations of method.

As for (ii), models draw on the warrant provided by a theory of technique, which licenses practical inferences on their behalf. This is enabled by the integration process discussed above. The theory of technique provides a conceptual framework, grounded in practice, that directs the inclusion of additional elements into a model and provides an interpretation of certain model features in terms of causal factors within the data acquisition process. This allows model users to spell out the consequences of its application in terms of action-guiding inference and to simulate patterns for fine-grained evaluation of the model.

Posed this way, the scientist's grasp of a model's pragmatic representational content appears as a matching exercise, where a model is sized up against domains of data acquisition. In reality, much of this happens organically as a result of the scientist's embeddedness within a research community. The historical materials covered in this dissertation—especially the patterns of mutual refinement—illustrate how the representational content of a model can accrue gradually and in close relation to a specific empirical program or collection of programs. A researcher trained within such a tradition may learn to use models that come pre-packaged with interpretations from these programs. In their unreflective use of these models, they move seamlessly between the model and a theory of technique without fully realizing they are performing model-entry and -exit moves. These reasoning skills, absorbed through training, give the model the appearance of having a direct and obvious relationship to a worldly phenomenon. In fact, these skills are the product of a long history of refinement in relation to empirical practice.

6.4.3. Scientific representation without denotation

I want to return to the claim that denotation is presupposed in talk of representation. My claim that the elements of a model are interpreted in terms of concepts within a theory of technique requires further scrutiny, as it may appear to sneak denotation back into the picture. That is, one might argue that this interpretation implicitly appeals to a denotative mapping meant to establish similarities between relevant features of the experimental model and its target interaction. A researcher must think the layered planes in the model are *really there* for it to make sense to rotate the crystal.

But this assumption is simply not necessary for models to possess representational content in the pragmatic sense I have proposed. For a model to possess this content, it must be integrated with a theory of technique in a way that permits model-entry and -exit moves. In other words, there must be a way to infer a statement expressed in concepts from the theory of technique from one based on modeling assumptions, and vice versa. The ability to move between model and technique depends on an interpretation of the model in terms of factors within the data acquisition process, but *not* in terms of a more fundamental ontology. Yes, the crystal model has layered planes, but what gives me reason to rotate the crystal is the dependency relation this establishes between the angle of the crystal in my apparatus and the response of my detector. What matters is whether the inferences drawn on this basis work or not, in the sense that the actions it warrants help or hinder scientists' goals. On this view, there is no further requirement to regard the model as literally or approximately true, in the sense entailed by Semantic Imperialism, and I do not see why there must be.

This remains the case when the model clearly does work. In Chapter 1, I argued that the inference from empirical success (i.e., "this makes a pattern that matches my data") to representational similarity (i.e., "the thing making the data looks just like the thing making my pattern") fails. For a particular system, there may be enough auxiliary reasoning or evidence to substantiate claims of an approximate match between model and target features, but this does not hold generally. Some consequences of the definition of pragmatic representational content will show why this is so:

- Two models of the same target will have the same representational content with respect to an empirical program if their action-guiding inferences are invariant under substitution of one model for the other. This is clearest in models that possess different mathematical structure, but make identical claims with respect to the dependency relations between factors within relevant data gathering contexts. An example is the use of wave mechanics or the path integral formalism to model the behavior of a quantum system. To the extent that an ontology may be read off these models, they differ in significant ways, and yet they may prescribe the same actions in data gathering contexts.
- One model of the same target will have different representational content when it is integrated with different empirical programs. Consider a protein model derived from a folding algorithm. The techniques for predicting data from this model will differ in significant ways depending on whether it is applied to a technique like X-ray crystallography or NMR spectroscopy. Each generates data from different forms of target-apparatus interaction, focused on specific features of the target, and will call for different methods of comparison. These methods have been known to produce different models of a given protein, meaning they may differ in their evaluation of the theoretical model (Cf. (Mitchell, 2019; Mitchell & Gronenborn, 2017)).

In both of these circumstances, a view of scientific representation that purports to "read off" the structure of a target from a model will see these as problem cases. From the first, it demands resolution in terms of which quantum formalism is more true or presents a more fundamental representation of the system. From the second, it demands to know which data gathering technique gives a more correct evaluation of the protein model. On the other hand, these are not problems for my view. The representational utility of the model is based entirely in its capacity to license inferences with respect to data gathering techniques. On this view, nothing more than this is required to understand how the model functions as an informative source of inferences about a target.

6.5. The dependence of target characterization on representational models

The notion of representation that emerges from this account is based on the warranted use of models as inferential tools in relation to an empirical program. Like other use-based accounts that appeal to factors like the purposes of model users, it treats representation as a relation between a model, target, and at least one other term. Unlike these other accounts, it restricts scientists' uses to those that are warranted in relation to the body of knowledge comprising an empirical program. This warranting is based on a scientist's grasp of the proper domains and consequences of application, and the reliability of data-gathering practices within this domain. These were the two relevance conditions presented in Chapter 2 in terms of **Applicability** and **Reliability**. The former has been elaborated in terms of the account of representational accuracy and content presented above. The latter has been elaborated in terms of the account of the consistency, coherence, and comparability of data-gathering practice from Chapter 3, and the account of the historical justification of these practices from Chapter 5. This account abandons the common assumption of ontic-first approaches that an explanation of scientists' surrogative reasoning depends on an account of the relation between a model and its target. The Paradox of Scientific Representation arose due to the inaccessibility of candidate ontic relations of this sort; by denying the need to appeal to an ontic relation, I avoid the paradox.

This leaves room for describing the relation between model and target in other terms drawn from the epistemology of modeling. I have discussed aspects of this relation in Chapter 5 in terms of the reciprocal influence of model development and target characterizations, and in Section 2 of this chapter in terms of the empirical constraints and incentives on model development. Here I will put things more precisely, drawing on prior sections to clarify the role that models play in target characterization. Recall the notion of superficial characterizations of a target phenomenon (Ch. 5, Sec. 2). These are the empirical qualities of a target and its relations to other factors within a data-gathering setup. These are given by causal generalizations within a theory of the technique that produces data from this phenomenon. The role of a representational model is to "fill in" this characterization. It does so by adding further structural details to the dependencies expressed in the theory of technique. Integration of a model with a theory of technique

allows a user to move from a description of the causal factors in these dependencies into the internal inference rules of the model and back again. With this framework on hand, I claim the following:

Take a model M and a target phenomenon T, where scientists follow a theory of technique E to produce data D from T. Then M can be used to characterize T under the following conditions:

(1) M can be successfully integrated with E

(2) M can be used to simulate a pattern S

(3) D is consistent and coherent with E

(4) D is comparable to S

This is an elaboration of the conditions laid out in Chapter 2, Section 3.3 in terms of the accounts of data reliability presented in Chapter 3 and the account of model integration presented above.

Understanding the role of pragmatic representational content has the following important consequence for target characterization: When a model is integrated with a theory of technique, this may inform scientists' understanding of the data-gathering procedure. Actions that were previously warranted for producing coherent data may lose their warrant; new actions may be warranted in their place. The model may similarly affect the actions taken to render data comparable to model derivations. That is, it may recommend certain forms of data processing that were not performed prior to the model's integration with the theory of technique. We saw this in the history of X-ray crystallography. Assumptions that crystal atoms contribute to diffracting power led modelers to calculate the effect of thermal motion on the intensity of reflected X-rays and add this to the intensity formula. When integrated into a theory of technique, this formula recommended controlling the temperature during a diffraction experiment and applying a correction to data on this basis.

In cases like this, the model plays an inferential role in the production of the data used to reliably characterize the target. The model features and target characterization are not independent terms whose relation is verified by experiment. Rather, the adequacy of the model is presupposed; its structural details fill in the superficial characterization of the target within a data-gathering process. This does not mean a model cannot be tested against data. It is still possible that the modifications it recommends to data

production and processing procedures result in data that fail accepted standards of accuracy when compared to model derivations. However, this means there are multiple ways of responding to instances of failure or success: Did the data production and processing really cohere with the content of the model? Was the model properly integrated? Were arbitrary adjustments made in the process of simulating results? Were the standards of comparison too lax? Too strict? Whether model features are accepted into the characterization of a target phenomenon depends on whether these kinds of adjustments of model to experiment are permissible. On the other hand, this means that the processes involved in target characterization do not provide an *independent* test of the model, as is implied by notions of comparison between model features to target features. To the extent that a relation between model and target is uncovered by these adjustments, it is not of the kind proposed by advocates of Ontic Priority and Semantic Imperialism. On the contrary, it is bound up with facts about the techniques scientists use to extract data from this target.

6.6. Concluding remarks

6.6.1. Theoretical models and surrogative reasoning

What does this mean for the status of more general theoretical models? On most account theoretical models are used as paradigmatic cases of representational models, despite their being abstracted from any of the specificities of data-gathering setups. My view can accommodate this. It entails a "bottom-up" view of a model's generality. The generality of a representational model of a given target (or set of targets) is a function of its accuracy with respect to this target (or these targets) across multiple domains of integration. An increase in generality may come about from the extension of a model developed in reference to a single empirical program to new domains. Or, more often, a model may be constructed from the start at a higher level of generality—for example, when it is intended to account for a broadly characterized target, based on complementary results from different techniques.

Once blood crystals were identified with *Blutroth*, for instance, they were understood to share their contents with a more general target—a component substance of blood that had already been characterized in terms of various other known interventions and outcomes. Similarly, blood crystals could be intervened on in multiple ways, producing different forms of data. Due to this broad characterization, a model of haemoglobin evaluated with respect to one technique was already a presumptive candidate for reasoning about haemoglobin in other contexts. X-rays, by contrast, were initially only produced and studied by means of vacuum tubes. Models of these rays were thus developed in closer correspondence with a particular data gathering practice. But here too, there was more to draw on, as the close causal connection between X-rays and the better-known cathode rays gave prima facie reason to regard the former as electromagnetic phenomena and to apply the associated resources of physical theories. In general, target characterization is rarely isolated to a single mode of intervention. While models may be developed in relation to target features characterized by a particular set of techniques, this does not strictly tether the inferences drawn from these models to those data domains.

A model based on a target characterization derived from a single program (i.e., produced under a specific set of preparation and intervention conditions) can be used to reason about target behavior elsewhere by inductive inference from one set of conditions to another. Warrant for this induction may come from different sources. One principal source in contemporary science comes from data processing. As discussed in Chapter 3, alterations to data are frequently intended to remove contributions from non-target causal factors. Many of these circumstantial factors are left out of any model-based simulations of experiment, often for the sake of computational tractability. For instance, prediction of structure factors from crystal models do not attempt to account for the effects of heat, polarization, absorption, and so on. On the side of the data, these contributions are all corrected prior to any comparison to a model. Both the model prediction and processed data are therefore abstracted from many of the local specificities of the original data acquisition process. If data have been shorn of local factors, scientists have increased warrant for treating a model's accuracy in one context as evidence of its accuracy elsewhere.

Perutz's structural model of haemoglobin provides another example of the generalized use of a model. Beginning in 1965, he and his team began to publish a series of papers on the structure and function of this protein, based on higher resolution structural models that were becoming available. The first of these combined four sources of information to construct a tentative atomic-resolution model of the haemoglobin molecule. The model's construction was based on X-ray crystallography data, supplemented with several other techniques. The primary source of evidence was the celebrated 1962 model of horse oxyhaemoglobin at 5.5 angstrom resolution. Because this resolution was insufficient for clear atomic-level details, these were filled in by reasoning from the following sources: Kendrew's higher resolution Fourier synthesis of myoglobin, the three-dimensional structure of which closely matched each of haemoglobin's four subunits; amino acid sequencing data for horse haemoglobin; and a 1956 study of horse methaemoglobin by Perutz using electron spin resonance. While the 1962 model gave the overall form of the molecule, sequencing data provided the order of amino acids along the protein chain, and the myoglobin model was used to estimate the orientation of each of their side chains. The 1965 model was thus based in complementary empirical results from multiple sources.

Here are two examples of Perutz drawing inferences from this model:

> Strong solutions of neutral electrolytes are known to weaken polar interactions and to strengthen non-polar ones. They would therefore be expected to weaken the contacts between like subunits and to strengthen those between unlike ones. This immediately explains why only symmetrical dissociation is observed; it requires splitting of only one of the pairs of non-polar contacts and of two polar contacts, whereas asymmetrical dissociation would require both pairs of non-polar and no polar contacts to be split (1965, pp. 664-665).

> It has been shown in Results section (e) above that the lattice forces are made up mainly of weak polar bonds. Accordingly, the temperature coefficient of the solubility of horse oxyhaemoglobin in water or in dilute salt solution should be positive. This is in agreement with observation (1965, p. 666).

In each case, Perutz is holding consequences of his model up against known empirical results. The first result concerns treatment of prepared samples of haemoglobin with strong salt solution. This splits the molecule from its four-subunit (or tetramer) form, consisting of two "α chains" and two "β chains," into two symmetrical pairs of αβ subunits. Perutz shows how exposure to strong salts would produce this

outcome by reference to the effect of concentrated salts on the types of interactions between amino acid residues that, according to the model, bind each of the molecule's subunits together. Since salts weaken polar interactions, and these are the interactions that bind the two α chains and two β chains together, the molecule splits into symmetrical αβ pairs, rather than asymmetrical αα and ββ pairs. The second result concerns the increased solubility of haemoglobin in water upon an increase in temperature. Again, Perutz reasons from properties of the molecular structure in the model: neighboring molecules in the crystal lattice appear to have packed together such that their contacts points involve only a few atoms likely held together by weak polar bonds. An increase in thermal motion will increase the likelihood of breaking these bonds, separating the individual molecules and increasing the bulk solubility.

Both examples involve reasoning about features of the model in relation to specific interventions on bulk preparations of haemoglobin. Perutz moves nimbly from these higher-scale treatments to their molecular-scale consequences (weakened interactions, increased thermal motions) and back to the observed outcome. In doing so, he integrates the model into his knowledge of these treatments and spells out their consequences. Because these are known effects achieved by simple methods, the model application here serves primarily to bolster confidence in the model. Were they not known, however, the model would provide novel action-guiding inferences to a chemist seeking to split or dissolve a sample of haemoglobin. Perutz was warranted in applying this model to chemical techniques, despite its basis in X-ray data, for historical reasons: the common preparation methods of haemoglobin prior to crystallization and the long-standing identification of blood crystals with uncrystallized haemoglobin provided warrant for the belief that Perutz's X-ray-based model could represent the behavior of the constituents of bulk preparations.

6.6.2. False, fictitious, and unverified targets

How do I account for models of systems for which there was no empirical test (e.g., gravitational waves); models for which there was no real-world target (e.g., Thomson plum pudding); or models that

scientists themselves regarded as fictitious (e.g., infinite populations assumed in the proof the Hardy-Weinberg equilibrium)? The latter two cases pose no challenge for my view. To claim these models are targetless or cannot be empirically scrutinized because they include erroneous or falsifying assumptions confuses properties of the assumptions in the model—their inaccuracy or incompatibility with accepted theoretical or empirical claims—with the target scientists sought to represent. Thomson sought to reason about the structure of the atom; Hardy and Weinberg sought to reason about the frequencies of alleles within a single population; in each case researchers had grounds for thinking *these* targets existed and were characterized by some set of empirical results. How, then, did the models help?

Frigg and Nguyen (2017) describe targetless models (their example is a model of phlogiston), as a puzzle for scientific representation: how are scientists able to reason about nonexistent targets? This can be a problem for views of representation as a substantive and naturalistic relation. Authors of these views (e.g., (French & Ladyman, 1999)) who sought to define scientific representation without reference to the goals of an intentional agent have struggled to account for inaccurate representation on this basis. There is no such problem on my account. The fate of Laue's model of X-ray diffraction provides an example. This model was judged less accurate than Bragg's model due to its inadequacies when compared to the result produced by Friedrich and Knipping. There were dissimilarities in the pattern it predicted and part of Laue's method for generating this prediction was not physically justified. Today, scientists do not claim X-ray diffraction is produced by secondary radiation from crystal resonators and we would speak of this model as "targetless." None of this prevented Laue from being able to use it as a tool for licensing inferences and simulating data patterns.

In the case of infinite populations, the mistake here is again in the reification of models along the lines prescribed by Semantic Imperialism. That is, the puzzle of fictitious models is a result of philosophers regarding these models as literal descriptions of the state of affairs that would obtain were the model true. This is not a problem for my view. On my account, the integration of models with a theory of technique is piecemeal; only certain assumptions or features of the model are chosen for modeling the actual causal processes producing data. Further, the model is not taken to stand in a one-to-one relation to this process.

That is, not every element or state in the model needs to be interpreted in terms of the empirical program. All that is required from an interpretation is that it be sufficient to permit of model-entry and -exit moves. Idealizations like infinite populations are techniques for improving the internal inference rules of models; they make powerful reasoning tools like calculus available. Including them as assumptions in models may carry implications for their appropriate domains of application; scientists may need to show that the populations to which these models are being applied are sufficiently large, for example. But none of this requires that all assumptions in a model be taken as literally true.

The case of gravitational waves is different, as there were no means of empirically checking inferences from these models for decades. But the claim that they are nevertheless a source of informative inferences about a target requires clarification: informative inferences drawn from a model to *what*? In Chapter 1, I described how the view of scientific representation as an abstract, inference-enabling relation presupposes that a model-independent characterization of a target phenomenon is available. This is not always true for scientific inquiry—gravitational waves being a case in point—and scientists are not always warranted in treating such models as informative with respect to any real-world target. There was substantial debate over whether Einstein's models were physically plausible, exacerbated when some of his "waves" were found to depend on a choice of coordinate system (Kennefick, 2007).[160] If a model-independent characterization of a target is not available, nor any form of empirical scrutiny, then it is unclear where inferences drawn from these models are landing. If scientific representation is to be distinguished from stipulative fiat made with reference to an obscure target, then we ought to maintain a more restrictive view of what counts as a representational (as opposed to merely speculative or theoretical) use of a model.

[160] A similar story can be told about the eventual acceptance of black holes as physical phenomena.

6.6.3. Empiricism, realism, and pragmatism in scientific representation

I have argued that the notion of pragmatic representational content does not presuppose any form of denotation relation between model and target features. Instead, a model integrated with a theory of technique functions as a vehicle for licensing inferences within a data-gathering practice. Suárez's inferentialism, the best-known use-based account of scientific representation, is explicitly quietist. He claims a general account of scientific representation can only provide some minimal necessary conditions and demurs from giving an explanation of surrogative reasoning. Contessa modifies Suárez's approach in order to give a non-quietist account, but in doing so assumes a denotative mapping between model and target and so does not avoid the Paradox of Scientific Representation.

My account provides an explanation of the ability to reason about targets using scientific models that dissolves the paradox. Like Suárez, I account for scientific representation in terms of the inferences that scientists draw from model to target. I explain how this is done in terms of the activities by which scientists draw together model-based inferences and data gathering practices. In this chapter, I have provided this explanation in terms of a model's integration with a theory of technique. I have shown how this model can account for the accuracy of scientific models and provided a corresponding notion of pragmatic representational content that bypasses appeal to a mapping between model and target features by linking a model to the inferential framework of a theory of technique.

This account offloads the question of the model-world relation raised by the philosophy of scientific representation to the contents of empirical programs. Does this work? Or does a theory of technique also presuppose a form of model-based representation? It is useful to address this question by way of comparison to van Fraassen's (2008) account of the model-world relation. Van Fraassen imagines a scenario where a group of scientists present a graph to an audience based on measures of deer population growth within the region. They show how their theoretical model predicts a structure that maps onto this graph and use this to claim that the model represents the deer population growth. An audience member then asks, "But how do we know that your model really represents the deer population growth?" This appears to be a matter of

understanding that the data structure is there for the "right reasons," an issue raised in response to homomorphic data structures in Chapter 1. In response, van Fraassen suggests scientists describe how they measured the parameters that their theory treats as relevant to population growth, used this theory to infer further parameters that could not be measured, and so on to construct their data model. But he imagines a metaphysically-inclined member of the audience who is still unsatisfied: "So what then is the relation between the data model and the phenomenon?" This is akin to the issue raised for my view.

Here is van Fraassen's response:

> The answer has to be that the data models represent the phenomena; but why does that not just push the problem one step back? The short answer is this: construction of a data model is precisely the selective relevant depiction of the phenomena *by the user of the theory* required for the possibility of representation of the phenomenon (2008, p. 253).

> Since this is *my* representation of the deer population growth, there is *for me* no difference between the question whether T fits the graph and the question whether T fits the deer population growth (2008, p. 256).

In cases of concrete application, in which an adequate mapping between model structure and data model is achieved, van Fraassen claims that there is no difference, for the user of the model, between the data extracted from a phenomenon and the phenomenon itself. For van Fraassen, this is a matter of the pragmatics of theory application. Scientific knowledge, as presented by theories, describe the world in terms of abstract structures. Applying that knowledge is akin to locating oneself on a map; it is an indexical judgment, the correctness of which depends on the use to which the structure is being put. Construction of a data model is analogous to the self-attribution of position with respect to a map—a way of locating oneself in the logical space of the theory—without which, he claims, it would be impossible to use representations at all. Thus, the model user must equate the target phenomenon with its data model or fall into a pragmatic contradiction: "What I cannot do is to both present the graph as representing something and say that perhaps it doesn't represent that at all." (2008, p. 257).

Contessa (2009) has responded that a researcher is still aware of the notion that there is a target or fact of the matter (i.e., the actual deer population) beyond the data they have gathered. Nguyen (2016) has argued against the underlying assumption that a researcher must believe that a representation is true to their

target in order to treat it as a representation of that target. By all counts van Fraassen is aware of this (Cf. (2010)). His concern, I take it, is specifically with how one can make any objective *claim* with respect to a "fact of the matter" beyond a mapping between model and data. If the application of scientific theories by using models is the way that we make such claims, then the problem stands. Van Fraassen, wary of weighing down scientific knowledge with the ontological bloat of correspondence, solves this by claiming a model user pragmatically equates data and phenomenon.

This is a problem of van Fraassen's own making. It is rooted in three key assumptions of his constructive empiricism: that scientific knowledge consists in the abstract structures presented by theories, that scientific representation is a matter of mapping empirical data into substructures of these theories, and that positing anything beyond these data to account for the aims of science is unwarranted metaphysics. He therefore claims the ability to measure and construct a data model from a target is akin to knowing an indexical proposition—a highly circumstantial matter that is outside the scope of general scientific knowledge. I disagree.[161] A substantial amount of scientific knowledge is devoted to ensuring that data is reliably connected to a target phenomenon. Though it requires some sensitivity to local contingencies, this knowledge is not reducible to indexical propositions; much of it consists in generalizations describing dependency relations between causal factors that recur across data gathering contexts, which I have called the theory of technique.

Though theories of technique can be overlaid with model-based reasoning, the introduction of such models is a downstream result. Theories of technique originate with the activities of cataloging the experimental conditions and causal factors that recur across data-gathering contexts, including the consequences of interventions on these factors. This can be seen in the early practice of blood crystallization or X-ray experiments. These practices did not depend on a representational model, in the sense that I have been using this term. There was no self-contained structure with internal inference rules that captured the entirety of the experimental setup and was manipulated as something over and above the experiment itself.

[161] See also Pincock (2011) on this point.

Practitioners did possess a specialized vocabulary for causes within this setup, but these were tied to empirically accessible features of their instruments and environments, and an account of the dependency relations between them. The import of these theories lay in their ability to license practical inferences, that is, to provide warrant to act with respect to these features. The actions warranted by a theory of technique, the processes of data production and interpretation that they comprise in sequence, forge the connection between data and phenomenon that Van Fraassen reduces to an identity. The introduction of more rigorous models of the experiment into a theory of technique, as seen in the expanding reflection formulae of X-ray crystallography, takes place only after practitioners have surveyed the parameter space of the experimental conditions, settled on stable procedures, and isolated an interaction within an encompassing set of factors and dependencies that is susceptible to more precise mathematical description. In my view, these preconditions are foundational contributions to scientific knowledge.

Like Van Fraassen, I hold that a representational model cannot, on its own, provide a direct account of its target beyond the comparison of derived results to data. These comparisons are the basis for the judgment of a model's accuracy. In addition, I claim that models inform the practice of data gathering through their integration with theories of technique. The historical reliability of these practices, the contribution they make to target characterization through the production of empirical results, and the coherence of the model in licensing practical inferences in this domain are further constituents of the ability to represent a target with a model. On this view, scientific representation is grounded in two basic features: the epistemic evaluation of models with respect to data and the pragmatic, action-guiding role they play with respect to data gathering practice. Model-based target characterization results from this joint representational role.

We can distinguish both Van Fraassen's empiricist account of representation and the realist assumptions he opposes from my pragmatist account. Van Fraassen's commitment to empiricism motivates him to restrict the representation relation to one that exists between a model and observational data. His equation of data and phenomenon is a consequence of this. Realist views instead posit a correspondence between model and target that goes beyond its relation to data. The problems of these views were the focus

of Chapter 1. Both stances mirror the standpoint of traditional epistemology, with the contents of scientific theories functioning as the stand-in for the ideas of the individual knower, observable data as sensory impressions, and unobservables as external reality. Like these epistemologies, a great deal of emphasis is placed on boundary between what is observable and what is not; and like these epistemologies, they diminish the role of practical activity as a constituent of scientific knowledge. Pragmatists, on the other hand, have traditionally preferred to replace talk of a fixed body of knowledge with a notion of inquiry as a self-correcting process. They begin from a fundamentally interactive standpoint[162] and interpret traditional philosophical dichotomies as reified instances of interactive processes. The view I have laid out here proposes something similar for the concept of scientific representation. I take scientific representation to be comprised of activities integrating model-based inferences with practices for gathering data from a target, where both sides of this dichotomy are subject to reciprocal influence. Modeling efforts are constrained and incentivized by empirical results, target characterizations are embellished and refined through the application of models.

This is different from what I am tempted to call light pragmatist views, such as those of Giere, Suárez, and perhaps Contessa or Weinberg. These views speak of representation as an activity, but tend to restrict their analyses to the cognitive domain: scientific representation is understood in terms of the inferences, interpretations, and aims of the scientists who use models. While these are important elements of scientific modeling, they follow much of the literature in promoting a focus on one side of the representation relation: models rather than targets. My account begins from the view that the practice of generating data from a target should play an equally important role in explaining model-based reasoning. I arrive at an account that emphasizes practical (and not just cognitive) activities for this reason. On this account, scientific representation is not entirely "in the head" of the user nor is it a relation "out there" between model and target, which can be defined in terms that are independent from empirical activities.

[162] Cf. Dewey: "Experience is primarily a process of undergoing […] Undergoing, however, is never mere passivity. The most patient patient is more than a receptor. He is also an agent—a reactor, one trying experiments, one concerned with undergoing in a way which may influence what is still to happen" (1980, pp. 8-9).

Rather, representation is constituted through broader scientific practices of model construction, reception, and evaluation in ongoing relation to empirical programs; whatever can be said of it qua relation between model and target is given through these practices. Because of this interplay with data-gathering practice, the characterizations of target phenomena are inferentially bound up with the models used to represent them. This does not make them less "real," it ties their reality to the new ways of successfully interacting with the world enabled by these models.[163]

References

Bernal, J. D., Fankuchen, I., & Perutz, M. (1938). An X-ray Study of Chymotrypsin and Haemoglobin. *Nature, 141*, 523-524.

Boesch, B. (2017a). There is a Special Problem of Scientific Representation. *Philosophy of Science, 84,* 970-981.

Brandom, R. (2010). Inferentialism and some of its Challenges. In B. Weiss, & J. Wanderer (Eds.), *Reading Brandom On Making It Explicit* (pp. 159-180). New York: Routledge.

Chang, H. (2022). *Realism for Realistic People.* New York: Cambridge University Press.

Contessa, G. (2007). Scientific Representation, Interpretation, and Surrogative Reasoning. *Philosophy of Science, 74*, 46-68.

Contessa, G. (2009). *Scientific Representation: Perspectives of Paradox*. Retrieved from Notre Dame Philosophical Reviews: https://ndpr.nd.edu/reviews/scientific-representation-paradoxes-of-perspective/

Dewey, J. (1980). The Need for a Recovery of Philosophy. In J. Boydston (Ed.), *The Middle Works of John Dewey, Vol. 10.* Carbondale: Southern Illinois University Press.

Fischer, E. (1907). Die Chemie der Proteine und ihre Beziehungen zur Biologie. In E. Fischer, & M. Bergmann (Ed.), *Untersuchungen über Aminosäuren, Polypeptide, und Proteine II* (pp. 1-21). Berlin: Springer.

French, S., & Ladyman, J. (1999). Reinflating the semantic approach. *International Studies in the Philosophy of Science, 13*, 103-121.

Frigg, R., & Nguyen, J. (2017). Models and Representation. In L. Magnani, & T. Bertolotti (Eds.), *Springer Handbook of Model-Based Science* (pp. 49-102). New York: Springer.

[163] I consider the view of scientific representation I advocate here to be broadly compatible with the approach to scientific realism advocated by Mitchell (forthcoming 2023), Chang (2022), and Woodward (2003).

Kennefick, D. (2007). *Traveling at the Speed of Thought.* Princeton: Princeton University Press.

Khalifa, K., Millson, J., & Risjord, M. (2022). Scientific Representation: An Inferentialist-Expressivist Manifesto. *Manuscript*.

Lehmann, C. G. (1851). *Physiological Chemistry, Vol. 1.* (G. Day, Trans.) London: Harrison and Son.

Mitchell, S. (2003). *Biological Complexity and Integrative Pluralism.* Princeton: Princeton University Press.

Mitchell, S. (2019). Perspectives, Representation, and Integration. In M. Massimi, & C. McCoy (Eds.), *Understanding Perspectivism* (pp. 178-193). New York: Routledge.

Mitchell, S. (forthcoming 2023). The Bearable Thinness of Being: A Pragmatist Metaphysics of Affordances. In S. Mitchell, & H. Andersen (Eds.), *The Pragmatist Challenge* (pp. 112-142). New York: Oxford University Press.

Mitchell, S., & Gronenborn, A. (2017). After Fifty Years, Why Are Protein X-ray Crystallographers Still in Business? *Philosophy of Science, 68*, 703-723.

Nguyen, J. (2016). On the Pragmatic Equivalence Between Representing Data and Phenomena. *83*, 171-191.

Osborne, T. (1908). Our Present Knowledge of Proteins. *Science, 28*, 417-427.

Perutz, M. (1965). Structure and Function of Haemoglobin: I. A Tentative Atomic Model of Horse Oxyhaemoglobin. *Journal of Molecular Biology, 13*, 646-668.

Pickering, A. (1995). *The Mangle of Practice.* Chicago: University of Chicago Press.

Pincock, C. (2011). Bas van Fraassen *Scientific Representation: Paradoxes of Perspective*. *British Journal for the Philosophy of Science, 62*, 677-682.

Suárez, M. (2004). An Inferential Conception of Scientific Representation. *Philosophy of Science, 71*, 767-779.

Tal, E. (2017). Calibration: Modelling the Measurement Process. *Studies in the History and Philosophy of Science Part A, 65*, 33-45.

van Fraassen, B. (2008). *Scientific Representation: Paradoxes of Perspective.* New York: Oxford University Press.

van Fraassen, B. (2010). Reply to Contessa, Ghins, and Healey. *Analysis, 70*, 547-556.

Weisberg, M. (2013). *Simulation and Similarity: Using Models to Understand the World.* New York: Oxford University Press.

Woodward, J. (2003). Experimentation, Causal Inference, and Instrumental Realism. In H. Radder (Ed.), *The Philosophy of Scientific Experimentation* (pp. 87-118). Pittsburgh: University of Pittsburgh Press.

Appendix A: Remarks on the Pragmatics of Measurement and Cross-checking

Here I present some simple formal results that are intended to illustrate some basic considerations involved in a researcher's decision-making. Elements of this formalism may be developed into a more powerful framework; for now, I will consider the simple case of deciding whether to perform a cross-check at a given stage of a procedure.

Suppose a researcher has carried out a procedure up to a particular step, partway between their starting point and an end point. The procedure is conceived as successive transformations of the experimental system's state. At the end point they aim to have reached a goal state. We will examine the pragmatic value of choosing to make a cross-check measurement at this stage in the procedure. Call c_1 the cumulative cost in time and resources of the steps taken up to this point, c_2 the cumulative cost of the further steps required until reaching the end point, and c_m the cost of carrying out a particular cross-checking measurement at this step. When carried out, the measurement will either output *Y*, telling the researcher that they are on track to reach their goal, or *N* for the contrary, with some probability of error. It is assumed that the researcher will stop and start over from the beginning if their cross-check outputs *N*.

In this case, the cross-check is worth carrying out when the following general condition holds:

$$\frac{c_m + (c_1 + c_2) * P(Y) + c_1 * (1 - P(Y))}{P(Y) * P(goal|Y)} < \frac{c_1 + c_2}{P(goal|{\sim}M)}$$

where *goal* is the event in which the researcher successfully reaches their goal, *M* is the event of making the cross-checking measurement and continuing the procedure, and ~*M* not making the measurement. On the left hand, we take the cumulative cost of the measurement and the steps to completion (if the measure outputs *Y*) and the steps already taken (if the measure outputs *N*) and divide these by the probability that the measurement outputs *Y* and the goal is reached. On the right, we take the cumulative cost of the steps to completion divided by the probability that the goal will be reached without a measurement being taken. These can be understood as the "cost per success" of carrying out the procedure with a measurement at this stage, or not.

Some algebra simplifies the above condition to:

$$\frac{c_m + c_1 + c_2 * P(Y)}{P(Y) * P(goal|Y)} < \frac{c_1 + c_2}{P(goal|{\sim}M)}$$

Note that the generality of the presentation makes it adaptable to different levels within a procedural hierarchy by simply relocating the beginning and end points that define the relevant terms.

We now consider the following assumptions:

A) *Non-alteration:* the measurement does not alter the state of the system, i.e.:

$$P(goal|{\sim}M) = P(goal|M) = P(Y) * P(goal|Y) + P(N) * P(goal|N)$$

B) *Error-free*: the measurement has no possibility of error, i.e.:

$$P(goal|N) = 0$$

C) *Cost-free*: there is no cost to performing the measurement, i.e.:

$$c_m = 0$$

These different assumptions will almost never be perfectly realized in any actual procedure, although they may be approximated in certain cases:

- *Non-alteration* may be approximately true in instances like those considered in the case study above where one takes a very small sample of the system and performs the cross-check on *this*, then the rest of the system is only affected insofar as a portion of it was removed. If the cross-check can be performed on a small enough portion, then the system is practically unaffected by the measurement. Generally, an empirical program can be drawn on to ensure that such measurements take place under conditions that limit the altering effects.
- *Cost-free* will be approximated in cases where the check is very cheap to perform. A simple example from this chapter is that of checking the optical density of a bacterial growth sample—a measurement that can be performed very quickly with minimal equipment.
- *Error-free* will be approximated in cases where a measuring apparatus is extremely well-calibrated and highly unlikely to produce false negatives or false positives. High precision measurements or

an apparatus whose resolution is much sharper than the value at which error becomes intolerable may approximate this case.

Several consequences of these assumptions will be shown. *Non-alteration* alone is not particularly illuminating to assume without any knowledge of how error-prone the measurement is. If we assume *error-free*, however, then the general condition can be reduced to:

$$\frac{c_m + c_1 + c_2 * P(Y)}{c_1 + c_2} < H = \frac{P(goal|M)}{P(goal|{\sim}M)}$$

If $0 < H < 1$, then it measures the extent to which the mere act of making the measurement (and otherwise continuing the procedure) hinders the researcher from reaching their goal. When *H* is close to 0, the check is overly inhibitive and not worth performing. $H > 1$ would be the somewhat unrealistic event that the measurement has some intrinsic corrective effect, i.e., helps in reaching the goal independent of any decisions made on its basis. If *H* is at some intermediate value between 0 and 1, the decision of performing a check becomes beneficial as (i) the cost of the measurement and the probability that it will output *Y* go down and (ii) the cost of completing the procedure from this step goes up. That is, if the measurement is relatively cheap and there is a high chance that the researcher has gone off track (equivalently, *P(Y)* is more informative), then a check may be worthwhile even if it has some hindering effect on reaching the goal, and especially so if the cost of completing the procedure is large. We can imagine a researcher in this position if they have completed a series of alterations of their system without any prior checks and are now facing a much more expensive series of final steps. In this case, an affordable check is worth doing as long as it doesn't hinder the researcher too severely.

When $H = 1$, then *non-alteration* holds as well and the check will be worth performing as long as

$$\frac{c_m}{c_2} < 1 - P(Y)$$

The right-hand side of the above equation becomes smaller as it becomes more probable that the measurement will indicate that the researcher is on the right track. If $P(Y) = 1$, then the researcher is guaranteed by this situation to be on the right track and there is no point in confirming this with a measurement of any cost. On the other hand, if *P(Y)* is small, then the measurement is more informative

and will be worth carrying out as long as it is not very costly relative to the cost of the ensuing steps. This makes sense: if there is a good chance that one could be on the wrong path to their goal, if continuing on this path will cost a lot, and if checking one's progress is cheap by comparison, then one should check. In general, a relatively expensive measurement will only be worth doing if it is very probable that prior steps have prevented the researcher away from reaching their goal (note that if one also assumes C, then it always makes sense to perform a cross-check: a cost-free, error-free, non-altering check is always worth doing).

Cost-free is again not especially illuminating to assume without any knowledge of how error-prone the measurement is, but when taken with *error-free* and the fairly uncontroversial assumption that $H < 1$, we arrive at the condition:

$$\frac{c_1}{c_2} < \frac{H - P(Y)}{1 - H}$$

Once again, if the measurement has a large hindering effect (H close to 0), then it is unlikely to be worth performing. Even if it has *some* hindering effect, it may be worth carrying out if there's a high chance that the researcher is not on the path to their goal (i.e., *P(Y)* is small) and/or the cumulative cost of the procedure is much larger than the costs already incurred (i.e., c_2 is much larger than c_1).

If we assume either that other costs are negligible compared to time spent, or that the cost of each step is roughly the same (in which case the cumulative cost of the procedure increases linearly), then c_1/c_2 becomes an approximate measure of how far the researcher has proceeded from the start to the goal. In this case the above inequality tells us that a cost-free, error-free check is usually worth performing at very early stages of a procedure as long as it is a little informative and not a complete hindrance, and usually *not* worth performing at very late stages unless it is extremely informative and a very small hindrance.

The quantity c_1/c_2 also arises when we assume *non-alteration* and *cost-free*, but allow for measurement error. In this case we get the condition:

$$\frac{c_1}{c_2} < \frac{P(Y) * [P(goal|Y) - P(goal|M)]}{P(N) * P(goal|N)}$$

This is a fairly unwieldy expression, but generally it tells us that a check is more likely to be worth doing when the probability of a false negative is low, the probability of a true positive is high, the probability that the measurement will tell the researcher they're on the right path is high, and/or the largest share of the costs for the procedure are yet to come. Much of this boils down to the obvious idea that, if one is performing a check that is not error-free, it is more worthwhile when errors are unlikely. If such a check is error-prone, this tells us that it is only worth doing before a large share of the costs of the procedure have been incurred, or at the earliest stages of an experiment.

Given that more cross-check measurements are generally desirable due to their epistemic merits (as discussed above), we can rephrase this analysis in terms of the following general heuristics:

1) Reduce, to whatever extent possible, the errors, costs, and altering effects of your measurements
2) Measurements have more pragmatic value when the probability of prior error is high
3) Measurements have more value when the cost of completion is high relative to incurred costs
4) Measurements have more value when their own cost is low relative to the cost of completion

The decision to perform a cross-checking measurement can thus be viewed in ideal circumstances as an optimization task with respect to these guidelines. This makes evident the pragmatic benefits of possessing a thorough theory of technique, as followers of heuristic (1) are greatly aided by an understanding of the system of interest, the instrument, and their interaction.

Appendix B: Two Additional Developmental Tendencies of Empirical Programs

Integration of techniques: Consistent data outcomes with low error rates are necessary for reliability. Crystallographers gradually reached a stage where they could regularly produce pure samples of haemoglobin as a reliable basis for further study, results that I described as points of stability. Blood crystallization procedures were incorporated into a variety of extended procedures: crystals were subject to chemical analyses, crystallographic microscopy, spectroscopy, X-ray diffraction. Each technique produces different forms of data with characteristic sources of error. However, since they have a common preparation, they are in principle comparable. Researchers intervening on the same object by different means, who can account for distinctive artifacts of each, may use inferences about phenomena drawn from one set of data to reinforce or cast doubt on inferences drawn from another set of data produced from the same preparation. Astbury's interpretation of his early fiber diffraction patterns in terms of recurring backbones and side-chains was partly informed by stereochemical results; Svedberg's sedimentation studies increased Bernal and Crowfoot's confidence that their diffraction patterns of insulin betrayed a large crystalline molecule. Section 3 discussed several examples in which results derived from one technique informed second-order concerns for another, leading to firmer convictions in some cases and altered procedures or revised interpretations of data in others. These stable through lines contribute to the heavily integrated character of laboratory work in fields like molecular biology (Cf. Knorr-Cetina 1999). This also helps explain how general theoretical models of phenomena may be subject to practices of piecemeal, localized integration when applied to concrete cases.

Procedural and epistemic sub-division: As procedures are scrutinized and honed to produce reliable results, they may undergo an increasing series of sub-divisions. Chapter 3 presented the procedures underlying an empirical program as segmented and hierarchically nested. Points of stability produced at the end of one stage in a procedure provide a stable input to a following stage. This allows for the incorporation of smaller procedures into larger ones and the integration of techniques, but it may also

leave distinct stages relatively insulated from one another, as reflected by an increased division of labor and degree of specialization among practitioners working at different stages of an overarching procedure. One example mentioned in Chapter 4 was the gradual separation of crystallization practices from the study of these crystals. This separation was extremely rare in the 19th century, when leading physiological chemists devoted substantial time to crystallization technique. But interest grew in the 20th century in studying a larger number of proteins crystals. Obtaining adequate crystals for diffraction studies was time-consuming and called for a special repertoire of skills, so much so that it became the role of individual members of a lab. Hence Perutz's lab brought in a specialist in heavy atom derivatives, Howard Dinitz, whose sole responsibility was crystallizing new forms of myoglobin and haemoglobin.

The production of larger quantities of data and the need to make them increasingly precise were further drivers of sub-division in Perutz's lab. The discussion of data D_1, …, D_m interpreted with respect to a range of claims about phenomena P_1, …, P_n belies the layers of processing involved in a program like protein X-ray crystallography. Consider the series of steps involved in constructing the first molecular model haemoglobin:

- One group produced diffraction data from haemoglobin and six heavy atom derivatives, collecting some 40,000 reflections.
- A second team of assistants measured the intensities of the reflection spots on hand-operated line-scanning densitometers. These output line graph readings, whose peak heights were measured by hand.
- The strongest reflections were given to Tony North, who made absolute intensity measurements with an Automatic Linear Diffractometer.
- All of this was then fed as raw numerical data into Kendrew's EDSAC II program for an initial Fourier analysis, with phase information underdetermined.
- Michael Rossmann was then tasked with designing a further EDSAC II program for determining and refining the heavy atom parameters using a least-squares method.

- These results were then passed to Hilary Muirhead, a PhD student, who used a further program to implement Blow and Crick's method of phase determination.
- Finally, a series of density contour maps were drawn, corresponding to parallel planar slices of the crystal's unit cell.

X-ray crystallography had come a long way from Bragg's early trial-and-error methods. Each of these sub-tasks was prone to its own sources of error, knowledge of which could not be easily and efficiently stored in one head. As the X-ray program became further extended, the knowledge embodied in this program was distributed across a larger number of practitioners. Even when a single practitioner carries out an extended procedure (as with Andrew in Chapter 3), they cannot be expected to possess deep knowledge of each piece of equipment or technique they draw on. While Andrew knew how to operate the AKTA purification machine and the effects it should have on his sample, persistent problems with its operation would require the help of a lab technician with more thorough understanding of the instrument. Thus, as empirical programs become segmented into further sub-sequences, incorporating a variety of different techniques, they are likely to see an increasing division of labor in their procedural or epistemic structure.

Bibliography

Abraham, E., & Robinson, R. (1937). Crystallization of Lysozyme. *Nature, 140*, 24.

Ackermann, R. (1985). *Data, Instruments, and Theory.* Princeton: Princeton University Press.

Alison, J. (2012). *The Road to Discovery: Detector Alignment, Electron Identification, Particle Misidentifcation, WW Physics, and the Discovery of the Higgs Boson.* Dissertation: University of Pennsylvania.

Allen, H. (1947). Charles Glover Barkla, 1877-1944. *Obituary Notices of Fellows of the Royal Society, 5*, 341-366.

Anstey, P. (2014). Philosophy of Experiment in Early Modern England: The Case of Bacon, Boyle, and Hooke. *Early Science and Medicine, 19*, 103-132.

Astbury, W. (1922). The Crystalline Structure and Properties of Tartaric Acid. *Proceedings of the Royal Society, A, 102*, 506-528.

Astbury, W. (1929). A New Integrating Photometer for X-ray Crystal Reflections, etc. *Proceedings of the Royal Society, A, 123*, 575-602.

Astbury, W. (1939). X-ray Studies of the Structure of Compounds of Biological Interest. *Annual Review of Biochemistry, 8*, 113-133.

Astbury, W., & Lomax, R. (1934). X-ray Photographs of Crystalline Pepsin. *Nature, 133*, 795.

Astbury, W., & Street, A. (1932). X-Ray Studies of the Structure of Hair, Wool, and Related Fibres. I. General. *Philosophical Transaction of the Royal Society of London, A, 230*, 75-101.

Astubry, W., & Bragg, W. H. (1925). Annual Reports on the Progress of Chemistry. In J. M. Bijvoet, W. G. Burgers, & G. Hägg (Eds.), *Early Papers on Diffraction of X-rays by Crystals, Vol. 2* (Vol. 20, p. 233). Utrecht: A. Oosthoek.

Austin, J. L. (1962). *How to Do Things With Words.* London: Oxford University Press.

Aydinonat, N. (2007). Models, conjectures and exploration: an analysis of Schelling's checkerboard model of residential segregation. *Journal of Economic Methodology, 14*, 429-454.

Bach, K. (2005). Context ex Machina. In Z. G. Szabo (Ed.), *Semantics Versus Pragmatics* (pp. 15-44). New York: Oxford University Press.

Baird, D., & Faust, T. (1990). Scientific Instruments, Scientific Progress and the Cyclotron. *The British Journal for the Philosophy of Science, 41*, 147-175.

Barkla, C. (1905). Polarised Röntgen Radiation. *Philosophical Transactions of the Royal Society, A, 204*, 467-479.

Barkla, C. (1967). Characteristic Röntgen Radiation. In *Nobel Lectures, Physics 1901-1921* (pp. 393-399). Amsterdam: Elsevier.

Barkla, C., & Martyn, G. (1913). An X-ray Fringe System. *Nature, 90*, 647.

Barlow, W. (1883). Probable Nature of the Internal Symmetry of Crystals. *Nature, 29*, 186-188.

Barnes, B. (1977). *Interests and the Growth of Knowledge.* New York: Routledge.

Bellesis, A. (2019). *Shadowing.* University of Pittsburgh Structural Biology Laboratory.

Bennett, J., & Kendrew, J. (1952). The Computation of Fourier Syntheses with a Digital Electronic Calculating Machine. *Acta Crystallographica, 5*, 109-116.

Bernal, J. D. (1927). A universal X-ray photogoniometer. *Journal of Scientific Instruments, 4*, 273-284.

Bernal, J. D. (1930). The place of X-ray crystallography in the development of modern science. *Radiology, 15*, 2-12.

Bernal, J. D. (1939). Structure of Proteins. *Nature, 143*, 663-667.

Bernal, J. D., & Crowfoot, D. (1934). X-Ray Photographs of Crystalline Pepsin. *Nature, 133*, 794-795.

Bernal, J. D., Crowfoot, D., Evans, R., & Wells, A. (1935). Crystallography (1934-1935). *Annual Reports on the Progress of Chemistry, 32*, 183-242.

Bernal, J. D., Crowfoot, D., Robinson, B., & Wooster, W. (1933). Crystallography. *Annual Reports on the Progress of Chemistry, 30*, 360-430.

Bernal, J. D., Fankuchen, I., & Perutz, M. (1938). An X-ray Study of Chymotrypsin and Haemoglobin. *Nature, 141*, 523-524.

Berntson, A., Stojanoff, V., & Takai, H. (2003). Application of a neural network in high-throughput protein crystallography. *Journal of Synchrotron Radiation, 10*, 445-449.

Berzelius, M. (1813). Sur la composition des fluides animaux. *Annales de Chimie, 31*, 26-72.

Bickhard, M. (1999). Interaction and Representation. *Theory & Psychology, 9*, 435-458.

Binning, G., Rohrer, H., Gerber, C., & Weibel, E. (1982). Surface Studies by Scanning Tunneling Microscopy. *Physical Review Letters, 59*, 57-61.

Boesch, B. (2017a). There is a Special Problem of Scientific Representation. *Philosophy of Science, 84*, 970-981.

Boesch, B. (2017b). The Means-End Account of Scientific, Representational Actions. *Synthese*, published online.

Bogen, J., & Woodward, J. (1988). Saving the Phenomena. *The Philosophical Review, 97*, 303-352.

Bogen, J., & Woodward, J. (1992). Observations, theories, and the evolution of the human spirit. *Philosophy of Science, 59*, 590-611.

Bokulich, A. (2018). Using models to correct data: paleodiversity and the fossil record. *Synthese, Special Issue: Abstraction and Idealization in Scientific Modeling*, Accessed online: <https://link.springer.com/article/10.1007/s11229-018-1820-x>.

Bokulich, A. (2022). Toward a Taxonomy of the Model-Ladenness of Data. *Philosophy of Science, 87*, 793-806.

Boorsboom, D. (2005). *Measuring the Mind.* New York: Cambridge University Press.

Boumans, M., & Hon, G. (2014). Introduction. In M. Boumans, G. Hon, & A. Petersen (Eds.), *Error and Uncertainty in Scientific Practice* (pp. 1-12). London: Pickering & Chatto.

Bourdieu, P. (1977). *Outline of a Theory of Practice.* Cambridge MA: Cambridge University Press.

Boyd, N. (2018). Evidence Enriched. *Philosophy of Science, 85*, 403-421.

Boyes-Watson, J., Davidson, E., & Perutz, M. (1947). An X-ray study of horse methaemoglobin. I. *Proceedings of the Royal Society, A, 191*, 83-132.

Bragg, W. H. (1912). X-rays and Crystals. *Nature, 90*, 219.

Bragg, W. H. (1912). X-rays and Crystals. In J. M. Bijvoet, W. G. Burgers, & G. Hägg (Eds.), *Early Papers on Diffraction of X-rays by Crystals, Vol. 1* (pp. 27-29). Utrecht: International Union of Crystallography.

Bragg, W. H. (1914). The Intensity of Reflexion of X-rays by Crystals. In J. M. Bijvoet, W. G. Burgers, & G. Hägg (Eds.), *Early Papers on Diffraction of X-rays by Crystals, Vol. 1* (pp. 103-110; 160). Utrecht: A. Oosthoek.

Bragg, W. H. (1915). X-rays and Crystal Structure. In J. M. Bijvoet, W. G. Burgers, & G. Hägg (Eds.), *Early Papers on Diffraction of X-rays by Crystals, Vol. 1* (pp. 111-114). Utrecht: A. Oosthoek.

Bragg, W. H. (1921). The Intensity of X-ray Reflexion by Diamond. In J. M. Bijvoet, W. G. Burgers, & G. Hägg (Eds.), *Early Papers on Diffraction of X-rays by Crystals* (p. 303). Utrecht: A. Oosthoek.

Bragg, W. H., & Bragg, W. L. (1913). The Reflection of X-rays by Crystals. *Proceedings of the Royal Society, A, 88*, 428-438.

Bragg, W. H., & Bragg, W. L. (1915). *X Rays and Crystal Structure.* London: G. Bell and Sons.

Bragg, W. H., & Glasson, J. L. (1907). On a Want of Symmetry shown by Secondary X-rays. *Proceedings of the Physical Society of London, 21*, 735-745.

Bragg, W. L. (1912). The Specular Reflection of X-rays. *Nature, 90*, 410.

Bragg, W. L. (1913). The Diffraction of Short Electromagnetic Waves by a Crystal. *Proceedings of the Cambridge Philosophical Society, 17*, 43-57.

Bragg, W. L. (1914a). The Structure of Some Crystals as Indicated by their Diffraction of X-rays. In J. M. Bijvoet, W. G. Burgers, & G. Hägg (Eds.), *Early Papers on Diffraction of X-rays by Crystals, Vol. 1* (pp. 59-71). Utrecht: A. Oosthoek.

Bragg, W. L. (1914b). The Analysis of Crystals by the X-ray Spectrometer. *Proceedings of the Royal Society, A, 89*, 468-489.

Bragg, W. L. (1920). The Arrangement of Atoms in Crystals. In J. M. Bijvoet, W. G. Burgers, & G. Hägg (Eds.), *Early Papers on Diffraction of X-rays by Crystals, Vol. 2* (pp. 34-39). Utrecht: A. Oosthoek.

Bragg, W. L. (1925a). A note on the paper 'The Atomic Structure Factor in the Intensity of Reflextion of X-rays by Crystals'. In J. M. Bijvoet, W. G. Burgers, & G. Hägg (Eds.), *Early Papers on Diffraction of X-rays by Crystals, Vol. 1* (pp. 298-299). Utrecht: A. Oosthoek.

Bragg, W. L. (1925b). The Interpretation of Intensity Measurements in X-ray Analysis of Crystal Structure. In J. M. Bijvoet, W. G. Burgers, & G. Hägg (Eds.), *Early Papers on Diffraction of X-rays by Crystals, Vol. 1* (pp. 304-306). Utrecht: A. Oosthoek.

Bragg, W. L. (1929). A Note on the Representation of Crystal Structure by Fourier Series. In J. M. Bijvoet, W. G. Burgers, & G. Hägg (Eds.), *Early Papers on Diffraction of X-rays by Crystals, Vol. 2* (pp. 391-403). Utrecht: A. Oosthoek.

Bragg, W. L. (1965). First stages in the x-ray analysis of proteins. *Reports on Progress in Phsyics, 28*, 1-14.

Bragg, W. L. (1975). *The Development of X-ray Analysis.* New York: Hafner Press.

Bragg, W. L., & Perutz, M. (1952). The structure of haemoglobin. *Proceedings of the Royal Society, A, 213*, 425-435.

Bragg, W. L., & West, J. (1928). A Technique for the X-ray Examination of Crystal Structures with many Parameters. In J. M. Bijvoet, W. G. Burgers, & G. Hägg (Eds.), *Early Papers on Diffraction of X-rays by Crystals, Vol. 2* (Vol. 69, pp. 383-385). Utrecht: A. Oosthoek.

Bragg, W. L., Darwin, C., & James, R. (1926). The Intensity of Reflexion of X-rays by Crystals. In J. M. Bijvoet, W. G. Burgers, & G. Hägg (Eds.), *Early Papers on Diffraction of X-rays by Crystals, Vol. 1* (Vol. 1, pp. 306-313). Utrecht: A. Oosthoek.

Bragg, W. L., Howells, E. R., & Perutz, M. (1952). Arrangement of Polypeptide Chains in Horse Methaemoglobin. *Acta Crystallographica, 5*, 136-141.

Bragg, W. L., Howells, E., & Perutz, M. (1954). The structure of haemoglobin. II. *Proceedings of the Royal Society, A, 222*, 33-44.

Bragg, W. L., James, R. W., & Bosanquet, C. H. (1921). The Intensity of Reflexion of X-rays by Rock-Salt. In J. M. Bijvoet, W. G. Burgers, & G. Hägg (Eds.), *Early Papers on Diffraction of X-rays by Crystals, Vol. 1* (Vol. 41, pp. 265-291). Utrecht: A. Oosthoek.

Bragg, W. L., James, R. W., & Bosanquet, C. H. (1922). The Distribution of Electrons around the Nucleus in the Sodium and Chlorine Atoms. *Philosophical Magazine, 44*, 433-449.

Brandom, R. (1994). *Making It Explicit.* Cambridge, MA: Harvard University Press.

Brandom, R. (2000). *Articulating Reasons: An Introduction to Inferentialism.* Cambridge MA: Harvard University Press.

Brandom, R. (2010). Inferentialism and some of its Challenges. In B. Weiss, & J. Wanderer (Eds.), *Reading Brandom On Making It Explicit* (pp. 159-180). New York: Routledge.

Bravais, A. (1866). *Etudes cristollagraphiques.* Paris: Gauthier-Villars.

Brewster, D. (1831). *Optics.* London: Longman, Rees, Orme, Brown, and Green.

Brigandt, I. (2010). Scientific Reasoning is Material Inference: Combining Confirmation, Discovery, and Explanation. *International Studies in the Philosophy of Science, 24*, 31-43.

Brown, A. (1864). On the Theory of Isomeric Compounds. *Journal of the Chemical Society of London, 18*, 230-245.

Bueno, O. (1997). Empirical Adequacy: A Partial Structures Approach. *Studies in the History and Philosophy of Science, 28*, 585-610.

Bueno, O., & Colyvan. (2011). An Inferential Conception of the Application of Mathematics. *Nous, 45*, 345-374.

Bueno, O., & French, S. (2011). How Theories Represent. *British Journal for the Philosophy of Science, 62*, 857-894.

Bueno, O., & French, S. (2018). *Applying Mathematics.* New York: Oxford University Press.

Bueno, O., French, S., & Ladyman, J. (2012). Models and structures: Phenomenological and partial. *Studies in History and Philosophy of Modern Physics, 43*, 43-46.

Buerger, M. (1942). *X-Ray Crystallography.* New York: Wiley & Sons.

Burnston, D., Sheredos, B., Abrahamson, A., & Bechtel, W. (2014). Scientists' Use of Diagrams in Developing Mechanistic Explanations: A Case Study from Chronobiology. *Pragmatics and Cognition, 22*, 224-243.

Callendar, C., & Cohen, J. (2006). There Is No Special Problem About Scientific Representation. *Theoria, 55*, 67-85.

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

Cartwright, N. (1999). *The Dappled World: A Study of the Boundaries of Science.* New York: Cambridge University Press.

Chakravartty, A. (2010). Informational versus functional accounts of scientific representation. *Synthese, 172*, 197-213.

Chang, H. (1995). Circularity and Reliability in Measurement. *Perspectives on Science, 3*, 153-172.

Chang, H. (2004). *Inventing Temperature.* New York: Oxford University Press.

Chang, H. (2012). *Is Water H2O?* New York: Springer.

Chang, H. (2022). *Realism for Realistic People.* New York: Cambridge University Press.

Collins, H. (2010). *Tacit and Explicit Knowledge.* Chicago: University of Chicago Press.

Compton, A. (1917). The Intensity of X-ray Reflection, and the Distribution of the Electrons in Atoms. *Physical Review, 9*, 29-57.

Compton, A. (1923). The Quantum Integral and Diffraction by a Crystal. *Proceedings of the National Academy of Sciences, 9*, 359-362.

Contessa, G. (2007). Scientific Representation, Interpretation, and Surrogative Reasoning. *Philosophy of Science, 74*, 46-68.

Contessa, G. (2009). *Scientific Representation: Perspectives of Paradox*. Retrieved from Notre Dame Philosophical Reviews: https://ndpr.nd.edu/reviews/scientific-representation-paradoxes-of-perspective/

Coopmans, C., Vertesi, J., Lynch, M., & Woolgar, S. (Eds.). (2014). *Representation in Scientific Practice Revisited.* Cambridge, MA: MIT Press.

Cox, E., & Crowfoot, D. (1936). Crystallography. *Annual Reports on the Progress of Chemistry, 33*, 196-227.

Cranswick, L. (2008). Busting out of crystallograph's Sisyphean prison: from pencil and paper to structure solving at the press of a button. *Acta Crystallographica, A, 64*, 65-87.

Crowfoot, D. (1938). The crystal structure of insulin I. The investigation of air-dried insulin crystals. *Proceedings of the Royal Society, A, 164*, 580-602.

Crowfoot, D. (1972). Dorothy Crowfoot Hodgkin - Biographical. In *Nobel Lectures, Chemistry 1963-1970.* Amsterdam: Elsevier.

Cullis, A., Dinitz, H., & Perutz, M. (1958). X-ray analysis of haemoglobin. In *Conference on Hemoglobin: 2-3 May 1957.* Washington DC: National Academies Press.

Curiel, E. (2021). *Math Does Not Represent.* Retrieved from Center for Philosophy of Science YouTube: https://www.youtube.com/watch?v=aA_T20HAzyY

Cussins, A. (1992). Content, embodiment, and objectivity: The theory of cognitive trails. *Mind, 101*, 651-688.

Dalton, J. (1803). On the absorption of gases by water and other liquids. In *Foundations of the Atomic Theory* (pp. 15-26). Edinburgh: William F. Clay.

Darwin, C. (1914). The Theory of X-ray Reflexion. *Philosophical Magazine, 27*, 315-333.

Daston, L. (2011). Empires of Observation: 1600-1800. In L. Daston, & E. Lunbeck (Eds.), *Histories of Scientific Observation.* Chicago: University of Chicago Press.

Daston, L., & Galison, P. (2007). *Objectivity.* New York: Zone Books.

Daston, L., & Galison, P. (2007). *Objectivity.* Cambridge MA: MIT Press.

de Chadarevian, S. (2003). *Designs for Life.* New York: Cambridge University Press.

de Chadarevian, S. (2018). John Kendrew and myoglobin: Protein structure determination in the 1950s. *Protein Science, 27*, 1136-1143.

Debye, P. (1913). Über die Intensitätsverteilung in den mit Röntgenstrahlen erzeugten Inferenz-bildern. *Verhandlungen der Deutschen physikalischen Gesellschaft, 15*, 738-752.

Debye, P. (1914). Interferenz von Röngtenstrahlen und Wärmebewegung. *Annalen der Physik, 43*, 49-95.

Debye, P., & Scherrer, P. (1916). Interferenz an regellos orientierten Teilchen im Röntgenlicht I. *Physikalische Zeitschrift, 17*, 277-283.

Deichmann, U. (2007). "Molecular" versus "Colloidal": Controversies in Biology and Biochemistry, 1900-1940. *Bulletin of the History of Chemistry, 32*, 105-118.

Dewey, J. (1980). The Need for a Recovery of Philosophy. In J. Boydston (Ed.), *The Middle Works of John Dewey, Vol. 10.* Carbondale: Southern Illinois University Press.

Dobrucki, J. (2013). Fluorescence Microscopy. In U. Kubitscheck (Ed.), *Fluorescence Microscopy: From Principles to Biological Applications* (pp. 97-142). Singapore: Wiley-Blackwell.

Donné, A. (1844). *Cours de Microscopie.* Paris: J.-B. Bailliere.

Dounce, A., & Allen, P. (1988). Fifty years later: recollections of the early days of protein crystallization. *TIBS, 13*, 317-320.

Drenth, J. (2007). *Principles of Protein X-ray Crystallography.* New York: Springer.

Dreyfus, H. (1991). *Being-in-the-World: A Commentary on Heidegger's Being and Time, Division I.* Cambridge MA: MIT Press.

Duane, W. (1925). The Calculation of the X-ray Diffracting Power at Points in a Crystal. *Proceedings of the National Academy of Sciences, 11*, 489-493.

Edsall, J. (1972). Blood and Hemoglobin: The Evolution of Knowledge of Functional Adaptation in a Biochemical System. *Journal of the History of Biology, 5*, 205-257.

Edwards, P. (1999). Global Climate Science, Uncertainty and Politics: Data-laden Models, Model-filtered Data. *Science as Culture, 8*, 437-472.

Elliot-Graves, A. (2020). What is a target system? *Biology & Philosophy, 35*, https://link.springer.com/article/10.1007/s10539-020-09745-3.

Elvira, V. (2017). Impact of Detector Simulation in Particle Physics Collider Experiments. *Physics Reports, 695*, 1-54.

Epstein, P., & Ehrenfest, P. (1924). The Quantum Theory of the Fraunhoffer Diffraction. *Proceedings of the National Academy of Sciences, 10*, 133-139.

Ewald, P. (1913). *Physikalische Zeitschrift, 14*, 471.

Ewald, P. (1914). Die Berechnung der Kristallstruktur aus Interferenzaufnahmen mit X-strahlen. *Physikalische Zeitschrift, 15*, 399-401.

Ewald, P. (1925). Die Intensitäten der Röntgenreflexe und der Strukturfaktor. *Physikalische Zeitschrift, 24*, 29-32.

Ewald, P. (1962). *Fifty Years of X-ray Diffraction.* Utrecht: International Union of Crystallography.

Fernet, E. (1858). *Du role des principaux éléments du sang dans l'absorption ou le dégangement des gaz de la respiration.* Paris: Martinet.

Finch, J. (2007). *A Nobel Fellow on Every Floor.* Cambridge: The Medical Research Council Laboratory of Molecular Biology.

Finlay, M. (1992). *Science, practice and politics: German agricultural experiment stations in the nineteenth century.* Dissertation: Iowa State University.

Fischer, E. (1907). Die Chemie der Proteine und ihre Beziehungen zur Biologie. In E. Fischer, & M. Bergmann (Ed.), *Untersuchungen über Aminosäuren, Polypeptide, und Proteine II* (pp. 1-21). Berlin: Springer.

Fitelson, B. (2001). A Bayesian Account of Independent Evidence with Applications. *Proceedings of the Philosophy of Science Association,* S123-140.

Forman, P. (1969). The Discovery of the Diffraction of X-rays by Crystals; A Critique of the Myths. *Archive for History of Exact Sciences, 6*, 38-71.

Foster, M. (1880). *A Textbook of Physiology.* New York: MacMillan and Co.

Frank, R. (1987). American Physiologists in German Laboratories, 1865-1914. In *Physiology in the American Context, 1850-1940* (pp. 11-46). New York: Springer.

Franklin, A. (1986). *The Neglect of Experiment.* New York: Cambridge University Press.

French, S., & Ladyman, J. (1999). Reinflating the semantic approach. *International Studies in the Philosophy of Science, 13*, 103-121.

Friedrich, W. (1913). Röntgenstrahlinterferenzen. *Physikalische Zeitschrift, 14*, 1079-1087.

Friedrich, W., Knipping, P., & Laue, M. (1913). Interferenzerscheinungen bei Röntgenstrahlen. *Annalen der Physik, 41*, 971-988.

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

Frigg, R. (2010b). Fiction and Scientific Representation. In M. C. Hunter (Ed.), *Beyond Mimesis and Convention* (pp. 97-138). New York: Springer.

Frigg, R., & Nguyen, J. (2017). Models and Representation. In L. Magnani, & T. Bertolotti (Eds.), *Springer Handbook of Model-Based Science* (pp. 49-102). New York: Springer.

Fruton, J. (1979). Early Theories of Protein Structure. *Annals of the New York Academy of Sciences, 325*, 1-20.

Fruton, J. (1999). *Proteins, Enzymes, Genes.* New Haven: Yale University Press.

Funke, O. (1851). Ueber das Milzvenenblut. *Zeitschrift für Rationelle Medicin, 9*, 172-218.

Funke, O. (1853). *Atlas of physiological chemistry.* London: Cavendish Society.

Funke, O. (1876). *Lehrbuch der Physiologie.* Leipzig: Leopold Voss.

Fuyuno, I. (2005). Lack of lab notes cast doubt on RNA researchers' results. *Nature, 437*, 461.

Galison, P. (1985). How the first neutral-current experiments ended. *Reviews of Modern Physics, 55*, 477-509.

Galison, P. (1987). *How Experiments End.* Chicago: University of Chicago Press.

Galison, P. (1997). *Image and Logic: A Material Culture of Microphysics.* Chicago: University of Chicago Press.

Gamgee, A. (1898). Haemoglobin: Its compounds and the principal products of its decomposition. In E. A. Schäfer (Ed.), *Text-book of Physiology* (pp. 185-260). New York: MacMillan.

Gangloff, J., & Allamel-Raffin, C. (2014). Commentary: Some Notions of Action. In L. Soler, S. Zwart, M. Lynch, & V. Israel-Jost (Eds.), *Science After the Practice Turn in the Philosophy, History, and Social Studies of Science* (pp. 58-66). New York: Routledge.

Giddens, A. (1984). *The Constitution of Society.* Berkeley: University of California Press.

Giegé, R. (2013). A historical perspective on protein crystallization from 1840 to the present day. *FEBS Journal, 280*, 6456-6497.

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

Giere, R. (2004). How Models Are Used to Represent Reality. *Philosophy of Science, 71*, 742-752.

Glasser, O. (1936). What Kind of Tube did Röntgen Use When He Discovered the X-ray? *Radiology, 27*, 138-140.

Godfrey-Smith, P. (2006). The strategy of model-based science. *Biology and Philosophy, 21*, 725-740.

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

Goldman, A. (1979). What is Justified Belief? In G. Pappas (Ed.), *Justification and Knowledge* (pp. 1-23). Dordrecht: Reidel.

Gooding, D. (1992). Putting Agency Back into Experiment. In A. Pickering (Ed.), *Science as Practice and Culture* (pp. 65-112). Chicago: University of Chicago Press.

Goodman, N. (1968). *Languages of Art.* New York: Bobbs-Merrill.

Graham, T. (1861). Liquid Diffusion applied to Analysis. *Philosophical Transactions of the Royal Society, 151*, 183-224.

Granick, S. (1942). Some properties of crystalline guinea pig haemoglobin. *The Journal of General Physiology, 25*, 571-578.

Hacking, I. (1981). Do We See Through a Microscope? *Pacific Philosophy Quarterly, 62*, 305-322.

Hacking, I. (1983). *Representing and Intervening.* New York: Cambridge University Press.

Hacking, I. (1992). The Self-Vindication of the Laboratory Sciences. In A. Pickering (Ed.), *Science as Practice and Culture* (pp. 29-64). Chicago: University of Chicago Press.

Haga, H., & Wind, C. H. (1899). Die Beugung der Röntgenstrahlen. *Annalen der Physik und Chemie, 68*, 884-916.

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

Harré, R. (2003). The Materiality of Instruments in a Metaphysics for Experiments. In H. Radder (Ed.), *The Philosophy of Scientific Experimentation* (pp. 19-38). Pittsburgh: University of Pittsburgh Press.

Hartree, D. (1925). The Atomic Structure Factor in the Intensity of Reflexion of X-rays by Crystals. In J. M. Bijvoet, W. G. Burgers, & G. Hägg (Eds.), *Early Papers on Diffraction of X-rays by Crystals, Vol. 1* (pp. 297-298). Utrecht: A. Oosthoek.

Haueis, P. (2020). A Generalized Patchwork Approach to Scientific Concepts. *University of Pittsburgh Center for Philosophy of Science Talk, February 11*.

Haurowitz, F. (1949). The Internal Structure of Protein Molecules. *Experientia, 5*, 347-354.

Haüy, R. (1784). *Essai d'une Théorie sur la Structure des Crystaux.* Paris: Gogué & de la Rochelle.

Haüy, R. (1822). *Traité de Cristallographie.* Paris: Bachelier & Huzard.

Havighurst, R. (1925). The Distribution of Diffracting Power in Sodium Chloride. *Proceedings of the National Academy of Sciences, 11*, 502-507.

Heidegger, M. (1963). *Being and Time.* New York: Harper & Row.

Herschel, J. F. (1831). *A Preliminary Discourse on the Study of Natural Philosophy.* London: Longman, Reese, Orme, Brown, and Green.

Hesse, M. (1963). *Models and Analogies in Science.* New York: Sheed and Ward.

Hey, S. (2015). Robust and Discordant Evidence: Methodological Lessons from Clinical Research. *Philosophy of Science, 82*, 55-75.

Hodgkin, D. (1950). X-ray Analysis and Protein Structure. *Cold Spring Harbor Symposium on Quantitative Biology, 14*, 65-78.

Hodgkin, D. (1990). Taking Penicillin to the R.I. In J. Thomas, & D. Phillips (Eds.), *The Legacy of Sir Lawrence Bragg* (pp. 91-95). Northwood: Science Reviews.

Holmes, F. (1993). Justus Liebig and the Construction of Organic Chemistry. In S. Mauskopf (Ed.), *Chemical Sciences in the Modern World* (pp. 119-136). Philadelphia: University of Pennsylvania Press.

Holmes, F. (1995). Crystals and Carriers: The Chemical and Physiological Identification of Hemoglobin. In *No Truth Except in the Details* (pp. 191-244). Dordrecht: Kluwer.

Hong, T. M. (2018). *Topics in LHC Physics.* Seminar: University of Pittsburgh.

Hoppe, F. (1862). Ueber das Verhalten des Blutfarbstoffs im Spectrum des Sonnenlichtes. *Archiv für pathologischen Anatomie, 23*.

Hughes, R. I. (1996). Models and Representation. *Philosophy of Science, 64*, S325-S336.

Hünefeld, F. L. (1840). *Der Chemismus in der thierischen Organisation.* Leipzig: F. A. Brockhaus.

Jung, J., Byeon, I., DeLucia, M., Koharudin, L., Ahn, J., & Gronenborn, A. (2014). Binding of HIV-1 Vpr Protein to the Human Homolog of the Yeast DNA Repair Protein RAD23 (hHR23a) Requires Its Xeroderma Pigmentosum Complementation Group C Binding (XPCB) Domain as Well as the Ubiquitin-associated 2 (UBA2) Domain. *The Journal of Biological Chemistry, 289*, 2577-2588.

Kahr, B., & Shtukenberg, A. (2012). Histories of Crystallography by Shafranovskii and Schuh. In J. Benedict (Ed.), *Recent Advances in Crystallography* (pp. 3-36). London: IntechOpen.

Karstens, B. (2014). The Lack of a Satisfactory Conceptualization of the Notion of Error in the Historiography of Science: Two Main Approaches and Shortcomings. In M. Boumans, G. Hon, & A. Petersen (Eds.), *Error and Uncertainty in Scientific Practice* (pp. 13-38). London: Pickering & Chatto.

Kassabian, M. (1910). *Röntgen Rays and Electro-Therapeutics.* Philadelphia: Lippincott Company.

Kaveh, S. (2021). *A Prescriptivist Account of Physical Theories.* Dissertation: University of Pittsburgh.

Kay, L. (1993). *The Molecular Vision of Life.* New York: Oxford University Press.

Keilin, D., & Hartree, E. F. (1935). The Combination between Methaemoglobin and Peroxides: Hydrogen Peroxide and Ethyl Hydroperoxide. *Procedings of the Royal Society of London, B, 117*, 1-15.

Keiser, E. (1901). *An Outline of the Progress of Chemistry in the Nineteenth Century*. Retrieved from Scientific American: https://www.scientificamerican.com/article/an-outline-of-the-progress-of-chemi/

Kendrew, J. (1987). *Talk by Sir John Kendrew on fifty years of Computing in Cambridge.* Cambridge Computer Laboratory.

Kendrew, J. (1990). Bragg's Broomstick and the Structure of Protein. In J. Thomas, & D. Phillips (Eds.), *The Legacy of Sir Lawence Bragg* (pp. 88-91). Northwood: Science Reviews.

Kendrew, J., & Perutz, M. (1948). A comparative X-ray study of foetal and adult sheep haemoglobins. *Proceedings of the Royal Society of London, A, 194*, 375-398.

Kendrew, J., & Perutz, M. (1948). A comparative X-ray study of foetal and adult sheep haemoglobins. *Proceedings of the Royal Society, A, 194*, 375-398.

Kennefick, D. (2007). *Traveling at the Speed of Thought.* Princeton: Princeton University Press.

Khalifa, K., Millson, J., & Risjord, M. (2020). Inquisitive Inferentialism About Scientific Representation. *Draft*.

Khalifa, K., Millson, J., & Risjord, M. (2022). Scientific Representation: An Inferentialist-Expressivist Manifesto. *Manuscript*.

Kidd, J. (1809). *Outlines of Mineralogy*. London: Oxford.

Kitcher, P. (1985). *The Nature of Mathematical Knowledge.* New York: Oxford University Press.

Klein, U. (2001). The Creative Power of Paper Tools in Early Nineteenth-Century Chemistry. In *Tools and Modes of Representation in the Laboratory Sciences* (pp. 13-34). New York: Springer.

Klein, U. (2003). *Experiments, Models, Paper Tools.* Stanford: Stanford University Press.

Knuuttila, T., & Loettgers, A. (2011). The Productive Tension: Mechanisms Vs. Templates in Modeling the Phenomena. In P. Humphreys, & C. Imbert (Eds.), *Models, Simulations, and Representation.* New York: Routledge.

Knuuttila, T., & Loettgers, A. (2013). Basic Science Through Engineering? Synthetic Modeling the Idea of Biology-Inspired Engineering. *Studies in History and Philosophy of Science Part C, 44*, 158-169.

Knuuttila, T., & Loettgers, A. (2016). Model Templates Within and Between Disciplines: From Magnets to Gases--and Socio-economic Systems. *European Journal for Philosophy of Science, 6*, 377-400.

Kohler, R. (1994). *Lords of the Fly.* Chicago: University of Chicago Press.

Kohler, R. (2008). *From medical chemistry to biochemistry*. New York: Cambridge University Press.

Kripke, S. (1982). *Wittgenstein on Rules and Private Language.* Cambridge MA: Harvard University Press.

Kuhn, T. (1962). *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

Kuorikoski, J., & Marchionni, C. (2016). Evidential Diversity and the Triangulation of Phenomena. *Philosophy of Science, 83*, 227-247.

Land, W. M., Volchenkov, D., Blasing, E., & Schack, T. (2013). From action representation to action execution: exploring the links between cognitive and biomechanical levels of motor control. *Frontiers in Computational Neuroscience, 7*, 1-14.

Landecker, H. (2007). *Culturing Life.* Cambridge, MA: Harvard University Press.

Lange, R. (2003). Technology as Basis and Object of Experimental Practices. In H. Radder (Ed.), *The Philosophy of Scientific Experimentation* (pp. 119-137). Pittsburgh: University of Pittsburgh Press.

Laszlo, P. (2001). Conventionalities in Formula Writing. In *Tools and Modes of Representation in the Laboratory Sciences* (pp. 47-60). New York: Springer.

Laudan, L. (1981). A Confutation of Convergent Realism. *Philosophy of Science, 48*, 19-49.

Law, J. (1973). The Development of Specialties in Science: The Case of X-ray Protein Crystallography. *Science Studies, 3*, 275-303.

Lehmann, C. G. (1851). *Physiological Chemistry, Vol. 1.* (G. Day, Trans.) London: Harrison and Son.

Lehmann, C. G. (1855). *Physiological Chemistry.* (G. Day, Trans.) Philadelphia: Blanchard & Lea.

Leonelli, S. (2019). What distinguishes data from models. *European Journal for Philosophy of Science, 9*, Access online, DOI: https://doi.org/10.1007/s13194-018-0246-0.

Levins, R. (1966). The Strategy of Model Building in Population Biology. *American Scientist, 54*, 421-431.

Liebig, J. (1839). *Instructions for the Chemical Analysis of Organic Bodies.* Glasgow: Griffin & Company.

Lloyd, E. (1988). *The Structure and Confirmation of Evolutionary Biology.* Princeton: Princeton University Press.

Loeb, J. (1912). *The Mechanistic Conception of Life.* Chicago: University of Chicago Press.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*, 492-527.

Longino, H. (1990). *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry.* Princeton: Princeton University Press.

Longino, H. (2013). *Studying Human Behavior.* Chicago: Chicago University Press.

Lonsdale, K. (1929). X-ray Evidence on the Structure of the Benzene Nucleus. *Transactions of the Faraday Society, 25*, 352-366.

Lynch, M., & Woolgar, S. (Eds.). (1990). *Representation in Scientific Practice.* Cambridge, MA: MIT Press.

Mayo, D. (2014). Learning from Error: How Experiment Gets a Life (of its Own). In M. Boumans, G. Hon, & A. Petersen (Eds.), *Errors and Uncertainty in Scientific Practice* (pp. 57-78). London: Pickering & Chatto.

Michler, G. (2008). *Electron Microscopy of Polymers.* New York: Springer.

Mikhalevich, I. (2016). Experimenting and Intervening: The Pragmatic Value of Intervention. *USML Colloquium Series, March 11*.

Mitchell, S. (2003). *Biological Complexity and Integrative Pluralism.* Princeton: Princeton University Press.

Mitchell, S. (2009). *Unsimple Truths.* Chicago: University of Chicago Press.

Mitchell, S. (2019). Perspectives, Representation, and Integration. In M. Massimi, & C. McCoy (Eds.), *Understanding Perspectivism* (pp. 178-193). New York: Routledge.

Mitchell, S. (2020). Instrumental Perspectivism: Is AI Machine Learning like NMR Spectroscopy. In M. Bertolaso, & F. Sterpetti (Eds.), *A Critical Reflection on Automated Science* (pp. 27-42). New York: Springer.

Mitchell, S. (forthcoming 2023). The Bearable Thinness of Being: A Pragmatist Metaphysics of Affordances. In S. Mitchell, & H. Andersen (Eds.), *The Pragmatist Challenge* (pp. 112-142). New York: Oxford University Press.

Mitchell, S. W. (1858). Observations on the Blood Crystals of the Sturgeon. *Proceedings of the Academy of Natural Sciences of Philadelphia, 10*, 2-4.

Mitchell, S., & Gronenborn, A. (2017). After Fifty Years, Why Are Protein X-ray Crystallographers Still in Business? *Philosophy of Science, 68*, 703-723.

Mitscherlich, E. (1843). *Über ein Goniometer.* Berlin.

Mohs, F. (1825). *Treatise on Mineralogy.* (W. Haidinger, Trans.) Edinburgh: Constable and Co.

Morgan, M., & Morrison, M. (Eds.). (1999). *Models as Mediators: Perspectives on Natural and Social Science.* New York: Cambridge University Press.

Mould, R. F. (1995). The early history of X-ray diagnosis with emphasis on the contributions of physics 1895-1915. *Physics in Medicine & Biology*, 1741-1787.

Mulder, G. J. (1838). Sur la Composition de Quelques Substances Animáles. *Bulletin des Sciences Physiques et Naturelles*, 104-119.

Nachman, B. (2016). *Investigating the Quantum Properties of Jets and the Search for a Supersymmetric Top Quark Partner with the ATLAS Detector.* Dissertation: Stanford University.

Neurath, O. (1931/1983). Physicalism. In O. Neurath, R. S. Cohen, & M. Neurath (Eds.), *Philosophical Papers 1913-1946* (pp. 52-57). Dordrecht: Reidel.

Nguyen, J. (2016). On the Pragmatic Equivalence Between Representing Data and Phenomena. *83*, 171-191.

Nienhaus, G., & Nienhaus, K. (2013). Fluorescence Labeling. In U. Kubitscheck (Ed.), *Fluorescence Microscopy: From Principles to Biological Applications* (pp. 143-169). Singapore: Wiley-Blackwell.

Northrop, J. (1930). Crystalline Pepsin. *The Journal of General Physiology, 13*, 739-766.

Northrop, J. (1964). The preparation of pure enzymes and virus proteins. In *Nobel Lectures, Chemistry 1942-1962* (pp. 124-134). Amsterdam: Elsevier.

O'Conner, C., & Weatherall, J. (2016). Black Holes, Black-Scholes, and Prairie Voles. *Philosophy of Science, 83*, 613-626.

Olby, R. (2012). *The Path to the Double Helix.* New York: Dover.

Osborne, T. (1902). The basic character of the protein molecule and the reactions of edestin with definite quantities of acids and alkalies. *Journal of the American Chemical Society, 24*, 39-78.

Osborne, T. (1908). Our Present Knowledge of Proteins. *Science, 28*, 417-427.

Ostwald, W. (1917). *An Introduction to Theoretical and Applied Colloid Chemistry*. (M. Fischer, Trans.) New York: Wiley & Sons.

Patterson, A. (1935). A Direct Method for the Determination of the Components of Interatomic Distances in Crystals. *Zeitschrift für Kristallographie, 90*, 517-542.

Pauling, L., & Corey, R. (1951). Configurations of polypeptide chains with favored orientations around single bonds: two new pleated sheets. *Proceedings of the National Academy of Sciences, 37*, 729-740.

Pauling, L., & Niemann, C. (1939). The Structure of Proteins. *Journal of the American Chemical Society, 61*, 1860-1867.

Perini, L. (2012). Image Interpretation: Bridging the Gap from Mechanically Produced Image to Representation. *International Studies in the Philosophy of Science, 26*, 153-170.

Perovic, S. (2017). Experimenter's regress argument, empiricism, and the calibration of the large hadron collider. *Synthese, 194*, 313-332.

Perutz, M. (1949). A X-ray study of horse methaemoglobin. II. *Proceedings of the Royal Society, A, 195*, 474-499.

Perutz, M. (1965). Structure and Function of Haemoglobin: I. A Tentative Atomic Model of Horse Oxyhaemoglobin. *Journal of Molecular Biology, 13*, 646-668.

Perutz, M. (1990a). How Lawrence Bragg Invented X-ray Analysis. In J. Thomas, & D. Phillips (Eds.), *The Legacy of Sir Lawrence Bragg* (pp. 71-86). Northwood: Science Reviews.

Perutz, M. (1990b). Mechanisms regulating the reactions of human hemoglobin with oxygen and carbon monoxide. *Annual Review of Physiology, 52*, 1-25.

Perutz, M. (1997). *The Medical Research Council Laboratory of molecular biology*. Retrieved from NobelPrize.org: https://www.nobelprize.org/prizes/uncategorized/the-medical-research-council-laboratory-of-molecular-biology/

Perutz, M., Rossmann, M., Cullis, A., Muirhead, H., & Will, G. (1960). Structure of Haemoglobin. *Nature, 185*, 416-422.

Pickering, A. (1992). From Science as Knowledge to Science as Practice. In A. Pickering (Ed.), *Science as Practice and Culture* (pp. 1-28). Chicago: University of Chicago Press.

Pickering, A. (1995). *The Mangle of Practice.* Chicago: University of Chicago Press.

Pincock, C. (2011). Bas van Fraassen *Scientific Representation: Paradoxes of Perspective*. *British Journal for the Philosophy of Science, 62*, 677-682.

Pincock, C. (2011a). *Mathematics and Scientific Representation.* New York: Oxford University Press.

Pincock, C. (2011b). Modeling reality. *Synthese, 180*, 19-32.

Polanyi, M. (1921). Faserstruktur im Röntgenlichte. *Die Naturwissenschaften, 9*, 337-340.

Polanyi, M. (1966). *The Tacit Dimension.* New York: Doubleday & Company.

Preyer, W. (1871). *Die Blutkrystalle.* Jena: Mauke's.

Putnam, F. (1994). Felix Haurowitz. *Biographical Memoirs, 64,* 134-163.

Putnam, H. (1981). *Reason, Truth, and History.* New York: Cambridge University Press.

Radder, H. (2012). *The Material Realization of Science: From Habermas to Experimentation and Referential Realism.* New York: Springer.

Rasmussen, N. (1999). *Picture Control.* Stanford: Stanford University Press.

Recanati, F. (2010). *Truth-Conditional Pragmatics.* New York: Oxford University Press.

Reichert, E., & Brown, A. (1909). *The Crystallography of Hemoglobins.* Washington, DC: Carnegie Institution of Washington.

Rheinberger, H. (1997). *Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube.* Palo Alto: Stanford University Press.

Rheinberger, H.-G. (2009). Recent science and its exploration: the case of molecular biology. *Studies in History and Philosophy of Biological and Biomedical Sciences, 40,* 6-12.

Rheinberger, H.-G. (2010). *An Epistemology of the Concrete.* Durham: Duke University Press.

Rice, C. (2019). Models Don't Decompose That Way: A Holistic View of Idealized Models. *British Journal for the Philosophy of Science, 70,* 179-208.

Rich, A. (2003). The double helix: a tale of two puckers. *Nature Structural and Molecular Biology, 10,* 247-249.

Ritter, C. (2001). An Early History of Alexander Crum Brown's Graphical Formulas. In *Tools and Modes of Representation in the Laboratory Sciences* (pp. 35-46). New York: Springer.

Rocke, A. (2000). Organic Analysis in Comparative Perspective: Liebig, Dumas, and Berzelius, 1811-1840. In F. Holmes, & T. Levere (Eds.), *Instruments and Experimentation in the History of Chemistry* (pp. 273-310). Cambridge MA: MIT Press.

Rocke, A. (2001). Chemical Atomism and the Evolution of Chemical Theory in the Nineteenth Century. In *Tools and Modes of Representation in the Laboratory Sciences* (pp. 1-12). New York: Springer.

Röntgen, W. (1897). Weitere Beobachtungen über die Eigenschaften der X-Strahlen. *Annalen der Physik und Chemie,* 18-37.

Röntgen, W. (1995). On a New Kind of Rays. *Veterinary Radiology & Ultrasound, 36,* 371-374.

Rouse, J. (2006). Practice Theory. In S. Turner, & M. Risjord (Eds.), *Handbook of the Philosophy of Science, Volume 15: Philosophy of Anthropology and Sociology.* New York: Elsevier.

Rule, G., & Hitchens, K. (2006). *Fundamentals of Protein NMR Spectroscopy.* New York: Springer.

Ryle, G. (1949). *The Concept of Mind.* Chicago: University of Chicago Press.

Sagnac, G. (1899). Transformations des rayons X par la matiere. *Éclairage électrique, 18*, 41-48.

Salkowski, E. (1903). *A Laboratory Manual of Physiological and Pathological Chemistry.* New York: Wiley & Sons.

Sapp, K. (2019). Interview. *University of Pittsburgh*.

Schank, J., & Wimsatt, W. (1986). Generative Entrenchment and Evolution. *Philosophy of Science, 1986*, 33-60.

Schickore, J. (2007). *The Microscope and the Eye.* Chicago: University of Chicago Press.

Schickore, J. (2017). *About Method.* Chicago: University of Chicago Press.

Schuh, C. (2007). *Mineralogy & Crystallography.* Tuscon: manuscript.

Seegelke, C., & Schack, T. (2016). Cognitive Representation of Human Action: Theory, Applications, and Perspectives. *Frontiers in Public Health, 4*, 1-6.

Séguy, G. (1897). Tubes for the production of Roentgen rays. *Nature, 55*, 296-297.

Seliger, H. (1995). Wilhelm Conrad Röntgen and the Glimmer of Light. *Physic Today, 48*, 25-31.

Sellars, W. (1953). Inference and Meaning. *Mind, 62*, 313-338.

Shearer, G. (1923). The Relation between Molecular and Crystal Symmetry as Shown by X-ray Crystal Analysis. *Proceedings of the Physical Society of London, 35*, 81-100.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing. II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127-190.

Sohncke, L. (1879). *Theorie der Krystallstruktur.* Leipzig: Teubner.

Sommerfeld, A. (1901). Theoretisches über die Beugung der Röntgenstrahlen. *Zeitschrift für Mathematik und Physik*, 11-97.

Stark, J. (1913). Zur Diskussion des Verhaltens der Röntgenstrahlen in Kristallen. *Physikalische Zeitschrift, 14*, 319-321.

Stebbings, S. (1937). *Philosophy and the Physicists.* London: Methuen and Co.

Steensen, N. (1669). The Prodromus to a Dissertation on a Solid Naturally Contained Within a Solid. In N. Steno, T. Kardel, & P. Maquet (Eds.), *Biography and Original Papers* (pp. 621-660). New York: Springer.

Stegenga, J. (2009). Robustness, Discordance, and Relevance. *Philosophy of Science, 76*, 650-661.

Sterelny, K. (2012). *The Evolved Apprentice.* New York: Oxford University Press.

Steuwer, R. (1971). Bragg's Corpuscular Theory of X-rays and y-rays. *The British Journal for the History of Science, 5*, 258-281.

Stokes, G. G. (1864). On the reduction and oxidation of the colouring matter of the blood. *Proceedings of the Royal Society of London, 13*, 355-384.

Stokes, G. G. (1898). On the Nature of the Röntgen Rays. In *Röntgen Rays* (pp. 43-65). New York: American Book Company.

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

Suárez, M. (2004). An Inferential Conception of Scientific Representation. *Philosophy of Science, 71*, 767-779.

Suárez, M. (2015a). Representation in Science. In P. Humphreys (Ed.), *The Oxford Handbook of Philosophy of Science* (pp. Accessed online, DOI: 10.1093/oxfordhb/9780199368815.013.25). New York: Oxford University Press.

Suárez, M. (2015b). Deflationary representation, inference, and practice. *Studies in History and Philosophy of Science, 49*, 36-47.

Sumner, J. (1926). The isolation and crystallization of the enzyme urease. *Journal of Biological Chemistry, 69*, 435-441.

Sumner, J. (1937). The Story of Urease. *Journal of Chemical Education, 14*, 255-259.

Sumner, J. (1964). The Chemical Nature of Enzymes. In *Nobel Lectures, Chemistry 1942-1962* (pp. 114-121). Amsterdam: Elsevier.

Suppe, F. (1989). *The Semantic Conception of Scientific Theories and Scientific Realism.* Chicago: University of Chicago Press.

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

Svedberg, T., & al. (1939). A Discussion on the Protein Molecule. *Proceedings of the Royal Society, B, 127*, 1-40.

Szabadváry, F. (1986). The History of Chemical Laboratory Equipment. *Periodica Polytechnica Chemical Engineering, 30*, 77-95.

Tal, E. (2012). *The Epistemology of Measurement: A Model-based Account.* Dissertation: University of Toronto.

Tal, E. (2017). Calibration: Modelling the Measurement Process. *Studies in the History and Philosophy of Science Part A, 65*, 33-45.

Tal, E. (2021). Two Myths of Representational Measurement. *Perspectives on Science, 29*, 701-741.

Taylor, W., & West, J. (1928). The Crystal Structure of the Chondrodite Series. *Proceedings of the Royal Society of London, A, 117*, 517-532.

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

Thompson, E. (1896). *Roentgen Rays and Phenomena of the Anode and Cathode*. New York: D. Van Nostrand Company.

Thomson, J. J. (1898). A Theory of the Connection between Cathode and Röntgen Rays. In G. Barker (Ed.), *Röntgen Rays* (pp. 69-72). New York: American Book Company.

Tibbets, P. (1990). Representation and the realist-constructivist controversy. In M. Lynch, & S. Woolgar (Eds.), *Representation in Scientific Practice* (pp. 69-84). Cambridge, MA: MIT Press.

Turner, S. (1994). *The Social Theory of Practices: Tradition, Tacit Knowledge, and Presuppositions.* Chicago: University of Chicago Press.

van Fraassen, B. (1980). *The Scientific Image.* New York: Oxford University Press.

van Fraassen, B. (2008). *Scientific Representation: Paradoxes of Perspective.* New York: Oxford University Press.

van Fraassen, B. (2010). Reply to Contessa, Ghins, and Healey. *Analysis, 70*, 547-556.

Van Fraassen, B. (2012). Modeling and Measurement: The Criterion of Empirical Grounding. *Philosophy of Science, 86*, 773-784.

Van Helden, A. (2010). Galielo and the telescope. In A. V. Helden, S. Dupré, R. v. Gent, & H. Zuidervaart (Eds.), *The origins of the telescope* (pp. 183-202). Amsterdam: KNAW Press.

Wallace, D. (2020). On the Plurality of Quantum Theories: Quantum theory as a framework, and its implications for the quantum measurement problem. In S. French, & J. Saatsi (Eds.), *Scientific Realism and the Quantum.* New York: Oxford University Press.

Walter, B., & Pohl, R. (1909). Weitere Versuche über die Beugung der Röntgenstrahlen. *Annalen der Physik, 29*, 356.

Warren, B., & Bragg, W. L. (1929). The Structure of Diopside. In J. M. Bijvoet, W. G. Burgers, & G. Hägg (Eds.), *Early Papers on Diffraction of X-rays by Crystals, Vol. 2* (Vol. 69, pp. 320-334). Utrecht: A. Oosthoek.

Waters, K. (2004). What was classical genetics? *Studies in History and Philosophy of Science Part C, 35*, 783-809.

Waters, K. (2019). An Epistemology for Scientific Practice. *Philosophy of Science, 86*, 585-611.

Watson, G. (2001). *The Double Helix.* New York: Touchstone.

Weisberg, M. (2013). *Simulation and Similarity: Using Models to Understand the World.* New York: Oxford University Press.

Weissenberg, K. (1924). Ein neues Röntgengoniometer. *Zeitschrift für Physik, 23*, 229-238.

Wheaton, B. (1983). *The tiger and the shark: Empirical roots of wave-particle dualism.* New York: Cambridge University Press.

Williamson, T. (2018). Counterpossibles. *Topoi, 37*, 357-368.

Willstätter, R. (1927). Problems and Methods in Enzyme Research. *Journal of the Chemical Society, 0*, 1359-1381.

Wilson, M. (2006). *Wandering Significance: An essay on conceptual behavior.* New York: Oxford University Press.

Wilson, M. (2018). *Physics Avoidance.* Oxford: Oxford University Press.

Wimsatt, W. (1980). Reductionistic Research Strategies and Their Biases in the Units of Selection Controversy. In T. Nickles (Ed.), *Scientific Discovery, vol. 2: Case Studies* (pp. 213-259). Dordrecht: Reidel.

Wimsatt, W. (1981). Robustness, Reliability, and Overdetermination. In M. Brewer, & B. Collins (Eds.), *Scientific Inquiry and the Social Sciences* (pp. 124-163). San Francisco: Jossy-Bass.

Wimsatt, W. (2013). The role of generative entrenchment and robustness in the evolution of complexity. In C. Lineweaver, P. Davies, & M. Ruse (Eds.), *Complexity and the Arrow of Time* (pp. 308-331). New York: Cambridge University Press.

Wittgenstein, L. (1953). *Philosophical Investigations.* Oxford: Blackwell.

Wollaston, W. (1809). Description of a reflective Goniometer. *Philosophical Transactions of the Royal Society of London, 99*, 253-258.

Wollaston, W. (1813). On the elementary particles of certain crystals. *Philosophical Transactions of the Royal Society, 103*, 51-63.

Woodward, J. (2000a). Data, Phenomena, and Reliability. *Philosophy of Science, 67*, S163-179.

Woodward, J. (2000b). Explanation and Invariance in the Special Sciences. *British Journal for the Philosophy of Science, 51*, 197-254.

Woodward, J. (2003). Experimentation, Causal Inference, and Instrumental Realism. In H. Radder (Ed.), *The Philosophy of Scientific Experimentation* (pp. 87-118). Pittsburgh: University of Pittsburgh Press.

Woodward, J. (2006). Some varieties of robustness. *Journal of Economic Methodology*, *13*, 219-240.

Woodward, J. (2015). Sketch of some themes for a pragmatic philosophy of science. *Draft*.

Woody, A. (2004). More Telltale Signs: What Attention to Representation Reveals About Scientific Explanation. *Philosophy of Science, 71*, 780-793.

Woody, A. (2014). Chemistry's Periodic Law: Rethinking Representation and Explanation After the Turn to Practice. In L. Soler, S. Zwart, M. Lynch, & V. Israel-Jost (Eds.), *Science After the Practice Turn in the Philosophy, History, and Social Studies of Science* (pp. 123-150). New York: Routledge.

Wu, J., O'Connor, C., & Smaldino, P. (forthcoming 2023). The Cultural Evolution of Science. In J. Kendal, R. Kendal, & J. Tehrani (Eds.), *The Oxford Handbook of Cultural Evolution.* New York: Oxford University Press.

Wyckoff, R. (1930). The Atomic Scattering Powers of Nickel, Copper and Iron for Various Wavelengths. *Physical Review, 36,* 1116-1120.

Yeo, R. (2014). *Notebooks, English Virtuosi, and Early Modern Science*. Chicago: University of Chicago Press.

Zinoffsky, O. (1886). Ueber die Grösse des Hämoglobinmolecüls. *Zeitschrift für Physiologische Chemie, 10,* 16-34.

Zsigmondy, R. (1966). Properties of colloids. In *Nobel Lectures, Chemistry 1922-1941* (pp. 45-57). Amsterdam: Elsevier.