THE BEAUTY OF SYMMETRICAL DESIGN:
THE ALLEGED EPISTEMIC ROLE OF AESTHETIC VALUE IN
THEORETICAL SCIENCE

By
Gregory J. Morgan

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
June 2004

© Gregory John Morgan 2004

All rights reserved
Abstract

The view that beauty is a sign of a theory’s truth is held by a number of scientists. I argue that this view has no a priori justification. This philosophical project is illustrated by the first written history of the first successful theories of virus structure—the 1956 Crick-Watson theory and the 1962 Caspar-Klug theory—two theories that are especially beautiful, but whose beauty plays no epistemic role or so I argue.

I define two senses of “sign of truth”: (1) A sign₁ of truth that renders a theory likely; and (2) A sign₂ of truth that increases the probability of a theory. After examining two causes of theoretical beauty, high degree of geometrical symmetry and William Whewell’s notion of coherence (or unification), I argue that there is no a priori reason to think beauty satisfies either of these two senses.

I argue that beauty cannot be an a priori sign₁ of truth since no features of beauty (or coherence or symmetry) guarantee that there will be a unique empirically adequate beautiful theory. Bayes’s Theorem says there are three components to a theory’s probability. I argue that in the short run or in the long run, beauty fails to systematically affect the prior probability of a theory, the likelihood, or the expectedness of the evidence, i.e., none of the three components of a theory’s probability.

That beauty is a sign₂ of truth is defended by James McAllister. He claims that the meaning of beauty in science tracks the properties of our best past theories. Thus, by induction over theories with these properties, there is warrant for thinking beauty is a sign₂ of truth, or so he claims. His position has serious problems. For example, most past successful theories, even beautiful ones, have been false. Further, he cannot account for the beauty of radically novel theories.

Finally I argue that beauty may give a reason to pursue a theory, but not in any global sense. Rather, any justification derives from society granting individual scientists freedom of inquiry, the right to determine the relative value of theoretical beauty.

Advisor: Peter Achinstein
Preface

Let me begin by thanking my first mentors in the philosophy of science, Lisa Lloyd, now at Indiana University, Paul Griffiths, now at the University of Pittsburgh, and Alan Musgrave, still at my alma mater, the University of Otago. These three scholars fixed my initial orientation in philosophy of science, an orientation from which I have deviated little over the last decade. At my first graduate school, the University of Pittsburgh, the late Wes Salmon, Brad Wilson, and John Norton influenced my approach to philosophy as did my fellow students and friends including Rachel Ankeny, now at University of Sydney, Carl Craver, now at Washington University St Louis, Greg Davis, now at Princeton University, Chris Martin, now at Indiana University, and Chris Smeenk, now at UCLA. Bob Olby fostered my interest in the history of molecular biology and Roger Hendrix eased me into laboratory life while showing me the divine side of the planet’s most abundant life form. Lindley Darden supported my history of molecular biology project and invited me into the District of Columbia History and Philosophy of Biology reading group. I thank Jessica Pfeifer, Ilya Farber, Eric Saidel, Ken Schaffner, Keith Benson, and Lindley Darden for their comments on some early drafts. Sondra and Milton Schlesinger volunteered to proof read the historical chapters, as did my collaborator Angela Creager. I thank Francis Crick and James Watson for their comments chapters 2 and 3. While presenting my philosophical project on the academic job market, astute comments from Dan Callcut, Palle Yourgrau, Michael Ruse, and Josh Gert prompted changes to some of my arguments. Audiences at the Nobel Forum in 2002 and at the History of Science Society meeting in 2003 gave useful feedback on the historical narrative. To my committee, Peter Achinstein, Michael Williams, Ed Lattman, Sharon Kingsland, and Nathaniel Comfort, let me thank you for your probing questions and your suggestions on how to improve sections for publication. Peter Achinstein’s careful reading of many developing drafts of this dissertation, his insightful comments, his support, and his wise counsel are greatly appreciated. It is said that imitation
is the sincerest form of flattery. Those familiar with Achinstein’s philosophical style will see that I have striven to emulate his rigorous approach to philosophical problem solving. 

I thank the United States National Science Foundation (NSF) for a Doctoral Dissertation Award (# 9910891) that allowed me to travel to the United Kingdom and interview several eminent virologists and the Johns Hopkins Department of Philosophy for the Sachs Memorial Fellowship that supported me in 2002. Sydney Brenner, Donald Caspar, Carolyn Cohen, Bob Connelly, Dick Crane, Francis Crick, David DeRosier, Lynn Elkin, John Finch, Alfred Gierer, Fred Holmes, Robert Horne, Hugh Huxley, Aaron Klug, Ruben Leberman, Barbara Low, Vittorio Luzzati, Brenda Maddox, John Maddox, Lee Makowski, Tony Minson, Alison Newton, Max Perutz, Ivan Raymont, Alex Rich, Willie Russell, Herbert Simon, Michael Stoker, James Watson and, Jo Wildy generously shared information about the history of science. Correspondence with Peter Kosso, Peter Lipton, and James McAllister has refined my view relative to theirs. Sylvia Alloway, Ed Applewhite, Jeremy Goldberg, Martin Kemp, Anne Massey, Magda Cordell McHale, Kenneth Snelson, and Kirby Urner helped me better understand the connections with art. Jeremy Norman generously allowed me access to his archive for molecular biology, as did Caltech Archives, Stanford University Archives, University of Maryland Baltimore County Archives, Royal Society Archives, Wellcome Institute for History of Medicine Archives, The Novartis Foundation (formerly The Ciba Foundation), Churchill Archives Centre, Tate Gallery Archive, University of Melbourne Archives, Cold Spring Harbor Laboratory Archives, and the Laboratory of Molecular Biology, Cambridge, Archive.

Finally I thank immediate family—my brother Stewart McKenzie Morgan, my parents, John Douglas Morgan and Barbara Ann Morgan, and especially my wife, Stacey Noelle Welch—for their unconditional support and my newly born son, Edward Nicolson Welch Morgan, for his timely arrival. True love is a beautiful thing.
This dissertation has purposefully been written so that the historical and philosophical components can be read separately if so desired. Those with only historical interests should read Chapters 2 and 3. Those with only philosophical interests should read Chapters 1, 4, 5, and 6.
# Table of Contents

## Chapter 1  Introduction

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1.1</td>
<td>Isaac Newton on Beautiful Simplicity</td>
<td>3</td>
</tr>
<tr>
<td>1.1.2</td>
<td>William Whewell on Beautiful Coherence</td>
<td>6</td>
</tr>
<tr>
<td>1.1.3</td>
<td>Beautiful Symmetry</td>
<td>7</td>
</tr>
<tr>
<td>1.2</td>
<td>Preliminary Analysis: Beauty in Science</td>
<td>10</td>
</tr>
<tr>
<td>1.2.1</td>
<td>Aesthetic Properties of Scientific Theories</td>
<td>10</td>
</tr>
<tr>
<td>1.2.2</td>
<td>Two Probabilistic Conceptions of “Sign of Truth”</td>
<td>13</td>
</tr>
<tr>
<td>1.2.3</td>
<td>Four Positions in the Conceptual Geography</td>
<td>18</td>
</tr>
<tr>
<td>1.2.4</td>
<td>Newton and Whewell Revisited</td>
<td>19</td>
</tr>
<tr>
<td>1.3</td>
<td>Plan of Dissertation</td>
<td>21</td>
</tr>
</tbody>
</table>

## Chapter 2  Early Structural Virology

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Point Symmetry in the History of Science</td>
<td>25</td>
</tr>
<tr>
<td>2.1.1</td>
<td>Point Symmetry in Plato’s <em>Timaeus</em></td>
<td>27</td>
</tr>
<tr>
<td>2.1.2</td>
<td>Spherical Symmetry in Aristotle’s Cosmology</td>
<td>29</td>
</tr>
<tr>
<td>2.1.3</td>
<td>Point Symmetry in Kepler’s <em>Mysterium Cosmographicum</em></td>
<td>31</td>
</tr>
<tr>
<td>2.2</td>
<td>The Beginnings of Structural Virology: Tobacco Mosaic Virus</td>
<td>33</td>
</tr>
<tr>
<td>2.2.1</td>
<td>The Early Tobacco Mosaic Virus Research of Donald Caspar</td>
<td>33</td>
</tr>
<tr>
<td>2.2.2</td>
<td>The Early Tobacco Mosaic Virus Research of James Watson</td>
<td>35</td>
</tr>
<tr>
<td>2.2.3</td>
<td>The Watson-Caspar Collaboration at Caltech</td>
<td>38</td>
</tr>
<tr>
<td>2.2.4</td>
<td>The Early Tobacco Mosaic Virus Research of Rosalind Franklin</td>
<td>43</td>
</tr>
<tr>
<td>2.3</td>
<td>The Beginnings of Structural Virology: Spherical Viruses</td>
<td>46</td>
</tr>
<tr>
<td>2.3.1</td>
<td>Five-fold Symmetry in Tomato Bushy Stunt Virus</td>
<td>47</td>
</tr>
<tr>
<td>2.3.2</td>
<td>The Crick-Watson Theory of Spherical Virus Structure</td>
<td>49</td>
</tr>
</tbody>
</table>
Chapter 3 The Caspar and Klug Collaboration

3.1 The Beginning of the Caspar-Klug Collaboration.................................65
  3.1.1 The Insight of the Physicist Richard Crane.................................69
3.2 Aaron Klug’s Work on Polio Virus.....................................................72
3.3 The Connection with Art: John McHale.............................................78
3.4 Buckminster Fuller and “Synergetic” Geometry................................83
  3.4.1 The Meeting Between Fuller and Klug.........................................87
3.5 The Continued Collaboration between Donald Caspar and Aaron Klug....87
3.6 Improved Electron Micrographs of Turnip Yellow Mosaic Virus...........89
  3.6.1 The Invention of Negative Staining............................................94
3.7 The Birth of Quasi-Equivalence.......................................................100
  3.7.1 The Selection Rule: \( h^2 + hk + k^2 \) ........................................109
  3.7.2 Two Projects in Theoretical Structural Virology:
      Geometrical and Physical.................................................................110
  3.7.3 The 1962 Cold Spring Harbor Meeting on Animal Virus Biology......111
3.8 Michael Goldberg’s Rediscovered Insight..........................................123
3.9 Historical Summary.............................................................................126
3.10 Some Preliminary Philosophical Lessons.........................................126

Chapter 4 Does Beauty Render a Theory Likely?

4.1 Beauty as a Sign of Truth.................................................................128
### Chapter 4

#### 4.2 The Form of the Aesthete’s Argument

- **4.2.1 Assessing the Aesthete’s First Argument**
- **4.2.2 Whewell on Coherence, Beauty, and Truth**

#### 4.3 The Problem of Beautiful Rival Theories

#### 4.4 The Aesthete’s Short-Run Rejoinder

- **4.4.1 Does Beauty Influence the Likelihood of a Theory?**
- **4.4.2 Does Beauty Influence the Prior Probability of a Theory?**

#### 4.5 Beauty and Belief

### Chapter 5

**Does Beauty Increase a Theory’s Probability?**

- **5.1 Beauty as a Sign of Truth**
  - **5.1.1 Four Ways Beauty could be a Sign of Truth**
- **5.2 William Whewell’s Notion of Coherence: An A Priori Sign of Truth?**
  - **5.2.1 Achinstein’s Rendition of Whewellian Coherence**
  - **5.2.2 A Priori Signs of Truth: The Dilemma**
- **5.3 Is Beauty an Empirical Sign of Truth?**
- **5.4 James McAllister’s Defense of Beauty as a Sign of Truth**
  - **5.4.1 Consequences of McAllister’s Model**
  - **5.4.2 General Objections to McAllister’s Model of Aesthetics in Science**
  - **5.4.3 Comparing McAllister’s Model with the Case Study**
- **5.5 Is the View that Beauty is a Sign of Truth Tenable?**
- **5.6 Is the Argument against an Epistemic Role for Beauty Too Good?**
- **5.7 Summary and Conclusion of Chapter 5**

### Chapter 6

**A Non-epistemic Role for Beauty**

- **6.1 Introduction**
- **6.2 Broadening the Project: from Means to Goals and from Belief to Action**
6.3  The Logic of Pursuit.................................................................209

6.4  A Brief Survey of Views............................................................211

6.4.1  Herbert Feigl: Beauty has No Role........................................212

6.4.2  Pierre Duhem: Beauty is a Goal of Science.............................212

6.4.3  Richard Feynman: Beauty as Physical Laws............................214

6.4.4  Henri Poincaré: Beauty as Motivation.................................215

6.4.5  Thomas Kuhn: Beauty as Subjective Value............................215

6.4.6  Taking Stock........................................................................219

6.5  A Decision Theoretic Analysis..................................................220

6.5.1  Goals and Means in Decision Theory....................................223

6.5.2  Beauty and Decision Theory................................................225

6.5.3  Two Decisions: by the Society and by the Individual...............228

6.5.4  Inter-value Realism and Inter-value Anti-realism....................229

6.5.5  Goals as Constraints on Utility Values...................................230

6.5.6  Ideal Deliberation: Individual Inter-value Anti-realism.............232

6.6  Poincaré Duhem and Kuhn Reconsidered..................................233

6.6.1  Poincaré Reconsidered.........................................................234

6.6.2  Duhem Reconsidered..........................................................235

6.6.3  Kuhn Reconsidered..............................................................237

6.7  Beauty and the Freedom of Inquiry: A Proposal..........................238

6.7.1  Freedom of Inquiry and Decision Theory...............................239

6.8  Summary and Conclusion..........................................................242

Appendix I  Archival Sources.........................................................244

Appendix II  Interviews with the Author.........................................246

Bibliography...............................................................................247

Scholarly Life..............................................................................271
List of Tables

Table 3.1  The structure of Caspar and Klug’s 1962 paper..........................119
Table 6.1  A decision theoretic matrix for theory pursuit..............................221
Table 6.2  A decision theoretic matrix for utility adoption...........................241
List of Figures

Figure 2.1  Radial density function of TMV relative to the density of water………………..40
Figure 2.2  Left: Caspar’s diffraction picture showing spikes in the 5-fold direction.  Right:  
Idealized data.................................................................48
Figure 2.3  Michael Stoker, Jim Watson, Milton Salton, and Francis Crick at the 1956  
CIBA symposium...............................................................52
Figure 2.4.  Madrid International Union of Crystallography meeting 1956.  From left to  
right, Anne Cullis, Francis Crick, Don Caspar, Aaron Klug, Rosalind Franklin,  
Odile Crick, and John Kendrew.  Image courtesy of Don Caspar.................57
Figure 3.1.  Two pictures that were printed in The Observer June 21, 1959, that probably  
captured the attention of John McHale....................................77
Figure 3.2  Transistor (1954) by John McHale.  Inspired by information theory……...83
Figure 3.3  Radio geodesic dome or “Radome”.....................................................86
Figure 3.4  Klug and Finch’s 1960 model of turnip yellow mosaic virus.....................93
Figure 3.5  Herpes simplex virus.................................................................98
Figure 3.6  Adenovirus...............................................................................99
Figure 3.7  “Empty” herpes virus particles.......................................................104
Figure 3.8  Dome assembled in Kabul Afghanistan in 1956.................................105
Figure 3.9  The 270-subunit tensegrity sphere...............................................107
Figure 3.10  Bonding diagram, drawn by Klug in February 1962, indicating how simple  
bonding rules and near equivalence can lead to icosahedral structures……...113
Figure 3.11  Number of capsomere clusters represented on a hexagonal grid..........125
Chapter 1 Introduction

Structure:

1.1 Introduction

   1.1.1 Isaac Newton on Beautiful Simplicity
   1.1.2 William Whewell on Beautiful Coherence
   1.1.3 Beautiful Symmetry

1.2 Preliminary Analysis: Beauty in Science

   1.2.1 Aesthetic Properties of Scientific Theories
   1.2.2 Two Probabilistic Conceptions of “Sign of Truth”
   1.2.3 Four Positions in the Conceptual Geography
   1.2.4 Newton and Whewell Revisited

1.3 Plan of Dissertation

1.1 Introduction

   It is not difficult to find scientists who claim that beauty is a sign of truth. James Watson speaking of the newly discovered model of DNA writes, “…like almost everyone else [Rosalind Franklin] saw the appeal of the base pairs and accepted the fact that the structure was too pretty not to be true.” (Watson 1968, p. 210; italics mine) In a similar vein, François Jacob says, “[Watson and Crick’s model of DNA] was of such simplicity, such perfection, such harmony, such beauty even, and biological advantages flowed from it with such rigor and clarity, that one could not believe it was untrue.” (Jacob 1988, p. 271) Jacques Monod referring to the Caspar-Klug theory of virus structure, the historical focus of this dissertation, claims, “A beautiful model or theory may not be right; but an ugly one must be wrong.” (Monod 1969, p. 23) Paul Dirac, writing of Einstein’s theory of relativity, claims, “It is the essential beauty of the theory which I feel is the real reason for us

The existence of such views, especially among such eminent scientists, raises a whole host of questions for the philosopher of science. Are these scientists correct? Is beauty a sign of truth? Can we determine what to believe without empirically testing theories? If beauty were a sign of truth, what sort of sign would it be, a priori or empirical? Would a connection between truth and beauty be an ontological claim about the world or merely a claim about how we reason? What exactly is it for a theory to be beautiful? Are there objective criteria for a theory’s purported beauty? Is there an epistemic role for aesthetics in science? If so, what is that role? If not, what is the role of beauty in science? Obviously, some of these questions are especially deep, and it would be difficult to give conclusive answers to all of them in a single dissertation. Nonetheless, I will attempt to make some progress in understanding the purported epistemic role of aesthetics in science. More specifically, I will examine the claim that beauty is a sign of truth. To telegraph my conclusion, I will argue that for the most part the above scientists are mistaken. Beauty is not a sign of truth, or at least it is not in any a priori sense. Whether there is an empirical connection between beauty and truth has yet to be settled and is a question for scientists not philosophers. On the other hand, there may be a role for beauty in science that is independent of the goal of truth.

The role of beauty in science has not been totally neglected by theorists of scientific methodology. Indeed many of the influential writers of scientific methodology, either

---

1 See Chao (1997) for more on the beauty of General Relativity and Engler (2001) and Dickson (forthcoming) for more on Dirac’s aesthetic views.
explicitly or implicitly, invoke aesthetic concepts in explicating their conception of proper scientific methodology. As their efforts are worth keeping in mind and also serve to justify further work on the role of beauty in science, let me mention several of the more famous attempts. That is, before beginning in a more analytical vein, I will briefly discuss the methodological writings of Isaac Newton, William Whewell, and the physicist Anthony Zee’s rendition of Einstein’s methodological strategies. These three thinkers focus on three different aesthetic properties that can imbue a theory with beauty: simplicity; coherence; and symmetry.

1.1.1 Isaac Newton on Beautiful Simplicity

Let us begin with Isaac Newton who makes simplicity a central principle in his system. Simplicity is perhaps the most widely invoked property that many deem aesthetic, although it is not inconsistent to advocate a conception of simplicity that is independent of aesthetic considerations. Nonetheless, J. B. S. Haldane expresses a common attitude:

In scientific thought we adopt the simplest theory which will explain all the facts under consideration and enable us to predict new facts of the same kind. The catch in this criterion lies in the word “simplest.” It is really an aesthetic canon such as we find in our criticism of poetry or painting (quoted in McAllister 1996, p. 105).

Newton codifies this attitude in one of his four rules of reasoning in Book III of the *Principia*. His first rule of reasoning in philosophy is:

We are to admit no more causes of natural things than such are both true and sufficient to explain their appearances. (Newton [1726] 1934, p. 398)

Although this rule of reasoning does not mention the word simplicity, in the next sentence Newton says, “To this purpose the philosophers say that Nature does nothing in vain, and more is in vain when less will serve; Nature is pleased with simplicity, and affects not the pomp of superfluous causes,” which makes it clear that Newton intended the first law to be about simplicity (Newton [1726] 1934, p. 398). Further, Newton’s personified nature who
is “pleased” with simplicity and lack of “pomp” suggests an aesthetic dimension to Rule I. Among other things, Newton uses Rule I to argue that gravity and only gravity (since more causes would be more than sufficient), is the cause that retains all the planets in their orbits (Scholium on Theorem V). Newton took satisfaction of this rule to be a sign of truth. To see why, consider his fourth rule of reasoning, which begins, “In experimental philosophy we are to look upon the propositions inferred by general induction from the phenomena as accurately or very nearly true …” (Newton [1726] 1934, p. 400). Since we often apply the first rule before the fourth (when we are dealing with the causes of phenomena), the proposition resulting from Rule I must have at least as high probability as the more general proposition resulting from Rule I and Rule IV. For example, if we are to “look upon” gravity as the only cause that retains all the planets in their orbits as true or nearly true, as Newton suggests, then we should also look upon the proposition that one planet, Venus say, is retained in her orbit only by gravity as true or nearly true also. Newton is thus committed to simplicity being a sign of truth.

On what grounds does Newton believe that simplicity is a sign of truth? Does he think it is an empirical fact about nature or does he think it is merely a methodological rule reflecting how we reason? Textual evidence suggests Newton thinks his rule of reasoning reflects empirical facts about nature. In the preface to the second edition of Newton’s *Principia Mathematica*, Roger Cotes writes about Newton’s method. He calls Newton’s approach “experimental philosophy.” Those who possess the experimental philosophy, he claims, “derive the causes of things from the most simple principles possible; but they assume nothing as a principle, that is not proved by the phenomena.” This quotation

---

2 Unfortunately, I. B. Cohen’s recent English translation of Newton’s *Principia* drops this flowery language (Newton 1999 [1726]). Interestingly, Mamiani (2002) argues that Newton’s rules of reasoning were based on Robert Sanderson’s *Logicae Artis Compendium* and rules contained in his *Treatise on Apocalypse* (c. 1672). If Mamiani is correct, Rule I derives from Newton’s belief that God’s word is never ambiguous. For Newton, God is the author of the Bible and of Nature and creates only the minimum number of meanings and causes. As Mamiani puts it, “In his search for a criterion of the truth, Newton made no
suggests that, as Newton himself says, the principles that underlie Newton’s system, presumably Newton’s Laws of Motion, are not derived a priori, but rather are empirical—they are derived from phenomena that are not given a priori. On the other hand, Cotes says nothing about Newton’s rules of “derivation” themselves: are they a priori or empirical? In truth-preserving deductive logic, the rules of deduction, such as modus ponens, can be known a priori. They do not depend on how the world actually is. Is the same true of the rules that govern Newton’s “deductions”? Newton says very little about this issue, but what he does say suggests his Rule I is not a priori. In the one sentence of commentary on Rule I, Newton appears to justify the rule with the following sentence I will quote again: “To this purpose the philosophers say that Nature does nothing in vain, and more is in vain when less will serve; Nature is pleased with simplicity, and affects not the pomp of superfluous causes.” In other words, because Newton believes that nature is simple and not ostentatious, he believes that our true theories of nature should be simple also. It could have been that nature affected the pomp of superfluous causes, but in fact she does not. Newton justifies his sign of truth—simplicity—on empirical grounds.

An obvious problem with justifying a methodological rule on empirical grounds is that it introduces, or at the very least appears to introduce, circularity. If we could ask Newton how he knows that Nature is simple, then he might claim that the deepest established scientific knowledge expresses simple facts and laws. For example, Newton’s own laws of motion are remarkably simple. But if we established these propositions by correctly using Newton’s rules of reasoning, then we necessarily must arrive at the most simple adequate explanations of the phenomena. It will not do to justify a principle of simplicity by invoking propositions gained via prior applications of the principle. No doubt distinction between science and theology” (Mamiani 2002, p. 391).
there are replies to this worry that one could make on behalf of the great man. Perhaps
some propositions can be gained without the use of Rule I. That is, perhaps propositions
could be gained by using only Rules III and IV that could be used to justify Rule I. (Rule
II follows from Rule I according to Newton.) We might then ask how we justify Rules III
and IV and a similar problem arises. Alternatively, it is possible that the phenomena are
such that an explanation of them that is simpler than any other is not a simple explanation.
Newton’s Rule I prescribes that we adopt the most simple explanation given the constraint
that the explanation adopted adequately explain the phenomena. It is at least possible that
the phenomena are so complex that no simple explanation is adequate and the only adequate
explanations are complex (i.e., involve many types of causes). However, even if this were
the case, Newton still has to justify why we should adopt the most simple of the complex
candidates “as true” and the circularity again raises its ugly head. I will return to the idea
of an empirical justification of an aesthetic methodological principle in Chapter 5.

1.1.2 William Whewell on Beautiful Coherence

William Whewell was one of the most prominent writers on the philosophy of
science in the 19th century. His momentous Philosophy of the Inductive Sciences is a
classic work in the philosophy of science. I will argue that Whewell believes that beauty is
an a priori sign of truth and that theoretical beauty for him consists in the coherence of
theories. Regarding the progress of science, he claims:

In [true theories] all the additional suppositions tend to simplicity and harmony; the new
suppositions resolve themselves into the old ones, or at least require only some easy

3 The Newton scholar Ted McGuire seems to agree with me on this point (McGuire 1995, p. 71). He
argues that Newton believes that the rules are “gathered from the whole” (from the Latin verb ‘colligiur’)
set of phenomena. On the other hand, George Smith argues that the rules are merely a way to manage
epistemic risk (Smith 2002, p. 162).
modification of the hypothesis first assumed: the system becomes more coherent as it is
further extended. … In false theories, the contrary is the case (Whewell 1847, p. 68).
While Whewell illustrates his claim with numerous episodes in the history of science, his
reason for believing that coherence (or at least more coherence) is a sign of truth is that he
believes it follows a priori. He seems to think it follows from the nature of truth: “Truth
may give rise to such a coincidence; falsehood cannot” (Whewell 1847, p. 71). For
Whewell, truth is coherent and simple (Whewell 1847, p. 72). False hypotheses cannot
explain all phenomena whereas true hypotheses can (Whewell 1847, p. 62). For him, the
true philosopher, if successful, “binds together” otherwise incoherent facts like stringing a
set of pearls, to use his simile. In fact, in a claim which arguably conflates psychological
inability and logical impossibility, Whewell goes as far as to say that once a system of facts
have been bound together, it is often impossible to see the facts as “incoherent and
detached” again (Whewell 1847, p. 52). Robert Butts claims that Whewell views facts thus
bound together as necessary and known by intuition (Butts 1968, p. 16). The claim that the
purported connection between beauty and truth is a priori is not open to the same objection
of circularity that I raised against Newton. Presumably, however, Whewell should be able to
provide a proof of the connection between beauty and truth or at least of the connection
between coherence and truth that would convince any rational agent. I don’t think he is able
to provide such a proof and there are additional problems with his proposal. I will return to
these issues as well as a fuller elaboration of Whewell’s position in Chapters 4 and 5.

1.1.3 Beautiful Symmetry

The final aesthetic property that I would like to consider is symmetry, which has
received less attention from philosophers than simplicity and coherence. 4 Thinkers from

4 See Rosen (1995), Stewart and Golubitsky (1993), and Shubnikov and Koptsik (1972) for good overviews
the Ancient Greeks onward have proposed particular symmetries to underlie the structure of the world and its objects. Not coincidentally perhaps, the same symmetries often have been viewed as beautiful. As Elaine Scarry, the Harvard scholar of aesthetics, puts it, symmetry is “the single most enduring recognized attribute” of beautiful objects (Scarry 1999, p. 96). It is thus not too surprising to find prominent thinkers who think that beautiful symmetries are signs of truth. In Chapter 2, I survey some of the more famous episodes in the history of science that have involved postulating beautiful symmetries, including Plato’s geometrical elements in his Timaeus, Aristotle’s spherical universe in his On the Heavens and Kepler’s platonic solids in his Mysterium Cosmographicum. Beautiful symmetries also have been claimed to play a central role in a methodology that allegedly springs from Einstein’s advances in physics. Physicist Brian Greene in his book, The Elegant Universe, describes the situation in recent theoretical physics:

… as we enter an era in which our theories describe realms of the universe that are increasingly difficult to probe experimentally, physicists do rely on such an aesthetic to help them steer clear of blind alleys and dead-end roads that they might otherwise pursue. So far, this approach has provided a powerful and insightful guide. In physics, as in art, symmetry is a key part of aesthetics (Greene 1999, p. 167).

Greene’s metaphor of a powerful and insightful guide who helps avoid cul-de-sacs suggests that through the use of symmetry, physicists can increase the probability that they have the true theory without experimentally testing the theory. The metaphor of a guide cuts across the contexts of discovery and justification. A (reliable) guide leads one to a theory whose probability of being true is high or at least increases the probability that one will find the truth. Herman Weyl, author of perhaps the most famous book on symmetry, expresses

Falkenburg (1988) argues symmetries in particle physics unify phenomena. In effect, she argues that symmetry and what Whewell calls coherence are closely related in particle physics. Mainzer (1997) makes the same claim more generally. In the similar vein, I make no claim that simplicity, coherence, and symmetry are conceptually or statistically independent notions. Whewell himself took his notions of
a similar sentiment: “My work always tried to unite the true with the beautiful; but when I had to choose one or the other, I usually chose the beautiful.” (Quoted in Chandrasekhar 1987, p. 65) I don’t think Weyl means that he would reject truth if it were certain he had it, but rather the choice is between what others take to be the truth independently of aesthetic considerations and what he takes to be beautiful. He thinks that his intuition about beauty is a better indicator of theoretical truth than more conventional empirical evidence. The third sentence of his famous book is: “Beauty is bound up with symmetry.” (Weyl 1952, p. 3)

Einstein motivates the special theory of relativity by appeal to its removal of an ugly asymmetry inherent in the electrodynamics of moving bodies. As he put it later, “For the theoretician such an asymmetry in the theoretical structure, with no corresponding asymmetry in the system of experience, is intolerable.” (Einstein 1983(1920), p. 12)

Although not usually interpreted this way, his famous declaration that God does not play dice also expresses intolerance to the idea that a symmetrical situation can lead to an asymmetrical outcomes through quantum fluctuations. Admittedly, the idea of intolerance need not be an aesthetic notion, but later commentators have interpreted Einstein’s attitude to be motivated by aesthetic principles. For example, motivated by Einstein’s successful appeal to symmetry principles in his development of the two theories of relativity, the physicist Anthony Zee devotes an entire book, Fearful Symmetry: The Search for Beauty in Modern Physics, to the role of symmetry as an aesthetic concept in physics. He begins his book with the provocative claim:

When presented with two alternative equations purporting to describe Nature, we always choose the one that appeals to our aesthetic sense. ‘Let us worry about beauty first and truth will take care of itself!’ Such is the rallying cry of fundamental physicists (Zee 1986, p. 3).

consilience and coherence to complement and perhaps replace Newton’s notion of simplicity.
The quotation does not explicitly mention symmetry, but a few pages later, Zee says, “…following the ancient Greeks, who waxed eloquent on the perfect beauty of spheres and the celestial music they make, I will continue to equate symmetry with beauty.” (Zee 1986, p. 9) Zee holds that one should appeal to the symmetry of the equations when choosing which equations to adopt. There is a conservative and a radical way to interpret Zee. First, when he says we should “choose an equation” that appeals to our aesthetic sense, he could mean choose an equation to pursue or to work on further. On this interpretation, beauty gives us a reason to “adopt” the equation but not necessarily a reason to believe it. Second, on the radical reading of Zee, when he says, “chose a equation” he means choose an equation to believe. Beauty gives us some reason to believe the equation. I think he means the latter radical position. Admittedly, Zee holds that we also have to test the equations against the world (Zee 1986, p. 98). However, by rejecting ugly possibilities, Zee and Greene suggest we increase the probability of reaching a true theory before we administer the empirical test.

1.2 Preliminary Analysis: Beauty in Science

1.2.1 Aesthetic Properties of Beautiful Theories

One goal for the philosopher might be to provide a definition of “beauty” or “aesthetic property.” This is one of the goals of philosophical aesthetics. Implicit within most scientists’ aesthetic judgments, I suspect, is an account of beauty that defines beauty in terms of a particular emotional response.⁵ One could try to make this account explicit, but as this is primarily a work in the philosophy of science, I do not intend to proceed in such a manner. Further, I suspect that beauty in science is like obscenity, difficult to define, even

---

⁵ Birkhoff (1956) develops a view of aesthetics in science along these lines.
though it is not difficult to identify. Instead I will assume that the theories that the large majority of scientists judge to be beautiful are, for the most part, the beautiful ones. That is, I assume that by and large scientists are correct about the extension of the term “beautiful theory.” Perhaps surprisingly, there is a lot of agreement about which theories are the beautiful ones amongst scientists. I also will assume that scientists mostly are correct about what makes a given beautiful theory beautiful. Nonetheless, insofar as a term’s meaning is dependent upon how the sentences that contain the term are inferentially related to other sentences, this dissertation sheds some light on the meaning of beauty in science. The primary goal of the dissertation, however, is to examine the purported epistemic role of beauty in science. My strategy will be to consider what properties beauty must have if it is to play an epistemic role in theoretical science; and then to argue that whatever scientists mean by beauty, within certain reasonable constraints, it does not have these properties.

Although I will not offer a definition of beauty in science, I will offer some analysis of the use of the term. As it stands, the term, “beauty,” as used in science is a fairly general term and is often used loosely to encompass more specific aesthetic properties such as simplicity, coherence, or symmetry. Scientists’ aesthetic judgments are typically not fine enough to make distinctions between “elegance”, “prettiness”, “loveliness”, “coolness” or “beauty.” Indeed often these terms or their associated adjectives are treated as synonyms. Since at one level my project is conceptual analysis of scientific discourse, I too, will not make fine distinctions between these subtly different aesthetic concepts and shall take “beauty” to be the exemplary aesthetic concept in science. This is not to say, however, that beautiful theories may not differ wildly in what causes them to be beautiful. On my view, there are a manifold of properties each of which if possessed by a particular theory can render a theory beautiful. A theory may be beautiful because of its simplicity, because

---

6 Urmson (1957) famously makes the same point about aesthetic situations.
7 Rota makes the same point about mathematical beauty and mathematicians (Rota 1997, p. 175).
of its coherence, because of the symmetrical structures it posits, or for some other reason. Although, I will purposefully not define theoretical beauty, my strategy will be to examine some of the principal causes of theoretical beauty in considering the purported epistemic role of beauty.

Let us consider what is entailed by the claim that a theory is beautiful. Since I am concerned with what we should believe and a convenient unit of belief is the scientific theory, let us frame the discussion in terms of theories and their properties. Overall, a scientific theory may be beautiful, ugly, or neither. A beautiful theory will have at least one property or relation between its properties whose possession renders the theory beautiful. Similar considerations hold for an ugly theory, although some ugly properties will simply be a distinct lack of a corresponding beautiful property. I call such properties, “beautiful properties” and “ugly properties,” respectively and the union of these two sets of properties constitutes a significant subset of all aesthetic properties. Possession of a beautiful property might be compensated by the possession of an ugly property and it is quite possible for a theory to possess both beautiful and ugly properties at once. To judge the overall beauty of a particular theory, we must judge the relative importance of each aesthetic property of the theory and consider the aesthetic value of the theory and its properties as a whole. Sometimes possession of a beautiful property will not be sufficient to render a theory beautiful. In other cases it will be. A theory’s beauty can even be over-determined, i.e., some theories exist that possesses multiple beautiful properties, possession of any one of which would individually be sufficient to render it beautiful. In fact, I claim that the beauty of most beautiful theories is over-determined. Beauty, like many of the properties responsible for it, admits of degrees. It is meaningful to talk of one theory being more beautiful or less beautiful than another as well as a theory being beautiful or ugly.

The philosopher of science will not be surprised to learn that some of the most plausible candidates for an aesthetic property of theories are explicitly discussed by scholars of scientific methodology. As I have stated briefly above, I will discuss three of the
more commonly invoked aesthetic properties: simplicity, symmetry, and coherence (or unification) with more emphasis on the last two for reasons of space. It would be contentious to claim that all instances of these properties constitute instances of an aesthetic property. For instance, some scholars discuss simplicity for example, making clear that they want to discuss a conception of simplicity that is not related to aesthetics at all. On the other hand, others see aesthetics inextricably tied to simplicity and other properties of theories. What I call aesthetic properties share a lot of features with what Bernard Williams calls “thick concepts” (Williams 1985, p. 129). In contrast to thin concepts like “right” and “wrong,” thick concepts like “cruelty,” for example, have for Williams both a descriptive and an evaluative dimension. Whether an action is cruel depends both on empirical factual features of the action and agents’ moral evaluations of the action. Although Williams discusses moral concepts, the concepts that express aesthetic properties also seem to share this duality. To some, whether a theory is simple or not seems to be merely a straightforward matter of fact, but to others it clearly is an evaluative judgment that depends on one’s aesthetic taste. A third position is that it is both evaluative and a matter of fact. I do not intend to settle this debate, but I would claim that there are cases in which symmetry, simplicity and coherence exhibit an evaluative dimension. The concept of beauty itself, as I treat it, is necessarily evaluative and is the aesthetic analog of the concept of good and so is a thin concept.  

1.2.2 Two Probabilistic Conceptions of “Sign of Truth”

To make some progress on whether beauty is a sign of truth, let us become clearer on what it means to be a sign of truth. The notion of a sign of truth is multiply ambiguous. We can use the language of probability to tease out the different conceptions. A standard way to show dependence between two characteristics of theories is to use conditional

---

8 On my view, Putnam mistakenly claims that beauty is a thick concept (Putnam 2002, p. 132).
probability. This is especially true of the dominant Bayesian approach to evidence (See for example Earman (1992) or Howson and Urbach (1989)). There seems to me to be at least two ways in which the claim that beauty is a sign of truth could be understood. First there is an *absolute* notion. Let us define a necessary condition of an absolute notion that I will call a “threshold sign of truth” or a sign₁ of truth:

If beauty is a sign₁ of truth, then, for any theory T, \( p(T/ T \text{ is beautiful}) > k \), where \( k \geq 0.5 \). Clearly, this is not a *sufficient* condition since it could be the case that \( p(T/\neg(T \text{ is beautiful})) = p(T/T \text{ is beautiful}) > k \), in which case beauty is irrelevant to the probability of a theory, but T is nonetheless likely. However, since my strategy will be to argue that beauty is not a sign₁ of truth by denying the consequent of this conditional, I need not consider sufficient conditions for a sign₁ of truth. Presumably, adding the condition \( p(T/\neg(T \text{ is beautiful})) \leq k \) to the above necessary condition would generate necessary and sufficient conditions for beauty being a sign₁ of truth. One weakness of the necessary condition is that it is undefined when \( p(T \text{ is beautiful}) = 0 \) since \( p(T/T \text{ is beautiful}) = p(T \& T \text{ is beautiful})/p(T \text{ is beautiful}) \). If you find this weakness troubling, then consider an alternative formulation:

For any theory T, if T is beautiful and beauty is a sign₁ of truth, then \( p(T) > k \). This second formulation arguably follows from the first.⁹ For the most part I will assume the first formulation because it allows one to use the machinery of conditional probabilities to analyze what is logically entailed by the definition. It also allows a closer contact with the

---

⁹ Using the proof system of Belnap (1993) augmented by additional theorems of the probability calculus:
1. For all Ti, \( (p(Ti/Ti \text{ is beautiful}) > k \)
2. \( \text{flag } Ta \text{ for universal generalization} \)
3. \( \text{Il } Ta \text{ is beautiful} \)
4. \( \text{Il } p(Ta \text{ is beautiful}) = 1 \)
5. \( \text{Il } p(Ta/Ta \text{ is beautiful}) > k \)
6. \( \text{Il } p(Ta \& Ta \text{ is beautiful})/p(Ta \text{ is beautiful}) > k \)
7. \( \text{Il } p(Ta \& Ta \text{ is beautiful}) > k \)
8. \( \text{Il } p(Ta) > k \)
9. \( \text{Il } If \text{ Ta is beautiful, then } p(Ta) > k \)
10. For all Ti, \( (If \text{ Ti is beautiful, then } p(Ti) > k) \)
philosophy of science literature on quantitative confirmation. Further, although the case of
p(T is beautiful) = 0 is outside the scope of the first definition and consequently is not well
defined, the second formulation offers little improvement. When p(T is beautiful) = 0, the
second formulation (sign₁) offers no constraints on the probability of T at all. And, this
second formulation is undefined when 0 < p(T is beautiful) < 1.

If beauty were an infallible sign₁ of truth, the value of k would be 1. Most
philosophers of science do not believe that infallible signs of truth exist for non-vacuous
theories. A closely related position is that there is no algorithm of discovery or deductive
logic of discovery. There would be an algorithm of discovery if there were infallible signs
of truth and a mechanizable way of generating a theory with such a sign. On the other hand,
many believe that there are methods for generating plausible hypotheses. That is, they
believe there is a middle path between the anarchism of Feyerabend and the aspirations of
some AI researchers for a mechanizable logic of discovery (Feyerabend 1975; Langley et al.
1987; Nickles 1987). A natural way to render plausibility in a probabilistic approach is to
say a theory is plausible when it has a probability higher than a certain threshold, k. Since I
am concerned not merely with plausibility but with what one should believe, I will consider
this type of sign when k = 0.5. If k < 0.5, then it is more reasonable to believe ~T than to
believe T.

There is a weaker notion of a sign of truth. While sign₁ of truth is absolute, the
second notion of sign of truth, which I call a sign₂ of truth, is relative:

Beauty is a sign₂ of truth if and only if, for any theory T, p(T/T is beautiful) > p(T).
The idea here is that by possessing beauty a theory’s probability is increased. This is
equivalent to saying that a theory’s lack of beauty is a relative sign of its falsity, i.e., p(~T/

We should however be somewhat cautious in accepting this proof since conditional proofs do not normally involve instances where a term does not have a truth-value. Further, step 4 is controversial.
This conception of sign also relies on conditional probabilities and is not defined when $p(T \text{ is beautiful}) = 0$. Let me mention another difficulty with this conception: the possession of a beauty property can be essential to the theory in question. Adding ugly properties or removing the relevant beautiful properties is not possible without creating an entirely new theory. However, it seems to be a mistake to preclude beauty being a sign of truth because the beautiful property in question is an essential part of the theory. To circumvent these weaknesses, a more general conception of sign$_2$ can be formulated:

Beauty is a sign$_{2,1}$ of truth if and only if $p(T_1) > p(T_2)$ where $T_1$ and $T_2$ have all the same epistemic virtues (excluding beauty) in the same degree, but $T_1$ is beautiful and $T_2$ is not.

And if we accept that beauty comes in degrees, a still more general conception can be formulated:

Beauty is a sign$_{2,2}$ of truth if and only if $p(T_1) > p(T_2)$ where $T_1$ and $T_2$ have all the same epistemic virtues (excluding beauty) in the same degree, but $T_1$ is more beautiful than $T_2$.

Like the situation with the definitions of sign$_1$ of truth, similar considerations show that there are virtues of each of these definitions of a sign$_2$ of truth. For the most part, I will assume the first conditional probability definition. There can be relative signs that are not threshold signs. For example, a sign could increase the probability of finding the truth without reaching the threshold $k$.

There are two different ways to interpret the claim that beauty is sign of truth (in either sense): epistemological and ontological. On an epistemological interpretation, the

\[ p(\neg T \mid \neg(T \text{ is beautiful})) > p(\neg T) \]
\[ p(\neg T \& \neg(T \text{ is beautiful})) > p(\neg(T \text{ is beautiful})) \]
\[ [1 - p(\neg(T \& \neg(T \text{ is beautiful})))[1 - p(T \text{ is beautiful})] > 1 - p(T)] \]
\[ 1 - [p(T) + p(T \text{ is beautiful}) \cdot p(T \& T \text{ is beautiful})] > 1 - p(T) - p(T \text{ is beautiful}) + p(T \text{ is beautiful})p(T) \]
\[ 1 - p(T) - p(T \text{ is beautiful}) + p(T \& T \text{ is beautiful}) > 1 - p(T) - p(T \text{ is beautiful}) + p(T \text{ is beautiful})p(T) \]
\[ p(T \& T \text{ is beautiful}) > p(T \text{ is beautiful})p(T) \]
\[ p(T \& T \text{ is beautiful}) > p(T) \]
reason that beauty is a sign of truth, if indeed it is, is because it reflects a way we have to reason, or perhaps the way we should reason about knowledge. In contrast, on the ontological reading, it is due to facts about the world in which we live that beautiful theories are likely (sign₁) or that beautiful theories are on the whole more likely than non-beautiful ones (sign₂). Although it seems natural to affirm or deny these two positions together, it is possible to endorse or deny only one reading. One could take a position analogous to Hume’s position on causation: acknowledge that psychologically we cannot help but to infer truth from beauty, but ontologically remain agnostic about the existence of a connection between beauty and truth. At the other extreme, one could argue that beauty does increase a theory’s probability, but hold that no theory envisioned by humans is ever beautiful and thus that beauty is irrelevant to human reasoning. Nonetheless, the more plausible position is to link the epistemological and ontological positions. On my view, if there is an ontological connection between beauty and truth, then this connection could be used to justify an epistemological rule that we are entitled to infer information about a theories probability in light of knowledge about its beauty. In the opposite direction, there is perhaps some reason to think that if we have a tendency to infer truth (or probability of truth) from beauty, and assuming we, as a species, are well adapted to our environment, then evolution by natural selection has endowed us with this tendency because there really is a connection between beauty and truth. On this just-so story, we are better adapted if our reasoning mirrors the structure of reality. While this approach is attractive to an aspiring philosopher of biology, the evidential basis for the selective history needed for a justification of this evolutionary claim is weak and very difficult to muster given the difficulty of studying the reasoning patterns of our Pleistocene ancestors. I will have more to say about the relevance of evolution for the epistemic role of beauty later.

\[
p(T/T \text{ is beautiful}) > p(T)
\]
1.2.3 Four Positions in the Conceptual Geography

Let us call someone who believes that beauty is a sign of truth in science an “aesthete.” There are many crosscutting distinctions that allow us to classify the possible positions that an aesthete might occupy. First an aesthete may promote beauty as a sign, a sign of truth, or both. Second, one can use the empirical/a priori distinction to dissect four additional positions. To illustrate these four positions, let us consider a sign of truth, although similar remarks apply to a sign of truth. The essential core inference that an Aesthete endorses for a particular beautiful theory Ta is the following:

Premise A. (T) If T is beautiful, then there is reason to believe that it is true.
Premise B. Theory Ta is beautiful
Conclusion: Therefore, there is reason to believe that Ta is true.

For each of the premises, one may ask the question: is it an empirical or an a priori claim? One’s answer to these questions defines one’s place in a simple conceptual space.

Position 1: The rationalist aesthete believes that the truth-values of both of the premises can be known a priori. Of course, what makes her an aesthete is that she believes that Premise A is true. As I will explore further below, a particular sanitized interpretation of William Whewell puts him in the rationalist aesthete camp. On this view of Whewell, it is an a priori truth that there is a connection between coherence and truth and one can tell a priori whether a given theory is coherent.

Position 2: The mixed aesthete believes that the first premise expresses an a priori truth, but whether a particular theory Ta is beautiful or not is an empirical matter. A more orthodox interpretation of William Whewell puts him in this camp.

Position 3: Another mixed position but the opposite of position 2. On this third position, the general rule linking beauty and truth is empirical, but it is an a priori matter whether a particular theory is beautiful. I think this is the most plausible of the four positions.
Position 4: The empiricist aesthete believes that both premises A and B are empirical claims. Those who hold that the general principle is empirical (Positions 3 and 4) often justify their belief inductively. For example, if beautiful theories in the past have been likely (or more likely than non-beautiful ones), then, they argue, we have some reason to think that current beautiful theories are likely. The philosopher James McAllister makes at least this claim in his book *Beauty and Revolution in Science* (McAllister 1996). I will discuss McAllister’s position further in Chapter 5.

1.2.4 Newton and Whewell Revisited

With a more technical definition of sign of truth in hand, let us examine Newton’s and Whewell’s positions again. Remember Whewell represents the aesthete who thinks the general connection between the aesthetic and epistemic realms is a priori and Newton represents the opposite view. Whewell has a fairly sophisticated position and I will discuss him in detail in Chapters 4 and 5. Given the complexities of his position, here I will merely illustrate his position with Peter Achinstein’s probabilistic rendition of Whewell’s notion of coherence. To keep the various positions separate let us call this rendition Achinstein’s Whewell. Following Achinstein (1991), let us say that a theory T consists of a set of hypotheses, $h_1, h_2, \ldots, h_m$. Achinstein first defines what is for a hypothesis $h_1$ to be coherent with the remaining hypotheses.

$h_1$ is coherent with $h_2, \ldots, h_m$ on B, if and only if $p(h_1/h_2, \ldots, h_m \& B) > k$, and $p(h_1/h_2, \ldots, h_m \& B) > p(h_1 / B)$. (Achinstein 1991, p. 130)

Then he introduces the notion of coherence for the set of hypotheses:

A set of hypotheses $h_1, \ldots, h_m$ is coherent, on B, if and only if each hypothesis is coherent with the other members of the set on B (ibid.).

Finally, theory T is coherent if and only if the set of hypotheses, $h_1,\ldots, h_m$ is coherent. Two things are noteworthy about Achinstein’s rendition of Whewell. First, it does not capture
(at least explicitly) the explanatory dimension of coherence to which Whewell alludes.
Second, it bears close resemblance to the above notions of a sign of truth. Relative to the background knowledge B, for a set of hypotheses T, the conjunction of all the members of T except $h_i$ is a sign of truth of $h_i$ and setting $k = 0.5$, it is also a sign of truth of $h_i$. Whewell himself thought that the coherence of a theory T was a sign of the truth of T.

Given the lack of explicit remarks, it is a debatable question whether Whewell himself thought that coherence was necessarily a beautiful property or that coherent theories were necessarily beautiful. He does argue that we should clearly separate the practical arts from the sciences and that incoherent phenomena “disturbs [Art] not” (Whewell 1847, p. 110). The implication being that incoherence disturbs science and should be avoided. On the other hand, for Whewell, whether coherent theories are beautiful might be a question for the science of the perception of beauty (in fine arts), or what he calls “Callaesthetics,” which he distinguishes from the practical arts (Whewell 1847, p. 568). Beyond these tantalizing suggestions, Whewell does not explicitly discuss the relation of coherent theories and beauty anywhere in his corpus to my knowledge. Nonetheless, the general tenor of the work suggests that he thinks science progresses when it becomes more coherent, harmonious, and beautiful. It is no accident that “beauty” is the last word of Philosophy of the Inductive Sciences. Let us extend Whewell’s position and suggest that a plausible Whewellian view would involve the claim that T is coherent if and only if T is beautiful. As Whewell also holds that if T is coherent, then T is true as suggested by his quotations, it follows on this Whewellian view that beauty is a sign of truth. Indeed to be charitable to Whewell, I will relax the assumption that any coherent theory is true and replace it with the more plausible view that any coherent theory is likely. Even with this relaxation the Whewellian still holds that beauty is a threshold sign of truth. The important point is that for Whewell the connections between coherence and truth is a priori and if we can judge the connection between coherence and beauty a priori also, which seems plausible given
Whewell’s view that science (including Callaesthetics) will eventually produce self-evident truths, then the Whewellian believes that beauty is an a priori sign of truth.

Given the definition of sign of truth, how does the Newtonian position fit into the framework? Although Newton does not discuss the connections among beauty, simplicity and truth, let me propose an argument that is consistent with the Newtonian position. That simplicity is a sign of truth can be decomposed into two separate claims that entail the view that simplicity is a threshold sign of truth.

(a) If T is simple, then T is beautiful
(b) If T is beautiful, then p(T) > k

Therefore, If T is simple, then p(T) > k

Unlike the Whewellian position, on the “Newtonian” position, (b) is an empirical question. Newton and Whewell illustrate two extreme views on how to justify signs of truth.

1.3 Plan of Dissertation

Is beauty a sign of truth? Contrary to the thinkers surveyed above, I will argue that it is not, at least not in any a priori sense. I will also consider the related question of whether beauty is a worthy independent goal of science. Rather than respond to the above thinkers individually, I will consider a concrete case that will help to illustrate why beauty is not an a priori sign of truth. My strategy is both descriptive and prescriptive. I detail a previously unresearched episode in the history of biology in which it appears that beautiful symmetries play an essential role in the development and justification of a successful theory, the Caspar-Klug theory of virus structure. I argue that contrary to first impressions, the beautiful symmetries postulated by the Caspar-Klug theory are not signs of truth. Many of my arguments also work against the more general claim that beauty is a sign of truth.

The next two chapters, Chapters 2 and 3, describe the development of Donald Caspar and Aaron Klug’s theory from its roots in the work of Francis Crick and James Watson until its presentation in June of 1962 (Crick and Watson 1956; Caspar 1962).
Chapter 2 details the state of theoretical structural virology in the 1950s. I describe the Crick-Watson theory of spherical virus structure and also what was known about tobacco mosaic virus (TMV), the best-studied virus at the time. Crick and Watson gave theoretical arguments for why spherical viruses have icosahedral (or 532) symmetry. Chapter 3 details the collaboration between Donald Caspar and Aaron Klug that gave birth to the Caspar-Klug theory published in 1962. Luckily for the historian of science, much of the interaction between the British Klug and the American Caspar is documented in extant trans-Atlantic letters. Caspar and Klug postulate a subtle form of symmetry, called “quasi-equivalence,” that solved the problem of how to construct icosahedral viral shells from more than 60 identical protein subunits. Although there have been subtle violations of the theory, the Caspar-Klug theory still remains the broad theoretical framework for structural virology today (See Casjens (1997)). Once the history is presented, it is clear that Caspar and Klug did not appeal to the beauty of their theory to justify it. This early theory of virus structure also illustrates how a beautiful theory may have several properties whose possession individually is sufficient to render the theory beautiful. The Caspar-Klug theory is coherent, simple, and postulates highly symmetrical structures.

The philosophy of science continues in Chapter 4 where I argue that beauty is not a sign of truth. That is, I argue that there are no good reasons to believe that P(T/ T is beautiful) > 0.5 and some reasons to believe that P(T/ T is beautiful) ≤ 0.5. The argument is illustrated by the Caspar-Klug theory, which I argue is an example of a beautiful theory. A serious problem with the claim that beauty renders a theory likely is that there can be, and often are, multiple incompatible competing beautiful theories. Obviously, beauty cannot be a sign of truth for more then two competing beautiful theories. For the case of beautiful symmetries, I argue that properties of symmetry cannot be used to surmount this problem. I also consider whether Whewell’s notion of coherence can be a sign of truth. I argue Whewell is mistaken because there is no a priori reason to think there is one unique coherent theory.
Chapter 5 considers the weaker claim that beauty increases a theory’s probability, i.e., that beauty is a sign of truth. I argue that it is unlikely that beauty can play this role by examining the role of beauty in the three components of a theory’s probability—the prior probability, the likelihood, and the expectedness of the evidence—as shown by Bayes theorem. The philosopher James McAllister claims that we can use induction to justify an epistemic role for beauty. I argue that his proposal has major problems. I argue more generally that a priori facts do not influence the probability of an empirical theory. Types of beautiful properties that can be determined a priori thus do not increase the probability of a theory.

The sixth and final chapter considers whether there is a positive role for beauty in science. More specifically, I consider whether there is a role for beauty in the “logic of pursuit”. While I am skeptical of a general role for beauty in this context, the prospects are better than for a sign of truth. I suggest that society grants individual scientists the right to value theoretical beauty in a way they see fit. Many scientists value beauty in such a way that pursuing a beautiful theory is more reasonable than pursuing an ugly alternative, ceteris paribus. I also articulate some positive lessons from the Caspar- Klug theory for the role of symmetry in scientific reasoning. In biology, one may justify the existence of symmetries under certain conditions when we have reasons for thinking that evolution by natural selection is likely to generate symmetrical solutions to symmetrical design problems.

Before I begin with the historical narrative in the next two chapters, let me defend my choice to examine the Caspar-Klug theory of virus structure. There are two principal reasons why I am writing about this biological theory. First, the successful theory of virus structure devised by Donald Caspar and Aaron Klug posits highly symmetrical virus particles. Biologists judge it to be an especially beautiful theory. It is uncommon for biologists to leave their aesthetic pronouncements in print, but nonetheless, the Nobel laureate Jacques Monod pronounced that the theory was “esthetically pleasing.” He
appealed to the nature of the postulated theoretical entities to justify this aesthetic judgment. “The discovery of the dodecahedron and icosahedron (apparently made around 600 B.C., in one of the Greek colonies of southern Italy) is considered by Herman Weyl as one of the greatest and most beautiful in the whole history of mathematics. The viruses apparently made this discovery much earlier.” (Monod 1969, p. 23) Second, surprisingly given the success of the theory, the history of Caspar-Klug theory is not yet written. Here then lies a golden opportunity for the post-Kuhnian philosopher of science: to combine history of science with philosophical analysis.  

Since writing these historical chapters, Crick (2003) has written a brief account of his work on virus structure. Note footnote number 7. Also a comprehensive biography of Franklin was published, which although containing several scientific errors, contains a useful discussion the Franklin-Klug collaboration (Maddox 2002). Makowski (1998a) and Makowski (1998b) and Cohen (1998) offer brief accounts of Don Caspar’s achievements. Caspar has briefly commented on Klug’s work in light of Klug’s Nobel Prize (Caspar and DeRosier 1982). There has been some previous consideration of virology and aesthetics. Koch and Tarnai (1988) consider the question of whether viruses, which have the potential to cause disease, are worthy of aesthetic appreciation. As I have hope is clear, their project has very little to do with mine. Kemp (1998) has a one page article on the visualization of viruses. Kemp (2000) collects together his essays on science, aesthetics, and visualization.
Chapter 2  Early Structural Virology

Structure:

2.1  Point Symmetry in the History of Science

2.1.1  Point Symmetry in Plato’s *Timaeus*

2.1.2  Spherical Symmetry in Aristotle’s Cosmology

2.1.3  Point Symmetry in Kepler’s *Mysterium Cosmographicum*

2.2  The Beginnings of Structural Virology: Tobacco Mosaic Virus

2.2.1  The Early Tobacco Mosaic Virus Research of Donald Caspar

2.2.2  The Early Tobacco Mosaic Virus Research of James Watson

2.2.3  The Watson-Caspar Collaboration at Caltech

2.2.4  The Early Tobacco Mosaic Virus Research of Rosalind Franklin

2.3  The Beginnings of Structural Virology: Spherical Viruses

2.3.1  Five-fold symmetry in Tomato Bushy Stunt Virus

2.3.2  The Crick-Watson Theory of Spherical Virus Structure

2.3.2.1  The 1956 CIBA Conference on the Nature of Viruses

2.3.2.2  The 1956 Crystallography Meeting in Madrid

2.3.3  The Analogy between Viruses and Ribosomes

2.3.4  Aaron Klug’s Work on Turnip Yellow Mosaic Virus

2.3.4.1  Russell Steere’s Electron Micrographs of Turnip Yellow Mosaic Virus

2.1  Point Symmetry in the History of Science

In Chapter 1, I quoted various thinkers who have claimed that beauty or properties responsible for it are signs of truth and clarified what I mean by two different senses of “sign of truth”. The cynic might counter that it is one thing for scientists to make claims
about the epistemic role of beauty and another thing to actually “practice what you preach.” If, contrary to what they sometimes say, all scientists treat beauty as epistemically irrelevant, then the position against which I am arguing – that we are justified in believing that beauty is a sign of truth – is something of a straw-man. The question then becomes, do scientists utilize beauty in deciding which theories to believe and which theories to work on? An answer to this question would appeal to the history of science. Let me now then sketch a few episodes in the history of science where scientists have indeed appealed to beauty to justify their theories. Some selection procedure is necessary since the history of science is vast and our scope is limited. One way to select historical episodes to show that beauty has played an important role would be to look at the diversity of different types properties that scientists have taken to be beautiful. Another way would be to look at theories that share a particular type of property that has endured throughout history as one judged to be especially beautiful. To achieve some focus, I will take the second option and focus on one type of aesthetic property that has repeatedly influenced scientists over the centuries. The choice of which property to scrutinize is somewhat arbitrary. While there has been much philosophical ink spilt over the nature of simplicity, the role of symmetry has been relatively neglected (although there is growing work in the philosophy of physics, for example, the work of van Fraassen 1989 and Ismael 1997). Furthermore, symmetry is interesting in its own right. It is abstract enough to be explicitly posited by scientists from many different disciplines. A longer-term hope is that my dissertation could form part of a comparative study of the role of symmetry across different disciplines. For these reasons, let us now turn to a brief sketch of the power of beautiful symmetries in science, focusing on what is know as cubic point group symmetry.

13 Hargittai and Hargittai (2000) detail a number of scientists’ uses of symmetry in their science, including Donald Caspar and Aaron Klug.
14 Hoffman (1990) suggests that dodecahedrane (C_{20}H_{20}) is beautiful because of its icosahedral point symmetry. As this was first synthesized in 1982, I have not included it in the historical survey that precedes
Thinkers from the Ancient Greeks onward have proposed particular symmetries to underlay the structure of the world and its objects. Not coincidentally, the same symmetries often have been viewed as beautiful (See, for example, Engler 1990 and Maizer 1995). As Elaine Scarry, the Harvard scholar of aesthetics, puts it, symmetry is “the single most enduring recognized attribute” of beautiful objects (Scarry 1999, p. 96). It is thus not too surprising to find prominent thinkers who think that beauty caused by symmetry is a sign of truth. Consider the following survey of the more famous episodes in the history of science that have involved postulating beautiful symmetries.

2.1.1 Point Symmetry in Plato’s *Timaeus*

Plato, in his dialogue, *Timaeus*, suggests a “likely story” of the creation of the universe. In Plato’s Universe, the four elements are associated with four figures that consequently have become known as “platonic solids”: the cube with earth, the octahedron with air, the tetrahedron with fire, and the icosahedron with water. Plato associates the remaining platonic solid, the dodecahedron, with the Universe as a whole. The five platonic solids possess a high degree of point symmetry. Broadly defined, a symmetry exists in an object when there is an operation that leaves a relevant property of an object invariant through a transformation.\(^\text{15}\) Plato is concerned with geometric symmetries. Geometric symmetry concerns transformations of space, such as rotations, mirror inversions, and translations, that leave features of an object invariant. For example, there are four axes of symmetry in each of the platonic solids such that if we transform the figure with a 3-fold rotation along one of the axes, its appearance is invariant. More specifically, point symmetry involves a set of operations, all of which always map at least one point in the

\(^{15}\) Harold Osborne points out that the classical meaning of συμμετρεται was something like “balance” and is related to but significantly different from the modern conception of symmetry that I have adopted here (Osborne 1986).
object back on itself. Of all the geometric symmetry operations applicable to the platonic solids, the “center” point is always mapped back on itself. Plato uses his symmetrical ontology to explain certain supposedly observed features of the world. Since the icosahedron, the tetrahedron, and the octahedron are composed of equilateral triangles, Plato provided a mechanism for transmutation of water, fire, and air. The 3-dimensional “elements” could be decomposed into triangles and reassembled into different elements. Although unsuccessful, Plato’s theory set the precedent by which properties of matter are to be explained by geometric features of its constituents.

I maintain Plato accepted highly symmetrical objects into his ontology of nature, in large part, upon aesthetic grounds. He did not imbue his theory with beautiful symmetrical posits because this is a goal of theorizing, but rather because he thought that the natural world actually contained beautiful symmetries. He thought that (1) symmetrical objects are the most beautiful and (2) at the most fundamental level the world contains all and only the most beautiful objects. The *Timaeus* is a creation story. Unlike in the Judeo-Christian creation story, Plato’s creator, the demiurge, is not an omnipotent God who creates *ex nihilo*, but rather a “rational craftsman,” who creates order and beauty from the primitive chaos by imposing forms on it. The forms are independent of the demiurge and not modifiable by him; hence he lacks omnipotence. The demiurge, however, chooses the most beautiful and thus the most symmetrical forms to impose upon the chaotic matter. As S. Sambursky says, “[Plato] was convinced, in true Pythagorean spirit, that certain symmetries prevailed in the structure of matter…” (Sambursky 1962, p. 30) The historian of science, David Lindberg puts it this way, “[the demiurge] struggles against the limitations inherent in materials with which he must work to produce a cosmos as good, beautiful, and intellectually satisfying as possible.” (Lindberg 1992, p. 40) Although Plato appears to assume that “most good”, “most beautiful” and “most intellectually satisfying” are co-extensive, that the world’s most fundamental elements are highly symmetrical can be sufficiently explained by the demiurge creating most beautiful world. The demiurge provides a mechanism to explain
why the world is built from the most beautiful objects. He links beauty to truth. A true
theory of the world will reflect the beauty and symmetry created by the demiurge. Thus, for
Plato, beauty is a sign of truth in the empirical world. “We must thus assume as a principle
in all we say that god brought [the four elements] to a state of the greatest possible
perfection, in which they were not before.” (Plato, [360 BC] 1965, p. 53) Crudely stated,
for Plato the platonic solids with their high degree of symmetry represent the greatest
possible perfection and are thus eternal and thus beautiful. Plato believes that beautiful
symmetries are signs of truth.

2.1.2 Spherical Symmetry in Aristotle’s Cosmology

As is widely known, Aristotle, following Eudoxus and Callipus, presented a
concentric sphere model of the heavens. In the Aristotelian world picture, the spherical
Earth sits at the center of the universe. The fixed stars, each of the planets, the sun, and the
moon are all attached to separate spheres that rotate uniformly with respect to the Earth.
The more-complicated wandering motion of the planets is explained by having several
spheres embedded within one another so that the net effect as seen by an observer on earth
involves both “retrograde” and “direct” motion. The sphere is the most symmetrical of all
the point groups. A sphere has an infinite number of symmetry elements. Any straight line
through the center of a sphere defines a rotation axis of any degree. Likewise there is a
plane of mirror symmetry defined by any plane that intersects the center of the sphere. The
requirement that heavenly motion is explained only in terms of spheres is essential to
Aristotelian inspired astronomy. Astronomers for two millennia following Aristotle
respected and upheld the requirement of spherical motion in astronomy. Indeed, one could
argue that the unconditional insistence on spherical motion defines the field of astronomy
until Kepler. Admittedly, Ptolemy introduces complicating devices such as the eccentric, the
epicycle, and the equant, but nonetheless, he still maintains the essential use of spheres.
Even Copernicus maintains the use of spheres in his “revolutionary” heliocentric system.
Aristotle justifies his claim of sphericity in the heavens in his *de Caelo* or “On the Heavens.” Aristotle mixes a priori argument with empirical claims. First, from a priori principles he argues that there are only two simple or basic motions: straight-line motion and circular motion (Book I, Part 2). Further, a simple motion must belong to a simple body. Unlike straight-line motion, where “up” motion is contrary to “down,” circular motion does not have contraries and cannot change or decay since for Aristotle, change subsists in contraries. That which cannot change is perfect. Therefore, circular motion is “perfect.” Aristotle assumes that the perfect is “naturally prior to” imperfect, which I interpret to mean that if the imperfect exists, then the perfect must exist also. We know that bodies whose nature is to move in the imperfect motion of a straight-line – Fire, Earth, Air, and Water – exist. Therefore, a body whose nature is circular movement also exists. The question then arises: what simple body exhibits circular motion? At this point, Aristotle resorts to an appeal to observation: “…our eyes tell us that the heavens revolve in a circle.” (Book I, Part 5) The fifth simple body that constitutes the heavens Aristotle calls “Aither.” Although Aristotle appeals to observation, it is the appeal to perfection that carries the ontological weight. Singer expresses Aristotle’s and subsequent justification succinctly:

Circular motion is perfect since the circle is the perfect figure. Circular movement represents the changeless, eternal order of the heavens. It is contrasted with the rectilinear motion, which prevails on this our changing and imperfect Earth. (Singer 1959, p. 54)

To make my interpretation clear, I am claiming that perfection is an aesthetic notion and that a sphere’s perfection derives from its high degree of symmetry. Because of the high degree of symmetry any rotation of the sphere leaves the figure the same—it does not grow or decay as things do in the sublunary region of the Aristotelian universe. Aristotle then advocates a metaphysical principle that assumes that perfection is naturally prior to

16 Aristotle does not consider clockwise motion the contrary of counter-clockwise motion as one might expect.
imperfection. However, his metaphysical position has epistemic consequences: a theory that
postulates circular motion is more likely than one that does not. For Aristotle, maximal
point symmetry is a sign of truth in the superlunar region.

2.1.3 Point Symmetry in Kepler’s *Mysterium Cosmographicum*

Johannes Kepler, forced by empirical considerations, abandoned the use of the
sphere in astronomy and replaced it with the ellipse to explain the orbit of Mars. He was,
nonetheless, from his earliest researches, fond of highly symmetrical objects such as
platonic solids. In a well-known case of a failed theory, Kepler proposed that the five
nested platonic solids where successively interspersed with the (then-known) six planets.
The nesting of platonic solids purported to explain the relative distances of the planets.
Saturn’s orbit circumscribed a cube, which was inscribed by the orbit of Jupiter; into this
was inscribed the tetrahedron and so on down to the orbit of Mercury and the octahedron.
The fit with observational data, even in Kepler’s day, required some adjustment. Kepler
added thickness to the orbs and then centered the orbs on the Sun, not the center of Earth’s
orbit. Nonetheless, Kepler believed in this arrangement for the remainder of his life.

Kepler wrote the *Mysterium Cosmographicum* (1596) when he was 25 years of
age. In the preface, he writes of his process of discovery:

Within a few days everything fell into place. I saw one symmetrical solid after the other
fit in so precisely between the appropriate orbits, that if a peasant were to ask you on
what kind of hook the heavens are fastened so that they don’t fall down, it will be easy
for thee to answer him. (quoted in Koestler 1959, p. 253)

While this quotation does not explicitly link symmetry and beauty, the biographer of Kepler,
Max Caspar, has suggested “the aesthetic” as one of Kepler’s five “manners of
approach to the examination of the world.” Caspar explains that he means by Kepler’s
“aesthetic” a manner “which finds the principle of the beautiful primarily in symmetry.” (Caspar 1993, p. 67) Kepler was convinced that the true theory of the world had to be beautiful.

Kepler’s justification of his scheme had both a priori and a posteriori aspects. First, Kepler tried to prove his scheme a priori: he argued that (1) God could only create a perfect world and (2) the five platonic solids and six planets fit together perfectly. Second, Kepler appealed to the calculated relative distances between the planets. Unfortunately, Kepler’s scheme did not exactly fit the available Copernican data. Although the fit for Mars, Earth, and Venus was fairly good, it diverged significantly for Jupiter and Mercury. Kepler’s response was to adjust his model somewhat and to question the veracity of the data. How could such a beautiful scheme, the “Mystery of the Cosmos,” obviously divinely inspired, be wrong, he thought? Twenty-five years later in the dedication to the second edition of the Mysterium Cosmographicum (1621), Kepler wrote:

For as if a heavenly oracle had dictated it to me, the published booklet was in all its parts immediately recognized as excellent and true throughout (as is the rule with obvious acts of God). (quoted in Koestler 1959, p. 254)

Kepler never abandoned his scheme, always siding with the beautiful model even in the light of conflicting data. Given the conflicting data, his reasons for accepting his model must be partially non-empirical. No doubt Kepler’s theological beliefs are relevant, but I think that the aesthetic sensibilities he attributed to God represent the aesthetic ideal envisioned by Kepler.

17 No relation of Donald Caspar the biophysicist discussed later.
2.2 The Beginnings of Structural Virology

Let us now turn to a previously unresearched episode in the history of recent science: the Crick-Watson and the Caspar-Klug theories of virus structure. As with the above cosmological theories, icosahedral symmetry plays an integral role in the development of these important theories of virus structure. However, the narrative will show that there is no evidence that these scientists appealed to the beauty of their theories to justify them. Let me start at the end. In June 1962, Aaron Klug presented a paper, co-authored with Donald Caspar, entitled “Physical Principles in the Construction of Regular Viruses,” at the Cold Spring Harbor Symposium “Basic Mechanisms in Animal Virus Biology” (Casper and Klug 1962). Their central claim was that spherical viruses are structured like miniature geodesic domes. The theory they presented was widely accepted until the early 1980s when the first deviation from their theory was discovered (in Caspar's lab) (Rayment et al. 1982).

In the remainder of Chapter 2 and the entire of Chapter 3, I will narrate the events leading up to this important theory of virus structure. First, I will describe the development of structural virology in the 1950s. It was within this context that Crick and Watson and then Caspar and Klug began their respective collaborations in virology. Second, I will trace Caspar and Klug’s process of scientific experimentation and theorizing that lead to their famous theory of virus structure from 1960 to 1962. The roughly linear narrative follows the career of Donald Caspar, the first author of the famous 1962 paper, and introduces new characters as they interact with him.18

2.2.1 The Early Tobacco Mosaic Virus Research of Donald Caspar

In 1953, Donald Caspar began his PhD in biophysics at Yale University. Earlier he had majored in physics at Cornell and following graduation had secured a job as a health physicist at Yale. In New Haven, Caspar witnessed the work of biophysics graduate
students and sensed an excitement that had been absent for him in pure physics. Many of the biophysics graduate students were using radiation to inactivate biological systems, such as viruses, to determine the sizes of targets sensitive to radioactivity. Knowing a little x-ray crystallography, Caspar thought this technique might be a more constructive tool for probing biological structures. Caspar’s knowledge of crystallography came from one of the masters of the field: when he was 20 years old, Caspar had taken a two week long clinic at the Brooklyn Polytechnic Institution organized by a friend of the family, Isadore Fankuchen, a collaborator of J. D. Bernal. Six years later at Yale, Caspar originally planned to analyze solutions of spherical viruses, but quickly realized that there was little information to be easily gained from a solution of randomly aligned virus particles. (In a crystal or a fiber, the particles are arranged with more regularity.) Instead, Caspar focused on the well-studied tobacco mosaic virus (TMV) (Creager 2002). Rod shaped TMV could be oriented in fibers. The fibers are not true crystals since, although the central axis of each TMV particle aligns parallel to the fiber length, each virus particle is randomly rotated around its central axis.

Often overshadowed by the discovery of the DNA structure, 1953 was also an important year for protein crystallography. In this year, Max Perutz, while attempting to discover the structure of hemoglobin, made an important technical breakthrough with the application of the method of heavy metal isomorphous replacement to a protein molecule. Perhaps the most general problem in determining a structure by crystal or fiber x-ray diffraction is known as “the phase problem.” To determine the nature of the diffracting substance, one needs to know both the intensities and the phases of the diffracted x-rays. Unfortunately, measuring x-rays with film or a Geiger counter captures only the intensities and destroys the phase information. Perutz, working on hemoglobin, had shown that if one

---

18 Events involving Caspar that are not otherwise footnoted are based on interviews with him.  
19 For more on the role of protein crystallography at Brooklyn Polytechnic, see Berol (2001).
could change the structure slightly with the addition of a heavy metal, one could compare the
diffraction patterns of the heavy metal derivative and original sample and estimate the phases
(Green, Ingram, and Perutz 1954). Hearing of Perutz's innovation, Caspar began to create a
heavy metal derivative of TMV. After trying various heavy metal compounds, he found
those of lead to work the best. Caspar then estimated the phases of his pattern. At low
resolution the TMV fiber is centro-symmetric and in this special case, determining the
phases requires merely determining the sign of each phase, i.e., whether it is negative or
positive. Further, a fiber diffraction pattern is continuous in one dimension (unlike a crystal
diffraction pattern that is discrete) and the shape this continuous pattern helps with the
determination of the signs. Caspar collected his data over the period 1953-4. In December
he left Yale for Caltech to join Jim Watson who was also interested in viral structure.

2.2.2 The Early Tobacco Mosaic Virus Research of James Watson

Jim Watson, after determining the structure of DNA with Francis Crick, was now
seeking the structure of RNA. TMV contains RNA and consequently Watson hoped it
would be the key to understanding the structure of RNA. Watson’s pursuit of RNA had
begun two years earlier. During the spring of 1952, Lawrence Bragg, head of the
Cavendish Laboratory, declared a moratorium on Watson and Crick's work on DNA. He
considered it ungentlemanly to compete directly against other English scientists on the same
problem. Maurice Wilkins and Rosalind Franklin were working on the structure of DNA at
Kings College, London and they had priority according to Bragg. Readjusting his focus,
Watson began studying TMV. This change was more in emphasis than in intellectual
orientation. As Watson himself said, “A vital component of TMV was nucleic acid, and
so it was the perfect front to mask my continued interest in DNA.” (Watson 1968, p. 67)

---

20 Watson mentions that he first went to Cambridge to study DNA and plant viruses (Watson [1968] 1992,
p. 245).
That TMV might be made of subunits had been hypothesized by the German scientist Gerhard Schramm as early as 1944 (Schramm 1944). Watson in his Double Helix speculates that Schramm’s work did not significantly influence British workers because of the wartime prejudice against German science (Watson 1968, p. 68).

Crick taught Watson the rudiments of the newly developed helical diffraction theory. He called his lessons “helical diffraction theory for birdwatchers!” (Watson was an avid bird watcher.) Using his newly gained knowledge, Watson could see evidence for helices in Fankuchen’s superb x-ray patterns taken in 1938 and published in 1941. He decided to get more experimental evidence for himself. Watson obtained purified TMV from Roy Markham at the Molteno Institute and with the help of Hugh Huxley obtained mediocre x-ray diffraction patterns from imperfectly oriented dry para-crystalline specimens. Unlike the earlier work by Bernal and Fankuchen, Watson and Huxley tilted the TMV specimen to obtain patterns that, in principle, indicate the number of subunits in the helical repeat. Using the newly developed theory of the diffraction of helical structures, Watson, with help from Crick, inferred that TMV was a helix that had an integral number of subunits in three turns of the helix, possibly 31 (Cochran, Crick, and Vand 1952; Watson 1954). The central idea of Cochran, Crick, and Vand (1952) was that there would be certain reflections in the diffraction pattern that are not permitted if the diffracting particle is helical. The useful result for Watson was that for a helix with \( n \) residues per repeat, a \( J_0 \) Bessel function contributes to the \( n \)th layer line. (The TMV diffraction pattern consists of a series of parallel “layer lines.”) Watson’s difficulty was that he could not distinguish between a \( J_0 \) and a higher order Bessel function on the basis of his data. His estimate of 31 subunits per 3 turns of the helix would be later corrected by Franklin to 49. Watson speculated about the reason for the helical nature of TMV. He drew an analogy with Frank’s theory of crystal
Frank thought that new molecules might be added to the “cozy corners” of a growing crystal. (Think of the corners at the base of each step of a helical staircase.) Watson saw that a growing helix could be thought of as a series of cozy corners.

Alongside speculations about the rod shaped TMV, Crick and Watson developed analogous hypotheses about the structure of spherical viruses. The problem of determining the structure of TMV involved the question of how to arrange subunits according to helical symmetry (i.e., a screw axis consisting of a rotation and a translation). Crick and Watson considered an analogous problem for spherical viruses. Given the appropriate projection, the arrangement of TMV subunits can be considered as a problem of determining the correct line group. Likewise, determining spherical virus structure was a matter of determining which point group best describes their symmetry. As mentioned earlier, a point group is a collection of symmetry elements (rotation axes, mirror planes, etc.) that intersect at a single point.

Caspar heard of Watson’s work on TMV and met him at the summer phage meeting at Cold Spring Harbor in 1954. He told Caspar about Crick and his ideas regarding spherical virus substructure and symmetry. Briefly, they hypothesized that the small “spherical” viruses must have cubic symmetry. Cubic symmetry involves having at least four 3-fold rotational axes – rotating the virus 120 degrees (360 degrees divided by 3) around any of the four different rotation axes yields an indistinguishable object. For example, all the platonic solids have cubic symmetry. Imagine a cube: if you look down the body-diagonal, you are looking down a 3-fold rotation axis and there are four such axes. In December 1953, a year before he finished his PhD, Caspar pitched a post-doctoral research proposal to George Beadle of Caltech. He proposed examining solutions of southern bean mosaic virus (SBMV), a spherical virus, using the low angle scattering apparatus (a

---

21 Watson interview.
22 Watson, personal communication, 1 April 2002.
point focusing monochromator) of Prof. J. DuMond of the Physics Department. Watson, now a Senior Research Fellow at Caltech supported Caspar’s application, indicating he was “willing and anxious to help Caspar to make his project a success.” Caspar’s appointment at Caltech was dependent on his gaining of an outside postdoctoral fellowship. Although Caspar did not receive an NRC or NSF fellowship, he was awarded a U.S. Public Health fellowship.

2.2.3 The Watson-Caspar Collaboration at Caltech

By September, most of Caspar’s PhD research was finished and in December of 1954, Caspar traveled to Caltech where Watson was working on the structure of RNA. Working with Alex Rich, Watson, a postdoctoral fellow of Max Delbrück, had attempted to make fibers of RNA since the fall of 1953 using techniques that had proved successful for DNA. Before coming to Caltech, Watson had tried and failed to take an informative diffraction pattern of plant RNA purified by Roy Markham (Watson 2002, p. 46). Results in Pasadena were also disappointing. During the months of January and February, Watson and Rich tried with little success to take informative x-ray photographs of purified RNA (Rich 1995). They obtained RNA from colleagues and also tried to measure the size of TMV RNA using the Spinco analytical centrifuge (Watson 2002, p. 130). The diffraction data did allow for the possibility of an RNA helix, but they were inconclusive and even consistent with RNA being a branched polymer (Rich and Watson 1954a; 1954b). The most that Watson and Rich could conclude was that RNA was “DNA-like,” perhaps having bases stacked on one another, but “this suggestion has many difficulties” since there was nothing analogous to Chargaff’s rules for RNA (Rich and Watson 1954c).

24 Memo from Max Delbrück, 23 December 1953, Caltech Archives, Bio Division 21.33.
25 In retrospect, Rich and Watson were considering ribosomal RNA that does contain some double-helical segments.
Watson and Rich could not determine if helical sections of RNA existed or even if it was a single or double chain molecule. They did manage to show “pretty reversible change in the RNA fiber length that occurs upon raising or lowering the relative humidity.” After six month of such “frustration,” Watson and Rich gave up working on “naked” RNA (Watson 2000, p. 25). In June of 1954, Watson put the situation this way: “Our work on RNA is a standstill. We need a cute idea or a much better x-ray photograph and neither possibility seems in the air. The main paradox is that the base ratios are not complementary (in some cases) while the x-ray diagram upon reflection is really quite DNA like.”

Perhaps the structure of TMV RNA would prove to be more tractable than naked RNA Watson hoped. It was in this context that Caspar arrived at Caltech. While there, Caspar analyzed his data on TMV: his goal was to calculate a cylindrically averaged radial mass distribution function for TMV. This involved calculating the signs (i.e., phases) of his peaks of intensity (Franklin was also calculating signs for her TMV data around the same time). The resulting mass density distribution showed peaks of density at a radius of 22Å and 42Å from the center of the TMV “cylinder” (see Figure 2.1).

---

27 Letter Watson to Delbrück, 1 June 1954, Caltech Archives 23.23.
Figure 2.1 Radial Density Function of TMV relative to density of water.
Most significantly there was no significant density in the center of the cylindrical TMV: the virus appeared to be hollow! Using this data, Watson and Caspar speculated about the structure of TMV RNA. At first Caspar took the innermost peak to be the RNA. They then speculated about how the RNA was wound within the protein shell: “For the RNA case we favor a 10-12 stranded model in which the RNA chains follow the same helical grid as the protein.” At the time, the best estimates of the molecular weight of the TMV RNA suggested that there must be more than one piece of RNA in each particle. On the Caspar-Watson model, as in later models, the length of RNA determines the length of the TMV particle. Watson also flirted with the idea that the RNA formed a double stranded ribbon in which two RNA strands were linked by pyrophosphate bonds between the phosphates of the RNA backbones. It is difficult to point their reasoning more succinctly than they did in the Caltech Biology annual report:

From a biological considerations it seems desirable to have the RNA chains run the entire length of the virus. Otherwise it is difficult to imagine why the virus length is constant, and why the infectivity always accompanies the parties of 3000Å length. The fact that the molecular weight of isolated TMV RNA falls rapidly from that of the intact core to units of molecular weight between $2 \times 10^5$ to $3 \times 10^5$ suggests that the core is composed of a number of RNA chains. … We consider therefore the arrangements with strand numbers of nine or greater as more attractive. All models with a single chain per lattice point must have the chains running in the same direction; otherwise the crystallographic sub-units would not be identical. Crystallographic evidence suggests that the TMV particle is non-polar, and the RNA core can only be non-polar if there are two chains per lattice point, running in opposite directions. Two chains are possible with a strand number of twelve, and this gives us a twenty-four-chain core in which the chains are fully

---

extended with the maximal phosphate-phosphate separation of 7.5 Å. (Biology Division Annual Report, Caltech Archives)

In retrospect, their estimated molecular weight of RNA was too low and TMV particles are polar with only one strand of RNA running through the entire length of the virus particle.

Watson was fully aware of the aesthetic consequences of his model. He wrote to Franklin describing it:

The main thing in favor of the P-O-P [pyrophosphate] model is that it is very very pretty steriochemically. But does nature always like to be pretty?30

Franklin was skeptical of the model: she replied that she thought the RNA might as well be a disordered core as far as the x-ray data are concerned.31

Caspar was also interested in working with spherical viruses, and traveled to Berkeley to obtain some samples of tomato bushy stunt virus (TBSV or BSV) from Art Knight. Unfortunately, as it turned out, the diffraction equipment at Caltech did not include a sufficiently powerful x-ray source to properly analyze virus crystals (Caspar would need a rotating copper anode x-ray source). Furthermore, Sturtevant had designed the Californian x-ray equipment to be earthquake proof and consequently one could not make the fine adjustments necessary to measure the closely spaced diffraction spots. Nonetheless, Caspar had logged 388.6 hours of x-ray tubes in Crellin Laboratory time by the time he left Caltech.32 Presumably most of this time was spent analyzing the southern bean mosaic virus (SBMV or SBM), although he also briefly examined tomato bushy stunt virus. With the limited equipment at Caltech, Caspar was able to discover little more about SBM than a rough estimate of the diameter (282Å) and even less about BSV.33

29 Letter Watson to Franklin, 9 April 1955, Churchill Archive.
30 Letter Watson to Franklin, 9 April 1955, Churchill Archive.
31 Letter Franklin to Watson, 10 June 1955, Churchill Archive.
32 Memo Sturdivant to Delbrück, 19 October 1955, Caltech Archives, Bio Division 21.23.
In the late summer of 1955, both Caspar and Watson traveled to England. Caspar arrived in Cambridge August 9. Caspar’s friend Harold Bloom arranged for a room for Caspar in Pembroke College, but Caspar discovered that the gas to the Cambridge laboratories was turned off over the inter-term break. Caspar decided to use the time before Michaelmas Term to tour continental Europe. He returned to London to meet Rosalind Franklin for the first time on September 12. They had been previously corresponding about the structure of TMV. Together they attended the opening performance of the Japanese Azuma Kabuki dancers and musicians in Covent Garden.

2.2.4 The Early Tobacco Mosaic Virus Research of Rosalind Franklin

Rosalind Franklin had been working on TMV since 1953 when she left King’s College and joined J. D. Bernal’s lab at Birkbeck College. The story of Franklin’s work on DNA at King’s is well known and I will not repeat it here. (See Watson 1968 and Sayre 1975) In March 1953, she began at Birkbeck, but there were delays due to difficulties in obtaining the necessary apparatus and samples. Finally, in November, Roy Markham of the ARC Molteno Institute sent her samples of TMV (as he had earlier supplied Watson in 1952). Randall, her former boss at Kings College, informed Franklin she was not to work on any helical biological material, but Bernal reassured her that virus crystallography was his “property” so she should go ahead. The helical TMV was to be the first of several viruses Franklin hoped to look at. The goal of determining the structure of viral RNA and its relation to protein was clearly of central importance to Franklin and her group. As she wrote in March 1956 in an attempt for continued Agricultural Research Council (ARC) funding:

36 Letter Markham to Franklin, Nov 23, Churchill Archive.
[Our] work is concerned with what is probably the most fundamental of all questions concerning the mechanism of living processes, namely the relationship between protein and nucleic acid in the living cell … The plant viruses consist of ribonucleic acid and protein, and provide the ideal system for the study of the in vivo structure of both ribonucleic acid and protein and of the structural relationship of the one to the other.38 Surprisingly her ARC funding was discontinued, although with the help of Caspar she was later able to gain money from the United States Public Health Service.

On the 13th September 1955, Caspar returned to Cambridge. After getting settled, Caspar was in search of more virus samples. Later in the month, he traveled with Peter Pauling in his sporty Porsche to Rothamstead Experimental Station to meet Fred Bawden and Bill Pirie. They gave him some BSV preparations in solution and crystalline form. They also told him that they had given Harry Carlisle crystals of BSV and turnip yellow mosaic virus (TYMV) years ago and that these samples should be still stored in the refrigerators at Birkbeck College. Caspar returned to Cambridge, started growing more BSV crystals, and analyzing the larger crystals on Tony Broad’s powerful rotating anode x-ray tube.

In November Caspar traveled down to London to look for the samples on ice at Birkbeck College. To his surprise, Franklin would not turn over the TYMV samples. Watson exaggerates when he claims that Caspar and Franklin got in a “verbal fight” over the crystals (Watson 2002, p. 188). She was saving TYMV for her colleague Aaron Klug.

Aaron Klug (1926- ) is a Lithuanian-born South African-raised British crystallographer. As a student at Durban High School he read the classic work Microbe Hunters by Paul de Kruif and consequently decided to study medicine. However, at the University of Witwatersrand, through the study of physiological chemistry, Klug discovered

38 Franklin (1956) “Note on the Future of the ARC Research Group in Birkbeck College Crystallography Laboratory” Folder 2/36, Churchill Archive.
that he wanted a deeper foundation so he shifted focus, first to chemistry, then to physics. Klug enrolled for an MSc at the University of Cape Town. Here he met R. W. James, a crystallographer originally from Bragg's school at Manchester. Klug learned crystallography from James, in large part by checking the proofs of James' classic book, *The Optical Principles of the Diffraction of X-rays*. In 1949, Klug left for Trinity College, Cambridge to work on “unorthodox” (e.g., protein) x-ray crystallography at the Cavendish Laboratory. Unfortunately there was no room in the MRC unit where Max Perutz and John Kendrew worked. Instead Klug completed his PhD on the kinetics of phase changes in solids under D. R. Hartree. At the end of 1953, Klug moved to Birkbeck College, London. Here Klug met Rosalind Franklin and began working with her on the structure of TMV. However he was not her post-doc, but had a Nuffield fellowship that allowed him some flexibility in choosing his research project. In fact, it was almost by chance that Klug began working with Franklin. Franklin and Klug shared adjoining rooms in the top of 21 Torrington Square, a house converted into laboratories. He had begun to work with Harry Carlisle on the structure of ribonuclease, but work was very slow and uninspiring. Serendipitously he met Franklin on the stairs of Torrington Square while she was carrying TMV diffraction photographs and was drawn to her “fascinating” work.39

By late 1955, Franklin had managed to obtain a radial density distribution for re-aggregated TMV protein using material supplied by Schramm. By comparing the density with that obtained by Caspar, she was able to conclude that the TMV RNA was neither in the center nor 20Å from the center as Caspar had thought, but lay 40Å from the center. Franklin wanted to publish the comparative analysis but first required that Caspar publish his results. Caspar procrastinated so Franklin wrote a first draft of Caspar’s paper based

39 Klug Interview.
on his dissertation herself. The first draft was finished February 10th. These results were also presented at the 1956 CIBA meeting (Franklin, Klug and Holmes 1957).

2.3 The Beginnings of Structural Virology: Spherical Viruses

Tomato bushy stunt virus (BSV) and turnip yellow mosaic virus (TYMV) were natural choices for the crystallographers Caspar and Klug. From the point of view of structural biology, BSV and TYMV were the two best-studied spherical viruses. J. D. Bernal and Isadore Fankuchen had first considered BSV from the point of view of x-ray diffraction in 1938. Bawden and Pirie prepared BSV crystals that proved to be too small for single crystal x-ray studies. Instead Bawden and Pirie took powder diffraction measurements to calculate the dimensions of the unit cell (Bernal, Fankuchen and Riley 1938). TYMV was also investigated by Bernal at Birkbeck, this time in collaboration with Carlisle. In 1948 they took powder photographs of TYMV purified by Kenneth Smith and Roy Markham (Bernal and Carlisle 1948). They also considered a nucleic acid-free “top component” which they determined was “for practical purposes” the same size as the virus particle. Bernal and Carlisle judged TYMV to be in a diamond-type lattice with 8 particles per unit cell. A diamond lattice is one in which the four nearest neighbors of any given particle are on the vertices of a tetrahedron in the same way that carbon atoms are bonded in diamond.

Klug wanted to work on spherical viruses in part so he would have his own project independent of Rosalind Franklin’s work. He chose TYMV despite the fact its crystal had a larger unit cell than BSV (700 Å compared with 380Å) and thus was probably a more difficult project. Klug was aware of Roy Markham’s work that had shown that during the purification of TYMV one could distinguish a “top component” as well as the infectious

---

40 Caspar Interview.
41 Letter Franklin to Watson, 10 February 1956, Norman Archive.
virus (Markham 1951). Top component was so named because it lay at the top of a sucrose gradient. It was hypothesized that the less dense top component was merely the protein component of the virus. Some had gone so far to suggest that the top component might be hollow viral shells that do not contain nucleic acid like the denser “normal” virus particles (Schmidt, Kaesberg and Beeman 1954). Klug hoped that in the longer term he would be able to compare the x-ray diffraction diagrams of normal virus and top component and infer some structural information about the viral RNA.43

2.3.1 Five-fold symmetry in Tomato Bushy Stunt Virus

Results with BSV came quickly for Caspar, but they were unexpected. First, using a partially disordered crystal, he obtained 10 “smudges,” which to a crystallographer surprisingly suggested 5-fold rotational symmetry. He repeated the results and showed that there were “spikes” in the diffraction that more conclusively indicated 5-fold symmetry (see figure 2.2) (Caspar 1956). The name “spike” was coined by Perutz who helped Caspar refine his manuscript.44

42 Klug Interview.
43 Klug Interview.
44 Caspar Interview, Perutz Interview.
Figure 2.2 Left Caspar’s diffraction picture showing spikes in the 5-fold direction. Right idealized data.
2.3.2 The Crick-Watson Theory of Spherical Virus Structure

Serendipitously, Caspar had x-rayed the virus crystal down one of the five fold axes of the virus. His results indicated that the virus has icosahedral symmetry (called 532 symmetry by crystallographers). Of the platonic solids, only the icosahedron (20 triangles) and the dodecahedron (12 pentagons) have 532 symmetry. These unexpected results supported Crick and Watson's hypothesis that viruses have cubic symmetry and Caspar showed his results to Crick, who happened to be working in the same small building known as "the hut." (Icosahedra and dodecahedra possess cubic symmetry as well as 5-fold symmetry.) With the knowledge of Caspar's result, Crick and Watson rewrote a theoretical article they had been working on and submitted it to Nature (Watson 2002, p. 123; Crick and Watson 1956). They had a draft of the article written the year before, which Crick shortened for Nature and adapted in light of Caspar's results. This article was adapted sometime after September 20 when Crick and Rich submitted a manuscript on collagen (Crick and Rich 1955; Rich 1998). They argued that "a virus possessing cubic symmetry must necessarily be built from a regular aggregation of smaller asymmetrical building bricks and this can only be done in a very limited number of ways." (Crick and Watson 1956, p. 474) These ways correspond to the three cubic point-groups. A point-group is a mathematical entity consisting of symmetry elements that intersect at a point. A virus with the symmetry of a cubic point group is necessarily made up of a multiple of 12 and a maximum of 60 identically situated subunits. Caspar wrote to his first mentor in crystallography in January 1956:

My results on Bushy Stunt are quite interesting: the single crystal X-ray photographs indicate that the virus has the point group symmetry 532 and thus it must be built out of

---

45 Watson Interview. See also Watson (2002) p. 123.
46 Crick Interview.
47 Watson Interview.
sixty asymmetric sub-units. I will be reporting this at the Easter meeting of the
International Union of Crystallography meeting in Madrid.\textsuperscript{48}

Caspar’s experimental results for BSV were published in \textit{Nature} immediately following
Crick and Watson's article (Caspar 1956).

Crick and Watson’s suggestion had been made previously as they acknowledge in a
footnote. In 1950, the crystallographer, Dorothy Crowfoot Hodgkin\textsuperscript{49} wrote the following
in a Cold Spring Harbor Symposium (Volume XIV):

Lately, remarkable photographs of stationary wet crystals [of BSV] have been obtained
by Carlisle and Dornberger (1948) which literally give millions of X-ray reflections
extending to spacings of 7.5 Å; these still give no evidence of crystal symmetry. If the
symmetry is genuinely cubic, two alternative conclusions are possible; the molecules may
be oriented statistically in all directions to simulate spherical symmetry or they may
\textit{consist of 12 n identical submolecules} each of weight say 750,000/n (Hodgkin 1950, p.
67; italics added).

Hodgkin’s student Barbara Low (1953) in a book edited by Hans Neurath and Kenneth
Bailey called “The Proteins”, wrote:

If the symmetry [of the lattice of tomato bushy stunt virus crystals] is not due to
twinning ... it must depend upon the presence in the virus particle of 12 n identical
subunits related to each other by the symmetry class of the cubic crystal class. (Low
1953, p. 314)

Although these analyses mention complicating factors such as crystal twinning not
mentioned by Crick and Watson, Hodgkin and Low’s comments did not have the same
impact as Crick and Watson’s treatment. Part of the difference is due to emphasis.
Hodgkin and Low make their remarks in passing as part of a larger review. On the other

\textsuperscript{48} Letter Caspar to Fankuchen, 28 January 1956, Norman Archive.
\textsuperscript{49} For an excellent biography of Hodgkin, see Ferry (1998).
hand, Crick and Watson devote an entire article to Hodgkin’s two-sentence remark and also make the connection with the cubic point groups more explicit. They also suggest that if icosahedral symmetry is present then there will be 60n subunits per virus particle. However, this is not a case of independent discovery. Watson and Crick were aware of the work of Hodgkin and Low and personally knew them both.

2.3.2.1 The 1956 CIBA Conference on the Nature of Viruses

In March of 1956, Crick presented their ideas to a group of virologists at a small CIBA colloquium (Crick and Watson 1957). Founded in 1949, The CIBA Foundation held small informal meetings of elite researchers from around the world in their elegant formerly residential establishment at 41 Portland Place. One unstated goal of the symposium originators, F. G. Young, was to “revivify” virology in England.  

This meeting involved British, American, Australian, South African, German, and French scientists interested in viruses. The major centers for basic viral research were represented: Cambridge, Berkeley CA, Birkbeck, and Tübingen. Franklin petitioned to have Schramm invited, but the Tübingen Max Planck-Institut laboratory was represented by Schäfer instead. Maurice Wilkins was scheduled to attend but it appears that his place was taken by Aaron Klug at the last minute. Of the 34 participants, six went on to win Nobel prizes.

At least half of the time in the symposium was allocated for discussion after each talk and this discussion was transcribed and published following each paper (Lee and Spufford 1993, p. 44). As the director of the foundation wrote at the time of its inception;

50 Letter Wolstenholme to Franklin, 21 June 1955, Churchill Archive 2/34; Letter Burnet to Lady Burnet, 13 March 1956, Burnet Papers, 2/18, University of Melbourne Archives.
51 Letter Franklin to Wolstenholme, 28 June 1955, Churchill Archive.
52 The CIBA/Novartis scrapbook has a printed list of participants in which Wilkins’ name is crossed out and Klug’s is written in by hand.
The intention is to bring together within its walls, in an atmosphere of a private home, small groups of scientists from all over the world, for moderately informal conferences, and, probably more fruitfully, for friendly private discussions when “off-duty.”

(Gordon Wolstenholme, quoted in Lee and Spufford 1993, p. 44).

Figure 2.3 Michael Stoker, Jim Watson, Milton Salton, and Francis Crick at the 1956 CIBA symposium.

Crick presented the first paper at the conference. Although much of the material had been previously presented in the *Nature* article, Crick did consider a new question: Why does a virus have subunits? His answer was both elegant and remarkably simple. To begin, Crick made a number of assumptions. First, he assumed that a virus consists of RNA surrounded by a protein “coat”. Second, he assumed that the amino acid sequence of protein coat is determined by the molecular structure of the RNA. Finally, he argued for a small constant “coding ratio”—the number of nucleotides that code on average for one amino acid. Crick then argued that given the relatively small size of a viral genome, and a coding ratio of say 3:1, then there is *not* enough information in the viral genome to code for a large number of non-identical subunits. Therefore, Crick concluded, there must be a single protein subunit (or a small number of them) repeated a number of times. This argument complements the earlier consideration of the number of subunits and how are they
arranged. It is also one of the first arguments (if not the first) to use the idea of quantifiable genetic information.

The CIBA symposium proved to be a meeting of the old and the new. The new were represented by Crick, Watson, Franklin, Caspar and Klug, and others who were also convinced that the new physical techniques would spawn a new molecular biology founded on information containing nucleic acids. The old were mainly traditional virologists, often medically trained and wedded to less informative immunological and biochemical techniques. The organizers of the meeting, M. van den Ende and F. G. Young, strove to bring these two groups together with the hope of meaningful exchange. The meeting was originally entitled the “Biophysics and Biochemistry of Viruses”—presumably the term biophysics referred to x-ray crystallography and electron microscopy and the term biochemistry referred to the more traditional virology. This intermixing of backgrounds is an explicit goal of the CIBA (now Novartis) Foundation colloquia:

[The colloquia] have a distinctive identity: limited to c. 25 invited participants, they are specifically designed to bring together scientists with differing backgrounds, attitudes and skills who have a common area of interest, and provide a suitable environment for the generation of new ideas (Novartis 2000, p. 4).

This particular colloquium, the 38th that CIBA had hosted, was a mixed success. A reviewer for Acta Medica Scandinavica wrote in 1957:

As Ciba Symposia go, this one may be considered as of medium quality. The sponsor’s intention was to bring together the physicists and chemists on the other hand and the biologists on the other for a heart to heart talk on subjects of mutual interest. From this point of view the meeting was only moderately successful, as the two groups largely confined their activities to their own respective spheres.
As Watson put it, “I do not think either Robley Williams or Knight from the Berkeley lab would have got excited [about our paper].”\(^{53}\) This is not to say that the material in the papers would not be useful to virologists and doctors alike. A reviewer in the *Lector Medicus* remarked

Some remarkable and exciting suggestions occur in the discussions, such as inducing the virus to enter the protoplast, of the use of Puck’s clonal studies for the purpose of manipulating virus-host interaction and there is much thought provoking material throughout the volume … this is one of the most important volumes that has emerged from this excellent organisation.

The review in the *British Medical Journal* (10.5.57) judged the Crick and Watson paper to be one of the three most impressive of the 17 papers presented. The other two being Gard’s paper on the kinetics of inactivation of polio by formaldehyde, and Bang and Issacs’ paper that showed electron micrographs of cells infected with mumps and influenza. Interestingly, even three years after the discovery of the DNA structure, Crick and Watson were not seen to be the authorities: A reviewer in the *Journal Pathology and Bacteriology* wrote, “The value of this publication is in great part due to the excellent record of informal discussions by such authorities as Burnet, Andrewes, Lwoff, Bawden, Wilson Smith and others.”

One must remember that even three years after the 1953 paper on the structure of DNA its importance was not fully appreciated. Nonetheless, the tide was turning. Macfarlane Burnet, an Australian participant in the symposium wrote to his wife after meeting with Watson and Crick: “The Watson-Crick formula for DNA is almost as famous now as Kekule’s ring formula of benzene.”\(^{54}\)

---

\(^{53}\) Watson Interview.

\(^{54}\) Burnet to Lady Burnet, 13 March 1956, Burnet Papers 2/18, University of Melbourne Archives.
Perhaps the most surprising result of the meeting was the news that nucleic acid alone could be responsible for an infection. Robley Williams mentions this work of his Berkeley colleague in an off-hand way towards the end of his paper on the electron microscopy of viruses:

It now begins to appear (Fraenkel-Conrat, 1956) that infection can be obtained from solutions in which essentially no full length TMV particles, either native or reconstituted, are present. The active solutions are believed to be pure RNA, and are infectious when rubbed upon tobacco plants in sufficiently high concentrations.

(Williams 1957, p. 31)

To the surprise of Watson and Crick, some of the participants did not appreciate the gravity of the finding or found it hard to accept. Bawden, in the question session, asked for quantitative results and more experiments to control for other enhancing or inhibiting substances in the inoculum. “Unless we have answers to such questions such as these, how are we to know what value to attach to the exciting statement at the end of your talk?” (Wolstenholme and Millar 1957, p. 35) Watson explained the root of the difference in opinion: “They were not at home with the concept that information flows unidirectionally from nucleic acids to protein and never backwards.” (Watson 2000, p. 28) The skeptical Bill Pirie said of Crick and Watson’s paper, “it was not as bad as I thought”—not exactly a glowing complement. Even the presenter of the important news, Robley Williams did not seem to appreciate the gravity of his finding and was consequently the subject of a practical joke by Jim Watson. Watson faked a telegram to Williams. It purported to be from Wendell Stanley himself and said, “TMV Protein infectious—be cautious.” (Watson 2002, p. 217) Even later when told of the hoax, Williams admitted he was still open to the possibility that genes may be made of protein. Williams’ skepticism is also apparent in the

55 Similar results were independently obtained by Schramm around the same time.
56 Crick Interview
question session after Crick’s talk. He asks, “But are you not going to get into geometrical difficulties if you say that the RNA codes all of one subunit? How does the RNA expose itself to the whole of the subunit?” To which Crick replies that the RNA takes an extended form to code for an extended form of the protein, which then folds up. This idea later became to be known as the “sequence hypothesis” and frames one the ongoing problems in molecular biology—the protein folding problem: what principles govern how the extended protein folds correctly?
2.3.2.2 The 1956 Crystallography Meeting in Madrid

A week later, Caspar, Klug, and Franklin traveled to Madrid to attend the International Union of Crystallography Easter meeting (See figure 2.4).

Figure 2.4. Madrid International Union of Crystallography meeting 1956. From left to right, Anne Cullis, Francis Crick, Don Caspar, Aaron Klug, Rosalind Franklin, Odile Crick, and John Kendrew. Image courtesy of Don Caspar.

Caspar presented a paper that combined Crick and Watson's theoretical speculations and his experimental results. The paper entitled “The Molecular Viruses Considered as Point-Group Crystals” illustrates how Crick, Watson, and Caspar now thought of viruses as a type of crystal (Caspar, Crick and Watson 1956). They conceived of the virus as consisting of identical subunits bonded together in identical ways. Each subunit is equivalent to every
other subunit, as is true of crystals. The negative analogy, to use Mary Hesse’s terminology, is that viruses are spatially bounded, whereas “space-group” crystals potentially extend to infinity (Hesse 1966). The maximum number of equivalently related subunits in a cubic point-group crystal is 60—think of the 20 triangular sides of an icosahedron each divided into three.

The idea that viruses behave like crystals goes back at least to Stanley’s well-publicized “crystallization” of TMV in 1935. What is new here is to view the components of a virus as crystallizing into a well-formed virus. Crick mentions this possibility at the CIBA conference: “The process of aggregation [of subunits into a virus] is one which you might reasonably call crystallization.” (Crick and Watson 1957, p. 17) The paper then developed a “formal crystallographic classification of viruses.” (Caspar, Crick, and Watson 1956) Caspar experienced some resistance from crystallographers to the idea that viruses possess 5-fold symmetries. As every student of crystallography learns, true 5-fold lattice symmetry is impossible. (The discovery of quasi-crystals that possess statistical five-fold symmetry was not known at the time.) Caspar was not proposing that the lattice itself had five-fold symmetry, but that the viruses, which sit on the cubic lattice points, have five fold symmetry. It is quite possible to crystallize a molecule with more symmetry in a low symmetry lattice. Nonetheless, five-fold symmetry in molecules was extremely rare and many in the audience remained unconvinced by Caspar’s presentation. One crystallographer who was more easily convinced was William Lipscomb who had discovered icosahedral symmetry in certain classes of Boron molecules.

---

57 Technically, Stanley did not create true crystals of TMV since they were not 3-D crystalline.
58 Creager (2002) argues that interest in the pieces of viruses was a 1950s development, with the exception of Germany where it occurred in the 1940s. Bernal however considered the possibility that TMV was some sort of finite crystal.
Klug presented after Caspar. He had been trying to discover the selection rule to express the Fourier Transform of the icosahedral point groups in polar coordinates. Only the abstract survives:

This problem arose in connection with spherical viruses. The transform for a general position in each of the point groups is expressed as a series of products of spherical harmonics and Bessel functions. The general character of the transforms will be discussed. (Klug 1956)

This project can be seen as applying the same methodology to spherical viruses that Cochran, Crick, and Vand (1952) did to helical viruses. Later Klug would extend this treatment of helical diffraction as well (Klug, Crick, and Wyckoff 1958). Klug’s novel applied mathematics illustrates the attempt to adapt an approach that was successfully applied in helical viruses to the spherical virus case. However, unlike the theory of helical diffraction, Klug’s theory proved very difficult to use, especially in pre-computational crystallography.

2.3.3 The Analogy between Viruses and Ribosomes

In 1956 Franklin and Klug flirted with the idea that microsomes (now called ribosomes) might have a structure similar to RNA viruses. The reason for thinking there might be a connection was that both viruses and microsomes were made of protein and RNA. The problem was that they needed some material to study. Klug wrote to Howard Schachman at Berkeley to see if he had any material. Unfortunately, he did not. Klug then wrote a letter to Bernal: “We need some advice rather urgently … Do you know of anybody who might be prepared to follow Schachman’s procedure?” Eventuality, Klug and Franklin were able to secure some fresh preparations of microsomes from Mary

59 Letter Klug to Schachman, 21 May 1956, Norman Archive.
60 Letter Klug to Bernal, 31 May 1956, Norman Archive.
Peterman and J. L. Simkin as well as from Schachman himself. They worked with an element of subterfuge, fearing that they might be scooped. If there were a connection between virus structure and microsomes it would be a major discovery. Franklin on her second trip to the United States wrote back to Klug:

I forgot to discuss with you how much I should tell people about our recent things. However, as Jim’s [?] better to tell too much than too little I told Jim about the microsomes. He seemed a bit surprised that we were able to get them … Jim intends to work entirely on microsomes.61

It was difficult to obtain conclusive data. By the end of July Franklin put their progress this way:

I really don’t think we have anything yet to justify publication. The most we could say is that the low angle reflections [?] indicate a regular material—but they are so weak that even this is hardly justified.62

Their first results were published first in 1958. Using powder diagrams, the Birkbeck group concluded that the RNA and protein in microsomes was very different than a mixture of RNA and protein suggesting that “the configuration of the RNA in some way closely conforms to that of the protein.” (Franklin et al. 1958) By 1960, they revised the earlier interpretation and showed that although the protein components of microsomes from different species are very similar, viruses and microsomes share only a “superficial resemblance.” (Klug, Holmes and Finch 1961, p. 87) The disanalogy between spherical viruses was further strengthened by Hugh Huxley who took high-resolution micrographs of microsomes that showed that they were not spherical at all.

61 Letter Franklin to Klug, 21 June 1956, Norman Archive.
62 Letter Franklin to Klug, 27 July 1956, Norman Archive.
2.3.4 Aaron Klug’s Work on Turnip Yellow Mosaic Virus

Klug’s progress on turnip yellow mosaic virus (TYMV) did not progress as fast as Caspar’s progress on tomato bushy stunt virus (BSV). As stated earlier, the larger unit cell of TYMV meant that longer exposure times were needed to get useful diffraction patterns. Klug however ran into a further difficulty. By June 1956, he was coming to the belief that there were 16 particles in the unit cell rather than the 8 proposed by Bernal and Carlisle. Furthermore, half the particles lay in a different orientation than the other half. (At lower resolution, the difference in orientation is difficult to see.) This meant that instead of getting 10 spikes of intensity as Caspar had with BSV, Klug and Finch had two superimposed sets, each of 10 spikes, which demanded more interpretation to “see” the 5-fold symmetry of TYMV.

To justify this doubling of the proposed number of particles in the unit cell, Klug sought to measure the nucleic acid content of the crystals using UV radiation. One needs to know the thickness of the crystal and the refractive index of the crystal of TYMV. Klug wrote to Oster of the Brooklyn Polytechnic to obtain this information. By July, Klug and Walker of Kings College had succeeded in measuring the refractive index of 1.49-1.50, which corresponded to a value of 19 particles per unit cell. It appeared that the unit cell of TYMV crystals did contain 16 particles. Consequently, Klug and Finch proposed a “double-diamond” arrangement for the TYMV crystal, so called because it consists of two intermeshed diamond-like lattices.

Two of Rosalind Franklin’s students—John Finch and Kenneth Holmes—also contributed to virus research. After studying physics at King’s College, John Finch decided to enter research in a biological/biophysical field. Looking through *Nature* for an appropriate job in the spring of 1955, he discovered that Rosalind Franklin at Birkbeck

---

63 Letter Klug to Oster, 25 June 1956, Norman Archive.
64 Letter Klug to Markham, 6 July 1956, Norman Archive.
College was advertising for a research assistant. After securing the job, he began work on calculating the relative humidity of TMV gels, work similar to that Franklin had done on DNA gels at Kings College.\textsuperscript{65} This work was never published. Ken Holmes joined the group in late summer, around the time Don Caspar arrived in Cambridge, wanting to work on spherical viruses. Klug, Caspar, and Franklin looked through the refrigerators at Birkbeck and found crystals of BSV and TYMV. Ken Holmes continued working on TMV while John Finch, in collaboration with Klug, began work on TYMV. Officially Bernal was Finch’s advisor, but although he was interested and would poke his head into the lab momentarily to utter “anything new?” he was not involved with the TYMV project. Franklin and later Klug would fill the role as Finch’s mentor. Finch began by taking x-ray photographs of TYMV using a Birkbeck-made x-ray tube and later switched to a Hilger x-ray tube that had better focus capabilities. Although he was quite happy as a technician, Franklin decided to begin him and Ken Holmes on parallel PhD programs.

2.3.4.1 Russell Steere’s Electron Micrographs of Turnip Yellow Mosaic Virus

In the summer of 1956, Franklin traveled to the United States and visited many biologists around the country. The most useful part of her trip was her stay at the Berkeley Virus Laboratory. Here she worked with Fraenkel-Conrat on heavy metal derivatives of TMV. Perhaps the most surprising results she saw in Berkeley were some EM photographs taken by Russell Steere. He used a freeze-shadowing replica technique, which was better able to illuminate the substructure of TYMV. Franklin wrote an excited letter to Klug:

The most important thing here is that I’ve recently seen some electron micrographs of TYM by Steere’s freezing-shadowing replica technique which show a magnificently clear surface structure. This only happened today, and seems very exciting, so excuse the

\textsuperscript{65} Finch Interview.
muddle. The prominent feature is an array of six knobs on each particle around a central one. ... The inter-knob distance is \( \sim 1/4 \) inter-particle distance, which is consistent with your 60A, but the thing does not look 5 folded ... If it is not 5-folded, the question arises, was the five foldedness in the RNA. ... Believe it or not, he says [Robley Williams] says he had a slide of this with him at Ciba but did not show it in case the effect was due to amm. sulphate!!

If TYMV has 532 symmetry and contained 60 subunits, then Franklin assumed one would to see a significant fraction of rings of five, i.e., when one looks down a 5-fold axis. Three days later she sent Klug some of the electron micrographs and a more considered opinion.

I have looked at these and others—particularly ones which show particles in a wide variety of orientations, and there seems very little doubt that the knobs lie at the vertices of a cubeoctahedron, i.e., 12 knobs. … Among hundreds of particles I have only found 2 that look remotely 5-folded and I certainly do not believe they are all like that. She wrote a similar letter to Don Caspar on the same day. The significant feature of the cubeoctahedron—a semi-regular solid formed by either truncating the corners of a cube or its dual the octahedron—is that it has 432 symmetry not 532 symmetry like Tomato Bushy Stunt. A week later Franklin was still thinking about Steere’s electron microscope photographs, however now from the point of view that the photographs pertained to the determination the unit cell of the crystal. When the viruses packed together they seemed to be in the same direction—not two different orientations as Klug now thought the unit cell contained.

By December Klug and Franklin had a paper on TYMV ready for publication that Francis Crick thought “reads very well indeed.” Crick suggested that they publish the

---

66 Letter Franklin to Klug, 27 July 1956, Norman Archive.
67 Letter Franklin to Klug, 30 July 1956, Norman Archive.
68 Letter Franklin to Klug, 5 August 1956, Norman Archive.
69 Letter Crick to Klug, 14 December 1956, Norman Archive.
paper in *Biochimica et Biophysica Acta* rather than *Nature* due to the long length and technical nature of the article. Klug, Finch, and Franklin submitted a short less technical paper to *Nature* on January 11 (Klug, Finch, and Franklin 1957b). This paper presented the crystal structure of TYMV with 16 virus particles per unit cell and suggested that like BSV, TYMV possessed 532 symmetry. A month later, the three submitted the longer paper to *Biochimica et Biophysica Acta* (Klug, Finch and Franklin 1957a). In the penultimate section, they address Steere’s electron micrographs. They suggest that the high percentage of ammonium sulphate (50% by weight) in the samples used by Steere may have introduced artifacts not present in the x-ray crystallography. Nonetheless, they offer no explanation of ammonium sulphate would cause virus particles to appear to have 432 symmetry. Both electron micrographs taken by Kaesberg using a different technique (shadow casting) and their work agree that the symmetry is 532 or icosahedral.
Chapter 3  The Caspar and Klug Collaboration

Structure:

3.1 The Beginning of the Caspar-Klug Collaboration
   3.1.1 The Insight of the Physicist Richard Crane

3.2 Aaron Klug’s Work on Polio Virus

3.3 The Connection with Art: John McHale

3.4 Buckminster Fuller and “Synergetic” Geometry
   3.4.1 The Meeting Between Fuller and Klug

3.5 The Continued Collaboration between Donald Caspar and Aaron Klug

3.6 Improved Electron Micrographs of Turnip Yellow Mosaic Virus
   3.6.1 The Invention of Negative Staining

3.7 The Birth of Quasi-Equivalence
   3.7.1 The Selection Rule: $h^2 + hk + k^2$
   3.7.2 Two Projects in Theoretical Structural Virology: Geometrical and Physical
   3.7.3 The 1962 Cold Spring Harbor Meeting on Animal Virus Biology

3.8 Michael Goldberg’s Rediscovered Insight

3.9 Historical Summary

3.10 Some Preliminary Philosophical Lessons

3.1 The Beginning of the Caspar-Klug Collaboration

Although Caspar and Klug were now in close contact, their first collaborative venture was forged in the light of a great tragedy. In August 1958, Rosalind Franklin was scheduled to talk in Bloomington, Indiana at a plant pathology meeting organized to celebrate the 50th anniversary of the American Phytopathological Society. Franklin
described the meeting as a “silly” meeting although she was looking forward to seeing her American friends and collaborators.\textsuperscript{70} She would have spoken on her recent work on TMV. Tragically, Rosalind Franklin died of ovarian cancer on April 16 1958 at the age of 37. She had known that she had cancer for two years, but it came as a shock to many of her acquaintances as she had kept her ailment secret from many. Under these unfortunate circumstances, Caspar and Klug began their collaboration. Caspar was invited by the committee to speak in Franklin’s place and he suggested to Klug that they collaborate. They began with some rough notes that Franklin had written. Caspar wrote to Klug:

I think perhaps we might start off with the idea of ‘molecular viruses’ and of subunits in viruses with some discussion of the early work indicating subunits. Next the concept of regular packing of subunits, symmetry and viruses as ‘point group’ or ‘line group’ crystals. Then a discussion of the power of X-ray diffraction methods in studying regular structures….\textsuperscript{71}

Caspar extricated himself from a previous engagement at an Army Chemical Warfare Symposium so he could travel to England to work with Klug on the paper.\textsuperscript{72}

They wrote a review of the x-ray diffraction of viruses and submitted it to be published toward the end of July\textsuperscript{73} (Franklin, Caspar and Klug 1959). The paper, which Bawden called “comprehensive and lucid,” reviewed much of the x-ray crystallography, a technique that Caspar and Klug felt was superior to alternative approaches.\textsuperscript{74} “X-ray diffraction is at present the only method available for studying the internal three-dimensional configuration of viruses and other highly organized, biological structures in the hydrated native state.” (Franklin, Caspar, and Klug 1959, p. 457) They noted that there were now more viruses known to exhibit icosahedral symmetry: as well as tomato bushy

\textsuperscript{70} Letter Franklin to Sayre, 8 October 1957, UMBC Archive.
\textsuperscript{71} Letter Caspar to Klug, 13 May 1958, Churchill Archive.
\textsuperscript{72} Letter Caspar to Klug, 14 May 1958, Norman Archive.
\textsuperscript{73} Letter Holton to Caspar, 24 July 1958, Folder 3/19, Churchill Archive.
stunt and turnip yellow mosaic virus, tobacco ringspot virus and perhaps coxsackie, an animal virus possess 532 symmetry. They also reported that Robley Williams using electron microscopy had also shown that a large insect virus, tipula iridescent virus (TIV), not only possessed icosahedral symmetry, but also was shaped like an icosahedron (Williams and Smith 1958). At the time, TIV did not seem to be part of the same class as the small RNA plant viruses—it was 100 times larger by weight.

As a review of the state of the field, the paper did not contain very much theoretical speculation beyond Crick and Watson’s 1956 ideas. However, they explain the structural similarity between BSV and TYMV in terms of efficiency:

The structural correspondence between these two significantly different virus particles is probably not fortuitous, but is likely a reflection of the fact that this type of cubic symmetry is a very efficient way for nature to build a compact particle out of smaller protein subunits (Franklin, Caspar, and Klug 1959, p. 457).

They do not mention competing explanations for the correspondence between the two viruses such as homology of descent. A further advance beyond the simplest Crick and Watson model was driven by recent biochemical work on the number of molecules in the viral shell of BSV:

End group analyses on TYMV, as well as for BSV, indicate that there are at least 120 protein molecules in the virus and that therefore the structural subunit (asymmetric unit) may contain two chemical subunits (Franklin, Caspar, and Klug 1959, p. 457).

Caspar and Klug refer to work done by J. Ieuan Harris and Art Knight and others from the Berkeley Virus Laboratory who used gastric enzymes to nibble the end residue from each protein subunit in the virus and then used the number of cleaved amino acids to determine the number of protein chains per molecule (Harris and Knight 1952; Niu, Shore, and Knight 1958). It was this work that began to suggest that the Crick-Watson theory was

74 Letter Bawden to Caspar, 2 Dec 1958, Bawden Papers, Royal Society Archives.
incomplete. As it turns out, rather than 120 protein molecules per subunit estimated by this procedure, there are 180 protein molecules per virus particle. Regardless, the important point is that a number greater than 60 subunits per virus calls for a richer picture of virus assembly and additional principles than those given by Crick and Watson. Presumably, there might be a multi-stage process of virus assembly. For example, first chemical subunits might come together in pairs and second the bonded pairs assemble into a viral shell, but clearly there are other options also. In ending the paper, based on the cases of TMV, BSV, and TYMV, Caspar and Klug indulge in generalization about how all small viruses (and perhaps other “particulate nucleoproteins”) are put together:

…it seems likely that the parts made by a type of subassembly process before being assembled to build the virus. The forces holding together the protein subunits in the virus particle are like those of a crystal. The configuration of the RNA is determined by its regular packing with the protein (Franklin, Caspar, and Klug 1959, p. 458).

Thus, a crystal of virus particles is constructed by at least three levels of processes akin to crystallization.

As is indicated in the above quoted concluding remark, at the time Klug and Caspar began to collaborate, Klug was concerned with how the RNA is arranged within the protein shell. Klug thought that the RNA might be arranged so as to have tetrahedral (23) symmetry or nearly so. He wrote to the famous geometer H. S. M. Coxeter:

I am … writing to ask whether it is possible to construct a continuous closed curve having tetrahedral symmetry … From trials, I rather suspect this is not possible, but wondered whether there is possibly a theorem on the subject. Coxeter replied that the point groups for knots are the trivial ones and the symmetry cannot be 222, 23, 432 or 532.

75 Klug Interview.
76 Letter Klug to Coxeter, 30 April 1958, Norman Archive.
Klug published his musings on these subjects with Finch in 1960 (Klug and Finch 1960). It was difficult to gain experimental evidence on the structure of RNA in TYMV. By comparing TYMV with RNA with TYMV without, Klug and Finch suggested that RNA possess “at least in part, some regular structure.” They suggest at most it could be apparent 23 symmetry or a lesser symmetry, which lead statistically from the crystal packing to the appearance of 23 symmetry. Alternatively they speculate that the “raison d'etre” of spherical viruses might be to contain many different RNA molecules. This suggestion would allow for the RNA to have higher symmetry than a single chain of RNA.

3.1.1 The Insight of the Physicist Richard Crane

Klug and Caspar mention the work of Richard (Dick) Crane as giving another reason for why viruses are made of subunits (Crane 1950). Watson had told Caspar about Crane’s work in 1954. Klug and Caspar do not commit themselves to whether Crane’s ideas would be a competing or a complimentary explanation to Crick and Watson (1957). The article that Crane called “mainly speculation, not research” was published in a companion popular science magazine for the American Association for the Advancement of Science (AAAS). In this article, Dick Crane had argued that assembly processes involving sub-processes would be more efficient than those that didn’t because mistakes that arose in a subassembly could be more easily discarded and not incorporated into the final product. A similar argument is found in Herbert Simon’s Sciences of the Artificial (Simon 1969, p. 90). Simon introduces a parable of two watchmakers: the successful Hora and the unsuccessful Tempus. Tempus makes watches out of 1000 parts and if interrupted during putting a watch together it falls apart and he had to begin from scratch. Hora, on the other hand, makes 10 subassemblies of 100 parts; each of the 100-part subassembly is itself

---

77 Letter Coxeter to Klug, 3 June 1958, Norman Archive.
made from 10 sub-sub-assemblies of 10 parts. If Hora is interrupted he often does not lose a watch but merely a subassembly or a sub-sub-assembly. Crane’s prescient idea is practically identical:

We may take a suggestion from the modern factory. The modern art of the mass production of automobiles, houses, and radios, and other articles by the technique of subassemblies may seem at first sight to be quite far removed from the question of the assembly of protein particles … [S]ignificant advantages of the subassembly method stand out: The culling out of defective units can be done at each level, so that a defective unit produced at one level does not get built into a larger unit and so make it defective in turn … (Crane 1950, p. 388)

Dick Crane earned his PhD in physics from Caltech in 1934. While on the physics faculty at University Michigan he temporarily became involved with the biophysics group that included Robley Williams and Cyrus Levinthal among others. His reading Oparin on the origin of life on Earth piqued his interest in the problem of replicating molecules. He wrote his influential article in 1950 while traveling to Pasadena by train for a sabbatical. The analogy with the factory floor above derives from Crane’s effort during WWII. He was involved with the development of the “proximity fuse.” The Johns Hopkins University’s Applied Physics Laboratory in Silver Spring, MD, developed the proximity fuse during the war—it was basically a small radio in the nose of an anti-aircraft shell that detected the presence of aircraft and exploded the shell. Miniature vacuum tubes were made for the fuse by rows of unskilled women. Mistakes made by any one woman were discarded and not used in the next assembly operation. Crane’s idea’s have more recently influenced researchers on the protein-folding problem and was reprinted in a 1983 AAAS volume. As Donald Wetlaufer the editor of the volume remarked, “So much of what Crane logically inferred, before the facts were known, has proven correct.” (Wetlaufer 1983)

79 Simon had not read Crane when he wrote his book. Simon personal communication.
Crane’s contribution to the early history of molecular biology goes beyond the importance of the efficiency of subassembly processes. His 1950 article contained a second important idea. He showed that any structure built from identical subunits making two determinate identical contacts each would be repetitive along a screw axis. In other words, structures made by adding bivalent subunits according to the rule that every “bond” between subunits is identical will lead to a helical structure. Crane illustrated his idea with matchbox models. Imagine a small spiral staircase made of matchboxes. Each matchbox is related to the next matchbox by the same angles and distances—the result is a spiral structure, or more correctly, a helix. Crick visited Michigan in the late 1940s and late 1950s and Crane and he talked. Watson also read Crane (1950). In many ways, 1950s molecular biology could be considered the decade of the helix: the alpha helix, the double helix, and the helical nature of TMV were all major discoveries. On the other hand, Crick has remarked, “[Crane’s] helix idea is certainly correct but also fairly obvious, so I am not sure how much influence it had.” Crick’s remark should be taken with a grain of salt since he had thought more about helices than almost any other person—in the decade he published two theoretical papers on helical diffraction (Cochran, Crick, and Vand 1952; Klug, Crick, and Wyckoff 1958). Crick and Watson’s 1956 papers and Caspar and Klug’s later work can be considered a generalization of Crane’s project. Whereas Crane considers a subunit that can bind two others and then considers linear chains, Caspar and Klug later consider subunits that can bind three or more other subunits in nonlinear lattices that form closed spherical structures.

80 Crane Interview.
81 Watson, personal communication, 1 April 2002.
82 Letter Crick to the author, 2 February 1999.
3.2 Aaron Klug’s Work on Polio Virus

In 1957, Franklin and her co-workers, Klug and Finch, as well as continuing to study TYMV, began to study poliovirus. Polio was the first animal virus crystallized. Schaffer and Schwerdt from the Virus Laboratory at U. C. Berkeley succeed in crystallizing the virus in 1955 (Schaffer and Schwerdt 1955). They purified virus from 15 liters of monkey kidney tissue culture. Later they crystallized Type 1 (Mahoney strain). Crystals large enough for single crystal x-ray diffraction took a year to grow. Some of these were given to Rosalind Franklin in 1957. Polio crystals were discussed at the International Poliomyelitis Conference in Geneva in July where Franklin heard Schwerdt give a talk on the crystals he had grown.\(^8\)

Some of the people working in the Birkbeck laboratories were not pleased that Franklin was working with poliovirus. To them, her work posed an unacceptable risk of infection. They convened a meeting regarding her use of polio at Birkbeck, which resulted in Franklin being banned from using polio in her lab. Consequently the crystals were taken from Birkbeck to the London School of Hygiene and Tropical Medicine (over the road) where there existed better facilities for dealing with highly infectious pathogens. Here was a safer place to mount the crystals. Franklin also needed an x-ray tube if she was to take x-ray diffraction pictures. She contacted Bragg at the Royal Institution which by then had a rotating anode x-ray tube copied from the design of the tube at Cambridge. Franklin asked Bragg if he would mind polio crystals being analyzed there and he was quite receptive to the work. Klug had to get permission to follow guideline rules, but there were no rules and so he invented his own and submitted them to an inspector.\(^4\) When crystals were finally successfully mounted in capillaries, they were transported in a Thermos across town to the Royal Institution.

\(^8\) Franklin Notebook entry, 10 July 1957, folder 3/14, Churchill Archive.
\(^4\) Finch Interview.
Franklin’s attempts to mount the polio crystals in capillary tubes in early to mid 1957 were unsuccessful. Her crystals would spontaneously dissolve in the capillary before she could get any diffraction pictures. Franklin attributed this instability of the crystals to an alkaline reaction occurring in the boro-silicate glass of the capillary. (Salts leaching out of the glass.) She then tried using acid treated capillaries, which delayed the dissolution of the crystals, but not by enough for her to get any data. In the month before her death, she wrote to Bawden suggesting that Pyrex tubes might be better. Franklin also wrote to R. W. Douglas of the department of Glass Technology at Sheffield who recommended she use “neutral glass.”

Klug discovered quartz capillaries to be more suitable for maintaining the crystals. Quartz does not have the same alkaline reaction as boro-silicate and is now widely used by protein crystallographers. Klug and Finch’s second technical innovation was to keep the crystals at 5°C while they were diffracting x-rays. They blew air cooled by dry ice over the capillary. The lower temperature decreased x-ray damage to the crystals and thus increased their useful exposure time. Klug and Finch began x-ray diffraction work on polio after the death of Rosalind Franklin. The first still photographs were “beautiful.” They then began collecting precession photographs. They had their results by December and a draft of the paper by February 1959.

Unlike TYMV and BSV, polio did not crystallize in a cubic lattice but rather an orthorhombic one. Nonetheless, Finch and Klug were able to infer that polio, like the plant viruses, possesses icosahedral symmetry. This result was especially significant since it

---

88 Finch Interview.
89 Letter Crick to Klug, 22 December 1958, Norman Archive.
90 Letter Klug to Crick, 13 February 1959, Norman Archive.
showed there were no major structural differences between animal and plant viruses.\textsuperscript{91} They speculate that polio and TYMV have the same structure, i.e., 60 subunits at the corners of a snub dodecahedron. They illustrate this point with a ping-pong ball model (See figure 3.1). Caspar and Crick in Cambridge built such ping-pong ball models as early as 1955. This putative structure was not given without some (as it turns out) prudent qualification:

The point group symmetry does not preclude the possibility of each structural (crystallographic) subunit being further subdivided in a number of smaller chemical units (Finch and Klug 1959, p. 1713).

Given the lack of biochemical data on the size of the polio chemical subunit, Klug and Finch assume the simplest model consistent with the x-ray data.

In the final paragraph, Finch and Klug speculate on more theoretical grounds. They classify Crick and Watson’s speculation as \textit{a principle of economy}. The virus uses one stretch of DNA again and again for up to 60 identical subunits (or perhaps a multiple of 60). They propose another “perhaps subsidiary” principle that they call a \textit{principle of efficiency}. This principle provides a reason why of the three cubic point groups, the icosahedral is preferred.\textsuperscript{92}

If [a virus particle] is required to ‘enclose’ space around a point by a symmetrical arrangement of small identical units, it can be shown that the ratio of the number of sub-units to the volume enclosed is the smallest if icosahedral symmetry is employed (Finch and Klug 1959, p. 1714).

Crick knew this point prior to his 1956 article. It probably formed one of the reasons that Crick thought it was “likely” that the symmetry was 532.\textsuperscript{93} Klug and Finch’s paragraph, however, makes this point explicit as Klug acknowledges in a letter to Crick.

---

\textsuperscript{91} This similarity has been shown to be also borne out at the molecular level. Plant viruses and animal viruses often share the same protein folds in their capsid proteins (Baltimore, 1985).

\textsuperscript{92} For more on the relation between economy or efficiency and symmetry see Toth (1986).

\textsuperscript{93} Letter Crick to the author, 3 March 2000.
The maximisation or minimization (of some relevant quantity) is nearly always more favourable for arrangements where symmetry is possible that for those in which it is not. The results are, what one might have guessed. But it does seem, though this point is not made explicit anywhere, that among the class of symmetric arrangements, those with icosahedral symmetry are the best. This is the basis of the statement made into the last paragraph of the enclosed paper.

Finch and Klug’s work was published in *Nature* in June of 1959. Immediately the popular press took notice of their work. Polio research in the 1950s was newsworthy. *The Manchester Guardian* printed a half page story, entitled “The Architecture of Viruses,” on June 30, 1959. The article was written by “Scientific Correspondent” John Maddox. The article included a large photograph of Franklin’s large TMV model (made for the Brussels International Fair) and also discussed the possible structure of polio:

In short, it appears that an individual poliomyelitis virus has the symmetry of an icosahedron … From this it has been possible to conclude that the virus is built up of sixty individual units arranged and held together in such a way that they form a shape with the correct symmetry. This is where guesswork of a kind is necessary, but the Birkbeck workers have put forward the suggestion that the sixty units of the which the virus is built are individual molecules each of which is practically speaking the same and possessed of no symmetry of its own. (p. 6)

However, it was a shorter article, “New Light on How Polio Starts,” which appeared on page 13 of *The Observer* on June 21st that probably was more important for the history of structural virology. This article did not discuss the helical viruses such as TMV, but

---

94 Letter Klug to Crick, 13 February 1959.
95 Note the use of the word “architecture” in the title before any connection was noted between domes and viruses.
96 Letter Maddox to Klug, 26 June 1959, Norman Archive. Maddox would later become the editor of *Nature*. 

75
concentrated exclusively on the structure of polio including a picture of a 60 ping-pong ball model and an icosahedron with three subunits distributed on each face (See figure 3.1).
The unnamed scientific correspondent wrote:

It is now becoming clear that geometry plays an important part in virus structure. The reason for the particular geometry displayed by the spherical viruses is probably that this icosahedral arrangement is the most economical way of “packing” the small protein units around a central core. This can be proved mathematically. (p. 13)

The mention of efficient packing is reminiscent of Buckminster Fuller who bases much of his theory of architecture on the closest packing of spheres.
3.3 The Connection with Architecture: John McHale

One of Buckminster Fuller's European popularizers, the pop artist John McHale (1922-1978), read one of the newspaper reports and noticed the potential similarities between viral structure and Fuller's geodesic domes. He wrote a letter to Fuller:

I held off writing for some time in case anything of importance came up to communicate about your visit … I enclose a cutting which may be of interest. I have written to the authors to ask where a complete text may be found—in ‘Nature’ 20th June. I mentioned the reason for my enquiry—that I thought it might be of interest to you—they replied ‘that they know of and admire your work ….[sic] and if you are interested would be glad to discuss their researches with you personally.’

Fuller came to London for the third week of July and he met with Klug and Finch (Marks 1960, p.44).

It is not too surprising that John McHale saw this connection between viral structure and geodesic domes once one understands McHale’s background. In the 1950s, he was a member of “The Independent Group,” (IG) a group of rebellious young artists and art critics based in London who promoted a new conception of art that focused on the aesthetic value of technology and popular culture (Massey 1984; Robbins 1990; Massey 1995). It was one of the Independent Group members, Lawrence Alloway, who coined the term “pop art” in the mid-1950s, although the direct influence of the Independent Group on later American pop artists such as Andy Warhol is controversial (Alloway 1966; Massey 1995). The Independent Group began a series of informal meetings under the auspices of the Institute of Contemporary Arts (ICA), although ironically in many ways the group reflects a reaction against elitist purist modernist art supported by the ICA and its leaders Herbert Read and Roland Penrose. The Independent Group organized a number of exhibitions
including the famous 1956 *This is Tomorrow* exhibition mentioned in most histories of early pop art (See, for example, McCarthy (2000)). Two of the Independent Group’s exhibitions are important for our purposes.

The first was called *Growth and Form* after D’Arcy Thompson’s same-titled 1917 book. 98 This exhibition was organized by Independent Group members Nigel Henderson and Richard Hamilton. On the initial draft proposal of 20 December 1949, Hamilton proposed:

> Large scale models and enlarged micro-photos to show the wide range and beauty of visual material of modern science. (Quoted in Massey 1984, p. 135)

Hamilton’s display incorporated films of sea urchin embryology, microphotographs, and x-ray micrographs, which were projected onto screens. The show consisted of seventeen categories ranging from atomic physics to astronomy. From a financial point of view, the exhibition was a failure: only 1140 people visited the exhibition and consequently it lost £223 (Massey 1995, p. 45). Some critics could not see the point of the exhibition:

> Instead of sending their [The ICA’s] members to the Natural History Museum, they have now decided themselves to display what appealed to their organisers’ aesthetic sense or seemed to them related to modern art (Burlington Magazine 1951)

Massey argues that it was the anti-teleological element of Thompson’s classic book that appealed to the young artists. Thompson rejected the idea that one explains the living world in terms of an end, purpose, or function, as Aristotle does with his final cause, and as one unorthodox reading of Darwin does. That is, one should not appeal to the function that a trait is selected for to explain it (fully). Rather Thompson proposes that we need only a more proximate causal explanation—roughly, the relation between growth and form—an

---

97 John McHale to Fuller, 3 July 1959, Chronofile, Fuller Papers, Stanford.
98 As Sharon Kingsland pointed out to me, it was probably the expanded 1942 edition of *On Growth and Form* that IG members read. Among other things, Thompson considers Albrecht Dürer’s use of perspective and thus a connection between biological form and art was already explicit in his work.
explanation that now would be the province of developmental biology. The Independent Group took themselves to be rejecting Aristotelianism in art: in particular, the idea artwork has an eternal essence. It is unclear how well they understood Aristotle’s worldview however, since their conception of Aristotelianism was shaped by a reading of the radical Alfred Korzybski’s “Non-Aristotelian Systems” (Korszybski 1933). They took Thompson to be also rejecting the Aristotelian idea that the natural world should be explained in terms of essences, or at least teleological essences, an anti-Aristotelian thesis congenial to their goals in the art world. The Independent Group’s exposure to Korzybski’s ideas derived from an interest in science fiction: A. E. Van Vogt’s 1948 classic *The World of Null-A* is based in “Non-Aristotelian” principles, hence the name “Null-A”.

In many ways, the exhibition *Growth and Form* was pivotal in the formation of the Independent Group. It explored new sources of art outside the pre-war modernist tradition and displayed the IG’s fascination with science and technology.

The exhibition spawned a book edited by Lancelot Law Whyte (Whyte 1951). The book, entitled *Aspects of Form*, “presents the first general survey of visual form, from physics through biology and psychology to art.” It included articles by prominent thinkers such as C. H. Waddington, Joseph Needham, Konrad Lorenz, and E. H. Gombrich. Ironically, the preface was written by Herbert Read and there is no mention of Richard Hamilton’s organization of the *Growth and Form* exhibition.

The second exhibition relevant to this narrative, *Parallel of Life and Art*, was shown in from September 11 until October 18, 1953 at the Architectural Association in London. One motivation of this exhibition was to challenge viewers’ perceptions of what was worthy of inclusion in an art exhibition (Massey 1995, p. 57). They displayed images drawn from scientific and technical publications including diagrams and photographs of radio valves, televisions and spacesuits. The photographs were not classically pretty but included pictures of benign tumors and rats. This disregard for conventional aesthetic taste formed the basis of what IG member Banham called “The New Brutalism”. Interestingly, included
in the exhibition were diffraction patterns taken by Birkbeck College scientists and an
electron micrograph of rabbit muscle fiber taken by the microscopist Robert Horne.99

John McHale was not a member of the IG at the time of the *Growth and Form*
exhibition, but he came to have similar views to *Growth and Form* organizer Richard
Hamilton. There was no official membership for the IG; rather it was a loose-knit group of
friends, roughly defined by repeated attendance at meetings. A. J. Ayer gave one early talk
to the IG on the principle of verification. Logical positivism was another tool the IG hoped
to use to sweep away grand aesthetic principles, replacing them with a new aesthetic based
on popular culture and technology. The first set of nine more organized seminars began in
October 1953. These seminars were open to just under 100 members of the ICA and the
fee-paying public. For example, on October 29, IG member Richard Hamilton spoke on
“New sources of form.” The abstract of the talk remains:

The growth of new canons of form through the continual extension of a visual horizon
under the impact of micro-photography, long range astronomy etc. Relevance of the
forms discovered to contemporary design procedures and to mathematical laws of
structures and statistics (reprinted in Massey (1995) as appendix 1).

It was about this time McHale, through his friend Lawrence Alloway, joined the IG. In a
manner similar to Hamilton, McHale expressed how he saw the relationship between
technology and the IG in a 1970s interview:

…the [period of] the fifties was characterized by a number of technological
breakthroughs. Our preoccupation, though on the surface we were concerned with
technology in many senses, was with the social implications of that technology …
technology expanded the human range, the possibility for an increased number of
choices for human beings, increased social mobility [and] increased physical mobility …

---

99 *Growth and Form* exhibition catalogue, Tate Gallery Archive. Horne will feature later in the narrative and it is unlikely that there was any mutual influence between Horne or the Birkbeck scientists and the IG at
A good deal of increase [came] through the media themselves, through photography, television, microscopes, and telescopes ... there was a feeling here of an expansion actually of what people were capable of feeling [and] what people were capable of doing.¹⁰⁰

By 1955, the IG meetings had been reconvened under the direction of John McHale and Lawrence Alloway. These later meetings show a continued emphasis on technology. For example, on March 8, 1955, McHale organized a talk by E. W. Meyer on Shannon’s information theory in which 22 people attended. Few, if any, of the members of the IG had the mathematical training to fully understand the technical side of the talk. However, the diagrams used by Meyer formed the basis of one of McHale’s most famous works, *Transistor* (see figure).

3.4 Buckminster Fuller and “Synergetic” Geometry

In one the meetings of the year 1953-54, the Independent Group considered the work of Buckminster Fuller (1895-1983). Unlike the public seminars organized by IG, their meeting to discuss Fuller was “closed.” Buckminster Fuller was a natural choice for study by the IG. First, he was on his way to becoming a prominent American architect/theorist and the IG tended to be fascinated by American post-war culture. Second, like the group, he thought that architecture should incorporate cutting edge science and technology. John McHale began to explore the work of Buckminster Fuller in more detail.
He wrote him a letter asking him the simple question “Did the Bauhaus influence your work?” to which Fuller replied that it did not with a 7000-word letter! (Hamilton 1984) Fuller’s letter was dated January 7, 1955 and McHale edited it and eventually had it published in *Architectural Design* (McHale 1961). This letter excited McHale to the degree that his “future seemed instantly mapped for him.” (Hamilton 1984) His study of Buckminster Fuller’s ideas led to the publication of one of the first books devoted to Fuller (McHale 1962).

Fuller’s distinctive style had its roots in his early life experiences. He was born into a prominent New England family. His first four years of his life were rather blurry. His defective sight was not discovered until later and some have suggested it was these early fuzzy years that focused Fuller on large-scale patterns (Rosen 1969, p. 6). Fuller began college at Harvard, but was asked to leave after he skipped his exams and went on a spending spree in New York City. He began working in a cotton mill, but after the United States joined World War I, Fuller enlisted in the Navy where he made a few minor inventions. After his service, he joined his father-in-law in a construction company, but found himself unemployed after his father-in-law lost control of the company. At this time, Fuller decided to devote his remaining life to improving the world through better design. Over the next few years he designed and built a streamlined car, the “Dymaxion” car, a mass-producible bathroom, and an aluminum house (Baldwin 1996). He also formulated a holistic and seemingly incomprehensible “synergetic geometry” and viewed his inventions as applications of his this “geometry” (Fuller 1975; Kenner 1973). It was, however, his geodesic domes that made Buckminster Fuller most famous.

It is difficult to give the necessary and sufficient conditions of being a geodesic dome, especially given the variety of enclosures designed and called geodesic domes by Fuller. One can, however, describe an exemplary case. The exemplary geodesic dome is

---

101 ICA Report for the year 1953-54, Tate Gallery Archives 955.1.1.15.
constructed from many approximately equilateral triangular faces (see figure 3.3). It approximates a tessellated icosahedron, although in most cases not all the triangles are identical. Many of the edges of the faces, when projected onto a sphere follow arcs of great circles. Fuller called his enclosures *geodesic* domes because the minor arc of a great circle is the shortest path between two points on the surface of a sphere, i.e., a geodesic.

In 1946, Fuller organized and incorporated the Fuller Research Foundation. He traveled around the country showing students how to build geodesic domes. In 1949, as a lecturer at Black Mountain College, he had students build a dome. It was there that he met the sculptor Kenneth Snelson.\(^{102}\) Fuller’s big break came in 1952, when Henry Ford II contracted him to build a ninety-three foot diameter dome at the Ford Motor Company plant in Dearborn, Michigan. After this publicly visible structure was built, Fuller’s ideas began to gain more interest. Prudently, Fuller patented his geodesic dome and could claim royalties on any dome built (Patent No. 2,682,235, June 29, 1954). Fuller also worked with the Department of Defense to build domes that would house radar equipment necessary for a missile early warning system. These domes had to withstand harsh arctic weather as well as be invisible to radio waves. Fuller designed a plastic radar dome, later shortened to “radome,” that was 55 feet in diameter, 40 feet high, and could withstand 220 mile-an-hour winds.

\(^{102}\) Snelson will be important later in the narrative.
He also built domes for the US Marines that could be airlifted and transported with a helicopter fully assembled. In 1956, the Department of Commerce asked Fuller to design a dome for the US Pavilion at the International Trade Fair to be held in Kabul, Afghanistan. The Kabul dome, as it became to be known, was assembled by largely illiterate native workers who blindly followed simple assembly rules.\textsuperscript{103}

\textsuperscript{103} It was later disassembled.
3.4.1 The Meeting Between Fuller and Klug

Buckminster Fuller was intrigued to hear about possible similarities between domes and viruses. A meeting was arranged in July 1959 between Klug, Finch, and Fuller. No written record of that conversation remains, but Fuller subsequently wrote:

A virologist is a very interesting kind of man. He was rarely trained to be a virologist; he was trained as a specialist—as a nuclear physicist, a mathematician, a chemist or a geneticist. But as virologists, they become one: an unusually comprehensive thinker—an integrator of widely held findings of science … In 1957 [sic] Klug asked if I could identify the geodesic-like protein shell of the polio virus. I was able to give him the mathematical explanation of the structuring. (Fuller 1969, p. 104).

This is perhaps a slight exaggeration on Fuller’s part. Since at the time, it was unclear what the exact details of the analogy were, the comparison with domes did not give Klug and Finch any further explanation of viral structure. That the placement of 60 spheres on the vertices of a snub-dodecahedron represented the closest packing is also a well-known mathematical result. The first meeting between Fuller and Klug did not begin a fruitful exchange.

3.5 The Continued Collaboration between Donald Caspar and Aaron Klug

After writing the phytopathology review article together, Klug and Caspar decided to continue their collaboration and expand the review article for Kenneth Smith’s *Advances in Virus Research*. Caspar was responsible for most of the material about helical viruses (exemplified by TMV) and Klug wrote much of the material on spherical viruses. Here I will only concentrate on the more theoretical discussion of spherical virus structure. Klug

Likewise it is exaggeration when Rosen writes, “Bucky’s geodesic dome formulas predicted exactly the geometric structure of every known virus.” (Rosen 1969, p. 163).
wrote this section in the beginning months of 1960. Although Caspar became sidetracked on the physical chemistry of protein interactions and considered breaking the increasingly long review into two articles, the manuscript was largely finished by late February. In a section on the symmetry and morphology of spherical viruses (section D3), Klug and Caspar, discuss the relationship between three types of subunit. First, they define the “chemical subunit” as an individual protein molecule. Second, they define the “crystallographic subunit” as the smallest asymmetric unit. Finally, they define the “morphological unit” as one of the “bumps” on the virus seen in electron micrographs. (These are sometimes called capsomeres.) One of the goals of their virus research was to see how these three types of subunit are related. First, one has to determine the number of each type of unit for a given virus. Second, one hopes to determine construction principles that underlie the formation of a viral shell.

Determination of the size and number of chemical, crystallographic, or morphological subunits in a virus does not explain how their units come together to form the shell. What we ultimately need to establish is the regular contact pattern between individual protein molecules. The regular shell is a consequence of this contact pattern; that is, the shell is assembled the protein molecules arranged in definite groups. Since we do not yet have techniques for the following the process of virus particle assembly we must deduce the construction principles from the symmetry and morphology of the finished product (Klug and Caspar 1960, p. 297).

In the next few pages, Klug and Caspar attempt to deduce the “contact patterns” by proposing “some tentative models”. The central idea is that the asymmetric unit can be composed of a number of smaller chemical units that then cluster to give rise to a different number of morphological units. The symmetry of the virus imposes constraints on how the asymmetric unit may be decomposed into chemical units. For example, to preserve

105 Klug to Caspar, 12 January 1960, Norman Archive.
icosahedral symmetry, identical types of chemical units situated around n-fold axes must be clustered into groups of n. Consider chemical groups on a vertex, that is, around a 5-fold axis. There must be five, ten, or fifteen, etc clustered around each vertex to maintain the 5-fold symmetry. Likewise, on the 2-fold axes (midpoints of the edges), there must be two or four or six etc. Since we know the number of each type of axis, we can predict allowed numbers of chemical subunits. Klug and Caspar consider a 120-unit model in more detail. On this model, the morphological subunits on the 12 vertices are divided into 5 and the subunits on the 30 edge midpoints are divided into two. There is no clustering on the 3-fold axes. Thus, the total number of chemical subunits is \((5\times12)+(2\times30) = 120\). (The astute reader will notice that this is not an allowed arrangement in the final Caspar-Klug theory.) This model yields 42 morphological units, 12 made from five chemical subunits and 30 made from two types of chemical units (See figure 3.4). Klug and Caspar also suggest that there could be more than one type of chemical subunit. For example, different types of chemical subunits may compose the groups of fives and the groups of twos.

3.6 Improved Electron Micrographs of Turnip Yellow Mosaic Virus

On the 10th of February 1960, while Klug was finishing up the manuscript for the Advances in Virus Research Volume, H. L. Nixon of Rothamstead Experimental Station wrote to Klug about some recent electron micrographs of TYMV that he and his colleague had taken. Nixon wrote to Klug because his interpretation of the pictures was “markedly at variance” with Klug’s 1957 proposal that TYMV consist of 60 subunits laying at the vertices of a snub dodecahedron. Nixon wrote of his observations:

Two aspects are visible, one with a ring of six subunits with one in the centre … This ring is in turn surrounded by eight other units, although eight are not usually visible there is room for them and I have little doubt that eight is the correct number. The other common aspect shows four subunits in a diamond shaped arrangement, with its centre
tilted slightly away from the observer. These 4 units can be seen to form parts of rings of 6 and 5. If one constructs a complete particle according to this pattern the result is a body with 532 symmetry, consisting of 32 subunits of two kinds.\textsuperscript{106} Nixon missed an inconsistency in his letter. Given the proposed 532 structure, the subunit that is surrounded by six others sits on a 3-fold rotational axis, and this implies that the ring of six should be surrounded by \textit{nine} others, not eight as he claimed. Klug replied a week later asking if Nixon meant that the proposed structure consist of two types of unit: 12 units at the corners of an icosahedron and 20 units at the corners of a dodecahedron. Klug called the first set of units “white” and the second “black” to correspond to a diagram that he included (which was, presumably, like figure 3.4 left). At first Klug was skeptical of Nixon’s result:

… this would imply that each white unit lies above the centre of a pentagonal ring of black units. I do not see any such arrangements in the picture you sent. In fact, without the cue of 532 symmetry from the X-ray, the e.m. pictures would look to an unprejudiced eye as though the particles were made up of 14 knobs at the vertices of a rhombic dodecahedron which has just the right features of lozenge-shaped rings of 1 + 6 neighbours.\textsuperscript{107}

In retrospect, one reason one does not see five knobs around one knob is that by looking down a virus’s 5-fold axes on an electron micrograph, one observes a superposition of the front and the back of the particle. Since the two sides are out of phase, the five knobs on the front do not reinforce the five on the back and one observes a more complicated projection. Of course at the time, the idea that an electron micrograph represented a superposition was not yet accepted. Instead many still thought that negative staining produced a one-sided “footprint” rather than a two-sided “superposition.” Nonetheless, Nixon remained

\textsuperscript{106} Letter Nixon to Klug, 10 February 1960, Norman Archive.  
\textsuperscript{107} Letter Klug to Nixon, 17 February 1960, Norman Archive.
convincing that it could not be a rhombic dodecahedron, as which Klug wrote:

I am pleased that your pictures cannot be explained by a rhombic dodecahedron, since
that would not have the right [i.e., 532] symmetry. I mentioned it only to make quite sure
that you could rule it out. As I pointed out in my last letter, the 20+12 unit structure
(which would, incidentally, be a rhombic triacontahedron if all rhomb edges were equal)
is quite compatible with the X-ray observations. 108

Klug also informed Nixon that a few days before he got his first letter, Hugh Huxley, of
University College, London, had shown Klug similar pictures and has independently come
up with a similar model of TYMV. In fact, Klug had given Hugh Huxley the samples of
TYMV. 109

While electron microscopy was revealing insight into the structure of TYMV, Klug
and Finch continued using x-ray crystallography to further probe the structure. In a 1960
paper modeled on Franklin and Caspar’s determination of the RNA position in TMV, Klug
and Finch compare the structure of infective TYMV particles and non-infective TYMV top-
component (Klug and Finch 1960). The hope was that the comparison of the TYMV
protein with and without RNA would yield information about the structure of RNA in
TYMV. Klug and Finch succeeded in growing single crystals large enough for x-ray
diffraction. The task of collecting intensity data for TYMV was formidable. There are
approximately 56,000 independent reflections to spacing of 5 Å and Klug and Finch resign
themselves to sample a much smaller subset of them. Virus crystals are degraded by the x-
ray exposure needed for a precession photograph and many photographs are needed to
ascertain useful information. Furthermore, the diffraction from top layer was much weaker
than for the complete virus and Klug and Finch were forced to use the powerful rotating

109 Hugh Huxley Interview.
anode x-ray tube at the Royal Institution. Nonetheless, Klug and Finch’s observations were able to show that the protein component of TYMV has 532 symmetry.

Klug and Finch addressed the electron micrographs of Huxley and Zubay, and Nixon and Gibbs. One could account for the electron microscope data and remain consistent with the x-ray data if the 32 “knobs” or “bumps” were arranged resembling a pentakis dodecahedron or a rhombic triacontahedron. These two semi-regular solids are two extremes in a continuous series: by varying the radii of the set 12 icosahedral knobs relative to the 20 dodecahedral knobs one can generate any one of the series. All members of the series have true 532 symmetry.

It might appear that TYMV violates one of the postulates of Crick and Watson’s hypothesis: on crystallographic grounds they proposed that all small spherical viruses are composed of a multiple of 12 subunits. Obviously the number 32 is not a multiple of 12. Klug and Finch suggest a way to resolve this apparent tension. If the particles have true 532 symmetry, then the knobs themselves can be further subdivided. These sub-sub-units Klug and Finch call “structural units”. For example, each of the 12 knobs that lay on five fold axes can be divided into 5 identical structure units and those 20 knobs on three fold axes can be divided into 3. Thus, in this example, the 32 knob structure consists of 120 structure units of two types \( (12 \times 5) + (20 \times 3) \). Klug and Finch also allow that each structure unit may itself be divided into “chemical units”—i.e., distinct polypeptides. This subdivision also addresses the data from the end-group analysis that suggested there were more than 60 protein molecules present in each viron.
Figure 3.4 Klug and Finch’s 1960 model of TYMV. The left represents the “bumps” or capsomeres seen by electron microscopists. It consists of 32 “bumps.” The right represents each “white” bump being divided into 5 structure unites and each “black” bump divided into 3 structure units each consisting of two chemical units, one black and one shaded.

Given the possibility of a subdivision of the structural units, the number of chemical units must be a multiple of 60. At the time, the best estimate of the number of chemical units was 150, although the amount of error was high enough that the real value could be 120 or 180. If the real value was 180, Klug and Finch favored the hypothesis that the knobs on the five fold axes consist of 5 chemical units and the knobs at the 3-fold axes contain 6 chemical subunits each. (The alternative would be 10 chemical units on the 5-fold axes and 3 chemical units on the three fold axes.) Klug and Finch’s model suggested that there were two different types of structure unit (those in fives, and those in threes or sixes). They acknowledge that even if the structural units lie in different positions, it is possible that they could be chemically identical.
It should be noted that even if all the chemical units are identical chemically, they are still differentially situated in the structure (i.e. they are not structurally identical) (Klug and Finch 1960, p. 208).

Caspar and Klug’s later work would be concerned with exactly how the units are differently situated in the structure.

After giving this structural model, Klug and Finch consider a question that immediately arises: what are the building units of the top component? For example, do the chemical units first form rings of 5 or 6, which then assemble or “crystallize” into the protein shell? Alternatively do the units form trimers and then assemble? In either case, they suggest that an empty shell might be “built by itself” without the need for a “further organizing principle or agency” such as a core of RNA. This set of intriguing questions signaled a shift from a structural approach to a dynamical approach.

3.6.1 The Invention of Negative Staining

Before continuing with the Caspar and Klug’s further collaboration, let us now look at the principal innovation that occurred in electron microscopy that allowed electron microscopists to take pictures with enough resolution to see viral substructure.\footnote{Rasmussen (1997) details how electron microscopy changed American biology over the period 1940-1960. Strangely, he omits mention of the invention of negative staining.} This innovation known as “negative staining” was discovered more than once.

In October of 1954, the MIT microscopist Cecil Hall submitted a paper describing how to determine the mass density of particles in electron microscope specimens (Hall 1954). Hall used BSV and TMV from Robley Williams with the goal of determining the optimal conditions under which viruses are take up stain (phosphotungstic acid) and thereby mass. A higher mass particle will produce more contrast in the micrograph. Almost as an
after-thought, Hall mentioned that if you did not wash the stain off the specimen and left it
surrounding the virus you could see lightly stained particles in a dark reagent:

Although the effect ... is the opposite to what is usually sought by the use of electron
stains, the visibility of particles of low scattering power can be enhanced as well, if not
better, by surrounding them with dense material rather than impregnating them with
dense material (Hall 1954, p. 10).

Hall’s discovery did not seem to be widely appreciated. Three years later Hugh Huxley
rediscovered the effect.

Hugh Huxley worked at University College, London. Although his principal
interest was in the structure of muscle, in 1956 Jim Watson asked him if one could stain and
thereby see the RNA of TMV in micrographs with the newly installed Siemens Elmiskop I
electron microscope. At first the results were disappointing, but Huxley noticed a
“curious effect” when the excess stain was not washed off (Huxley 1956). Huxley noticed
that he could achieve the effect either with either 40% phosphotungstic acid (PTA) or in a
dilute solution of potassium chloride (KCl). Using this “outlining” technique, Huxley
showed that TMV appeared to have a hollow core of 20-30Å. Huxley ended his two-page
paper with the following observation:

The outlining technique would appear to be a quite useful one for this type of specimen,
particularly as it is so simple and gives excellent contrast and resolution (Huxley 1956,
p.261).

Nonetheless, although Huxley rediscovered the technique he admits that he “obviously
underestimated its use.”

In late 1956 or early 1957, in Cambridge, Sydney Brenner, with an eye to
solution of the genetic code, began to study the tail fibers of bacteriophage (Judson 1979, p.

---

111 Hugh Huxley Interview.
112 Hugh Huxley Interview.
He wanted to use the Siemens Elmiskop electron microscope in the Cavendish Laboratory, the first production model of the microscope which was installed in 1954 by direction of Lawrence Bragg and V. E. Cosslett. Brenner worked with Robert Horne, an electron microscopist, to develop a stain that was more alkaline than the stains currently in use whose low pH disrupted the phage particles and proteins. Some years earlier, Brenner had used Indian Ink to alkalize bacteria for use in light microscopy. Now they neutralized many different heavy metal salts, discovering that potassium phosphotungstate (KPT) worked the best. They refined and perfected the technique. On the day the Russians launched Sputnik 1 (October 4, 1957), Brenner and Horne took their first micrograph of “negatively stained” T2 bacteriophage. Brenner and Horne also looked at the plant viruses TMV and TYMV. They submitted a paper to Virology, but the paper was rejected because the technique was “too new” to allow critical assessment (Horne and Wildy 1979, p. 107). As Brenner put it at the time, “it was probably the first paper ever to be rejected for being too novel.” Instead they published in Biochimica et Biophysica Acta (Brenner and Horne 1959). Publication of Brenner’s work on bacteriophage was held back pending the production of the new Journal of Molecular Biology.

In December 1958, Peter Wildy, a medically trained pathologist, approached Robert Horne to see if one could use the EM to quantify the number of herpes viruses he had purified. Wildy had spent 1953-54 with Macfarlane Burnet at the Walter and Eliza Hall institute in Melbourne, Australia. There he had become interested in herpes simplex virus and the mode of its growth (Wildy 1954). He returned to the problem of herpes multiplication in 1958-9, while an MRC external worker in Michael Stoker’s lab in

114 Sydney Brenner Interview.
115 Robert Horne Interview.
116 Robert Horne Interview.
Cambridge. On January 10, 1959, Wildy wrote in his otherwise bland lab book “e.m. photographs of great interest”.
Figure 3.5 Herpes simplex virus. Micrograph taken by Horne and Wildy.
Wildy put it this way in his biographical Curriculum Vitae\textsuperscript{117}:

Initial observation with herpes [see figure 3.5] were [sic] so promising that we immediately resolved to examine a variety of animal viruses simultaneously and brought in all the virologists we could find in Cambridge. Between January and June 1959 we [Horne and Wildy] examined particles from all the groups of vertebrate viruses then known. We were able to discover a unity in the structural pattern which went beyond the conventional boundaries limiting virologists at the time.

Indeed by 1961, sugar beet virus, poliovirus, herpes simplex, myxoviruses, polyoma viruses, bacteriophage φX174, and parainfluenza virus had been seen through the Siemens Elmiskop I microscope. Many articles of the newly formed \textit{Journal of Molecular Biology} concerned the electron microscopy of viruses. Perhaps the most impressive were the micrographs of adenovirus. Brenner and Horne went to a sports store on Kings Parade and bought 252 table tennis balls (See figure 3.6). They built the model on the bench in “the hut.”

![Figure 3.6 Adenovirus. Left micrograph taken by Brenner and Horne. Right 252 table-tennis-ball model built by Brenner and Horne.](image)
3.7 The Birth of Quasi-Equivalence

On November 11, 1960, Caspar wrote to Klug with some exciting news. Given the electron microscopy and biochemical data, it was becoming clear that nature violated the Crick-Watson constraint that there was a maximum of 60 identical subunits per virus. Caspar had discovered the beginnings of a solution.

I have found a general solution to the structural problem: How can a particle with icosahedral symmetry be built out of $60n$ identical structure units? The solution is a selection rule: $n = 3^p q^2$ where $p = 0$ or $1$ and $q$ = any integer.\(^{118}\)

The key insight was to see that the angles or “contact points” between the identical structure units are not “absolutely equivalent” for values of $n > 1$, but vary around a mean value. “The unique idea here is non-crystallographic equivalence.” Caspar suggested that his idea was analogous to Pauling’s use of a non-integral helix to solve the structure of the alpha helix. Klug wrote in the margins of the letter that the true analogy is with Crick’s coiled coil.\(^{119}\) The selection rule did not seem to be violated by the then known viral structures. Caspar then asked Klug to join him in a collaborative project. He included a rough draft of his notes and suggested that they rewrite his ideas. Three days later Caspar wrote to Klug again suggesting a number of possible models for BSV structure. He suggests that it could be $n = 9$ a structure (with $9 \times 60 = 540$ subunits) where there are 270 “chemical” subunits consists of two very similar halves. Alternatively, it could be $n = 4$ with 240 ($4 \times 60$) subunits). This theoretical breakthrough reoriented Caspar’s research agenda. “Now that there is a unifying principle to work with I am more interested in the spherical viruses and will be glad to collaborate with you in anyway possible on BSV.”\(^{120}\) Caspar also hoped to work with Jim Watson on polyoma virus. A collaboration between

---

\(^{117}\) CV in the possession of the author.
\(^{118}\) Letter Caspar to Klug, 11 November 1960, Norman Archive.
\(^{119}\) See also the postcard Klug to Caspar, 17 November 1960.
\(^{120}\) Letter Caspar to Klug, 14 November 1960, Norman Archive.
Klug and Caspar seemed natural since, as Caspar noted, his own breakthrough fitted well with Klug’s earlier ideas on the subject.

Klug replied on the 17th that he would very much like to collaborate with Caspar and that in fact he had recently been trying to classify all the structures that can be made up of linear combinations of 30, 20, 12, and 60 “on the basis of both density and uniformity of packing.” Klug explained that the idea was to triangulate the sphere to get triangles as nearly equilateral as possible and in passing mentioned Buckminster Fuller. Klug was hesitant to include another of his ideas in the manuscript, what he called the “statistical model”—the idea was that there may be subunits that can sit on either 5-fold or 6-fold axes—since he thought that data might soon be available to decide the issue. As for polyoma virus, Klug’s interpretation of the electron micrographs allowed for the possibility that there might be all rings of 6 not 12 rings of 5 and the rest rings of 6.

Before Klug’s letter reached him in Boston, Caspar had found one of the first books written on Buckminster Fuller, The Dymaxion World of Buckminster Fuller (1960), by Robert Marks. Caspar found the book “very enlightening” and wrote to Klug, “In retrospect, I expect you will find it surprising that you did not recognize how well Fuller’s geodesic structures do, in fact, represent virus structure.” Caspar thought Fuller had already deduced the selection rule for icosahedral structures (Marks 1960, p. 46). Fuller devises his rule by looking at the close packing of spheres over the tetrahedron, cube, octahedron and what Fuller calls the vector equilibrium, but is otherwise known as the cuboctahedron. The subsequent formula applies to icosahedral shells (Fuller 1969, p. 105).

\[ P = 10F^2 + 2 \]

121 Letter Klug to Caspar, 17 November 1960, Norman Archive, underline in original.
122 These remarks of Klug’s foreshadow the fate of polyoma virus, which was shown to consist entirely of rings of 5 and thus was the first clear violation of the Caspar-Klug rules of 1962.
123 Letter Caspar to Klug, 18 November 1960, Norman Archive.
F stands for the “frequency” of the figure, or the number of spheres that can be packed along each edge. P stands for the number of spheres on a one-layer shell. To see the connection with Caspar’s derivation, notice that what Fuller calls a sphere is divided either into 5 or 6 subunits for Caspar. The 12 spheres on the five fold axes are divided into 5 and the remainder into 6. Thus,

\[ n = 6(10F^2 + 2 - 12) + (5 \times 12) = 60F^2 \]

Thus Fuller’s symbol “F“ represents Caspar’s symbol “q” and we notice that Fuller’s formula describes one of Caspar’s two classes (the P = 0 class). Caspar noted that Fuller is concerned with building shells out of triangles that need not compound into integral numbers of hexagons and pentagons. Another shift occurred in interpreting Fuller’s geodesic domes: rather than identify vertices as viral subunits, Caspar identified the triangular faces as subunits. Technically this shift involves moving from one figure to its dual.

Klug response to Caspar’s exuberance was a little more reserved. He wondered about what the justification could be for why virus subunits cluster into fives and sixes, if indeed they do.

My difficulty is still the same one that led me to identify the structure units with the vertices of the triangulations rather than with the centres as you have done. In the first type (my early ideas) the structure units are packed as densely as possible on the surface of the sphere, i.e. make as many contacts as possible and share out the stresses at any one point in the framework in the most favorable way (Buckminster Fuller’s principle). In the second type (i.e. your idea) each structure unit has only three neighbours so that if one joins up the centres of the units by lines one has a polyhedron with trihedral vertices.\(^{124}\)

\(^{124}\) Letter Klug to Caspar, 2 December 1960, Norman Archive.
Klug was aware that polyhedra with trihedral vertices span the maximum volume with the minimum material and therefore that there was some reason from efficiency for preferring this second class (rather than its dual). A further problem was whether one should allow sub-clustering (as Caspar had done for his suggestion for BSV above) and whether removing this constraint would lead to many more difficulties. “There are so many aspects to this problem and so many loose points that I have difficulty in deciding in what detail to deal with them.”

The groups of 5 or 6 postulated by Caspar were intended to capture the results from electron microscopy that showed “bumps” on the viral surface. In June of 1960, Peter Wildy, William Russell, and Robert Horne had submitted a paper on the electron microscopy of herpes virus (Wildy, Russell, and Horne 1960). In terms of showing virus substructure, these pictures were the most spectacular yet. After looking at the micrographs in Wildy, Russell and Horne’s paper (especially Wildy’s figures 4 and 5c), Klug wrote to Caspar wondering whether the “clusters” have an independent existence and thus Caspar’s structure units would bear little relation to reality. For example, Klug wondered if what Wildy calls capsomeres are in fact helices (See figure 3.7). The alternative that Caspar had been developing was that the capsomeres were merely morphological units made up of the more fundamental structural units.

---

125 Letter Klug to Caspar, 2 December 1960, Norman Archive.
Figure 3.7 “Empty” Herpes virus: figure 5c from Wildy, Russell, and Horne (1960). The arrow points to a cluster or a capsomere. Compare with a complete herpes shell in Figure 3.5.
By January 1961 Klug too had received Marks’ book on Buckminster Fuller.\textsuperscript{126} After looking at many pictures of domes, Klug gleaned a couple of conceptual advances. First Klug learned that some domes had been assembled by unskilled workers following simple building rules.\textsuperscript{127} For example, the United States Department of Commerce had used a dome as a pavilion at the Kabul International Trade Fair in 1956.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3.8}
\caption{Dome assembled in Kabul Afghanistan in 1956.}
\end{figure}

The components of the 100-foot diameter dome had been flown in by DC4 and assembled within 48 hours by workers following a system of color-coded building rules. Presumably viruses also might assemble by “following” simple building rules. (At the time, Caspar and Klug had been experimenting with the following bonding rule: If you label each triangular subunit’s three edges, a, b, and c, then c bonds only with c, a bonds b, and b bonds a.) The

\textsuperscript{126} Letter Klug to Caspar, 11 January 1961, Norman Archive.
color-coded assembly system was analogous to specific bond types between protein subunits. Crick had even enunciated a related principle—"virus assembly should be simple enough that a child could do it." Second, Klug spent some time looking at the 270-subunit tensegrity sphere on page 196-7 of Marks’ book (See figure 3.9). He wrote to Caspar: “Also pictures R8-11 give a very good illustration of building a structure with 60n units, each of which has some ‘give’ in it. This, of course, he achieves by using wheels and runners, whereas the protein units would require a like degree of flexibility. Anyhow it is a striking illustration of how very nearly equivalent the units are.” Klug’s idea was that the slight differences in each unit’s relations to its neighbors are compensated in the overall structure by differentially winding the wires around “wheels,” as one does when tightening the ropes on a yacht’s sail. As it turned out, Klug learned later that what he had thought were *rotatable wheels* in Fuller’s dome were, in fact, *static washers* and Fuller had calculated all the differences by hand and carefully built them into the structure in advance. In April, John McHale sent Klug, who was home in South Africa, Buckminster’s manuscript “Energetic Synergetic Geometry”. In this manuscript Fuller revealed how he made the “270 Strut Isotropic Tensegrity” sphere. There were 5 types of precalculated strut or boom, not one type with “give” in it as Klug had inferred. Klug calls this event a “misprision” after Caspar’s friend the famous literary critic Harold Bloom who discusses the term in his 1973 book *The Anxiety of Influence* and his 1975 book *A Map of Misreading* (Laszlo 1986, p. 50). Bloom contends that later writers are influenced by earlier writers by misconstruing and reinterpreting the earlier texts.

\[127\] Letter Klug to Caspar, 21 February 1961, Norman Archive.  
\[128\] According to Caspar, Crick formulated this principle in 1955.  
\[129\] Letter Klug to Caspar, 21 February 1961, Norman Archive. Note that 270 is not divisible by 60. If each of the 270 symmetric “subunits” were divided into 2, one would have a 540-subunit structure.  
\[130\] Buckminster Fuller “An Introduction to Energetic Synergetic Geometry” MS, p. 212-224. Robert Horne was given a similar MS by Fuller also. Some of this material was published in his *Synergetics* (1975). See for example p. 394.
Figure 3.9 The 270 Subunit Tensegrity Sphere.
At the time Klug was looking at Marks’ book, more electron micrographs of viruses were becoming available. After looking at electron micrographs of turnip crinkle virus taken by Bob Haselkorn, Klug wondered if it might consist of two shells: a 42 “knob” shell nested within a 92-knob shell. Klug thought that combined with the biochemical evidence that the 134 knobs each have about 3 histidines each that this supported his “statistical model”. On this model the building unit is made of 3 or 6 subunits. Each building unit has at least three fold symmetry and can adopt two orientations: if a sphere is divided into 10(n-1) hexagonal cells and 12 pentagonal cells (think of a soccer ball), then the statistical unit could fill either type of cell. The orientations of the unit are in some sense guided by a “core.” Caspar was skeptical of the double shell and raised the possibility that the internal “shell” was an artifact resulting from a re-aggregation process. Of the statistical model, Caspar wrote that he agreed about the possibility of such a structure, but that it would not necessarily have more of a tendency to form an icosahedral-shaped shell than one made of 60n subunits as Caspar had suggested.

By February, Klug was aware of two other groups who were also working on a theory of spherical virus structure. The Italian physicist Mario Ageno (1915-1992) had sent a manuscript to Francis Crick who passed it to Klug. In his manuscript, Ageno derives the formula $10(n-1)^2 + 2$, which is closely related to Buckminster Fuller’s formula mentioned above. His paper, however, shows his biological naiveté: he argues that the icosahedral shape is to be preferred over other polyhedra because of “surface tension”.

132 Letter Klug to Caspar, 3 February 1961, Norman Archive.
133 Letter Caspar to Klug, 7 February 1961, Norman Archive. This letter is also significant in that Caspar suggests, in passing, that carbon might also have an icosahedral morphology as was later discovered in the fullerenes.
135 Ageno, M. “Some Remark [sic] in the Shape of Viruses” MS, Norman Archive.
The second group now working on the structure of viruses was closer to home. George Hirst, editor of the journal *Virology*, had commissioned Robert Horne and Peter Wildy to write a paper on the structure and symmetry of viruses (Horne and Wildy 1961). Robert Horne worked in Cambridge and Wildy at the newly formed Virology Institute at the University of Glasgow. Klug learned that they too had derived the formulae $10q^2 + 2$ and $30q^2 + 2$ as well as a skew type $10(m^2 + m) + 12$ and a further type based on the icosidodecahedron. Klug thought their paper was mainly descriptive and did not compete directly with their more theoretical project, but nonetheless the existence of two groups working on similar problems provided more urgency for Caspar and Klug to finish up and submit a manuscript.  

3.7.1 The Selection Rule: $h^2 + hk + k^2$

Klug discovered that the selection rule derived by Caspar was incomplete. The formulae ruled out certain possible solutions. For example, it did not consider enantiomorphic structures. Spurred on by this demonstration, Caspar worked on the hexagonal lattice and was able to show that the complete selection rule was given by the simple but elegant formula:

$$N = h^2 + hk + k^2 = Pn^2$$

Where $h$ and $k$ can be any integer and $P$ refers to a particular class of polyhedra.  

This simple formula described the parameters of any possible spherical virus structure. Klug called it a geometrical classification of icosahedral structures in contrast to a physical problem that remained: why are so few of the possible types realized in nature?  

---

137 Letter Caspar to Klug, 13 March 1961, Norman Archive.
3.7.2 Two Projects in Theoretical Structural Virology: Geometrical and Physical

Over the summer of 1961 Klug went home to South Africa. Caspar refereed the paper by Horne and Wildy on symmetry in virus structure from the point of view of electron microscopy. In August he wrote to Klug to discuss how they should continue. Two distinct research projects persisted. First, a geometrical project: to enumerate and classify icosahedral structures. Caspar’s derivation of a general formula had made significant progress on this project. Second, a physical project: to explain why are so few of the possible structures are realized in nature. Caspar suggested they write two papers based on this division.\textsuperscript{139} The first paper would be applicable to boron molecules,\textsuperscript{140} viruses, and geodesic domes. The second would be focused more specifically on virology.

Caspar came to London to work with Klug for three weeks in mid August. On his return to Boston he was visited by Peter Wildy who shared the news that Tom Anderson had managed to take electron micrographs of single isolated “morphological” units: they appeared to consist of groups or rings of 6 smaller “structure” units. No rings of 5 were seen as were suggested in Caspar’s model. On the other hand, Klug’s statistical model, which allows rings of six to sit statistically on the (statistically) five fold axes, did not require rings of five. For the remainder of 1961 Caspar and Klug made few conceptual advances. In January 1962, Klug put the situation this way:

I am afraid I have got into a complete malaise with the icosahedral virus paper. I seem unable to take it up again although on all sides of us people are busy dealing with points that have arisen. For instance have you seen Pawley in the January Acta Crystallographica and Ageno’s paper in … Nouvo Cimento. Also I see that Horne and Wildy mention the connection with Buckminster Fuller which I am sure they learned

\textsuperscript{139} Letter Caspar to Klug, 8 August 1961, Norman Archive.
\textsuperscript{140} Caspar had been speaking with Bill Lipscomb who would win the Nobel Prize for his work on Boron.
only from us.141 (If you will remember, I spoke about it at the Biophysics Conference in July 1959 and you presumably showed Marks’s book to Wildy). Therefore, despite the doubts I have indicated previously, I think it would be a good idea to publish our thoughts as soon as possible as we had intended last summer.142

One thing that Caspar and his new post-doc Ken Holmes had been concerned with since the summer was an analysis of the Dahlemense strain of TMV. This strain gave interesting results in the diffraction pattern. Additional meridianal and near meridianal diffraction maxima appeared on the layer lines halfway between those given by the common strain (Caspar and Holmes 1969).143 Don discussed these data with his laboratory partner Carolyn Cohen.144 She saw that the results could be described by a periodic perturbation in the TMV helix and pointed Don to the recent MIT PhD dissertation of Carrol Johnson.145

In his dissertation, Johnson considered the diffraction effects of various types of periodic distortions of a helix. Johnson noticed that distortion in the axial direction led to changes in the meridian of the diffraction pattern. The relevant finding for Caspar and Klug’s theoretical aspirations was that due to the periodic perturbations of Dahlemense TMV helix, the rod shaped virus essentially consisted of identical subunits lying in non-identical environments. It provided a real example that Caspar and Klug’s speculations about spherical viruses might be on track.

3.7.3 The 1962 Cold Spring Harbor Meeting on Animal Virus Biology

Later that month, Klug was invited to present at a June meeting at Cold Spring Harbor on “Basic Mechanisms in Animal Viruses.” Robley Williams was to chair a

141 Robert Horne had shown Wildy his copy of Buckminster’s unpublished notes, which Fuller had given to Horne sometime between 1959 and 1961. Horne Interview.
142 Klug to Caspar, 9 January 1962, Norman Archive, italics mine.
143 Note that although this paper was published in 1969 it was first received at the JMB on 25 August 1965. However, the first consideration of this strain began much earlier.
144 For more on the Cohen- Caspar laboratory see Cohen (1998).
session on the “Structure and Intracellular Location of Viruses,” which along with Klug, scheduled Peter Wildy to talk on electron microscopy. Klug saw this as an “appropriate place to present some of our ideas on icosahedral structure.” He wrote to the organizers requesting that Caspar be included as a co-author. Casper suggested that Klug come to Boston for a least a month so that they had time to write the Cold Spring Harbor paper. Meanwhile Klug continued to think about spherical viruses. He spoke with Crick to “try out” some ideas with him:

I noticed that even he did not see at first the implications of near equivalence and I had to make a diagram [see figure 3.10] to convince him that by following a simple bonding rule, one could make the structure.

The central idea is that with two types of bonds, those that bind the subunits into fives or sixes and those that bind the groups of fives and sixes together, one can assemble icosahedral structure with more than 60 subunits.

---

145 Cohen Interview.
147 Letter Caspar to Klug, 29 January 1962, Norman Archive.
148 Letter Klug to Caspar, 6 February 1962, Norman Archive.
Figure 3.10 Bonding Diagram, drawn in Klug’s hand, in February 1962 indicating how simple bonding rules and near equivalence can lead to icosahedral structures. Only one face of the icosahedron is shown in each case.
There are many different sizes of structure that can be assembled using these rules. How do the subunits “know” which particular size to assemble into? In other words, what determines the correct ratio of the number of clusters of six and clusters of five for each species of virus? Caspar and Klug’s insight was to attribute a property to the subunit that they called “built in curvature.” This property guaranteed that there was only one unique size of closed structure that could be built from subunits with a particular degree of built in curvature. Crick remained skeptical:

The other point that worried him [Crick] was whether the built-in curvature of the subunits made the strict icosahedral structure come out. If you look through our correspondence at the time when you were building the 540 unit (92 cluster) shell I queried [Feb 3, 1961] whether the units could not have assembled themselves in a less symmetrical form. You replied no [Feb 7 1961], but it is clear that one can build isolated fragments of shell or even pieces of plane sheet which will not close up to make the complete structure. Francis [Crick] wondered whether there was not some actual rule of assembly which forced five-rings to form in the appropriate places. 149

The appropriate places for the five-rings would be on the fold-fold vertices of an icosahedron. Klug also spoke with Buckminster Fuller in two “very instructive” sessions. Unfortunately, the meeting served mainly to prompt Klug to further develop his statistical model, a model not essential to the final theory. The idea was that with a single type of subunit based on a rhombic triancontahedron that can sit on an edge, on a face, or a vertex, one can build nested icosahedral shells.

Over the month Klug saw better electron micrographs of viruses such as BSV and canine hepatitis virus. Klug wondered if the idea of nesting shells would explain the structure of canine hepatitis virus. Perhaps there was an inner protein shell around which
was wound the nucleic acid, which in turn was surrounded by an outer protein shell. The overall effect would be a polyhedral sandwich with protein “bread” surrounding the nucleic acid. Caspar suggested that perhaps the inner shell “assembles automatically” and serves as the “scaffolding” for the outer shell.150

By March, Caspar had also spent many hours speaking with Buckminster Fuller. Harvard had awarded Fuller the 1962 Charles Eliot Visiting Professorship in Poetry and Caspar arranged for them to meet at the Children’s Cancer Research Foundation in the Harvard Medical School where Caspar worked.151 They talked about tensegrity, the idea that lies behind the 270-strut dome considered by Caspar and Klug earlier. (See figure 3.9.) Although there is some controversy over the origins of the concept, the American sculptor Kenneth Snelson constructed the first tensegrity structure/sculpture in 1948 while he was a student of Buckminster Fuller's at Black Mountain College, North Carolina. There are a number of his sculptures dotted around the United States. For example, his “Needle Tower” is exhibited at the Hirshhorn Sculpture Garden on the Mall in Washington DC (Snelson, Schultz, and Fox 1981). The basic idea of a tensegrity structure is to isolate the components of the structure that are under tension (wires) and those under compression (struts). Additionally the tension components for continuous loops—they have “tensional integrity” hence the name. Tensegrity structures are in equilibrium and will return to the original state after deformation. Fuller popularized Snelson's idea by incorporating it into his “energetic geometry”, and often gave people the impression that it was one of his own ideas. Caspar wrote to Klug: “It is clear to me that our model for the virus shells are tensegrity structures.”152 Klug had come to a similar conclusion after looking at the 270-strut model. What attracted Caspar to tensegrity structures was that like viruses, “the

149 Letter Klug to Caspar, 6 February 1962, Norman Archive.
150 Letter Caspar to Klug, 12 March 1962, Norman Archive.
151 Caspar Interview.
152 Letter Caspar to Klug, 12 March 1962, Norman Archive.
structure units naturally arrange themselves in as near equivalent environments as possible.” Klug acknowledged the applicability of the analogy between viruses and tensegrity structures, but cautioned the limits of its fruitfulness:

…I have always thought of the application of tensegrity to virus structure as more a geometrical principle rather than of an engineering principle, putting emphasis on the fact that identical units can be used in almost identical surroundings with only very simple building rules for the assembly. I think it would be unwise to carry the analogy further since the protein units could certainly not be represented by rods roughly parallel to the near-spherical surface.

Klug had discussed this connection between tensegrity and near equivalence but Fuller either did not understand or did not see the importance of the connection.

Through March, Caspar continued to have conversations with Buckminster Fuller. He built new more complex tensegrity models and presented them at a seminar, at which Francis Crick “raised no objections” to Caspar’s ideas. The building of new models delayed Caspar from meeting his self-imposed deadline for a rough draft of the manuscript. Klug argued that they should not overemphasize the principle of tensegrity in their paper since many tensegrity structures were not built from 60 n units and many were not even approximately spherical (such as Snelson’s needle tower). Indeed “tensegrity” merely denoted a separation of tension and compression components in a structure. On the face of it, the separation of tension and compression had little to do with protein structures or a structure that can “build itself” as a virus does. Caspar replied that the significant analogy between virus structure and tensegrity structure was they both exist as minimum structures. If one deforms a tensegrity structure and then removes the source of

---

153 Letter Caspar to Klug, 12 March 1962, Norman Archive.
154 Letter Klug to Caspar, 16 March 1962, Norman Archive.
155 Letter Klug to Caspar, 16 March 1962, Norman Archive; letter Klug to Caspar, 7 April 1962, Norman Archive; Klug Interview.
deformation, the structure will spring back to its original shape.\textsuperscript{157} Furthermore, tensegrity could be applied to other model-building problems in molecular biology such as Crick’s coiled coil.

Klug’s ideas were also evolving. He had re-examined some electron micrographs of ECBO virus. What was surprising was that while one could see “bumps” on the periphery of each virus, bumps in the centers (i.e., the faces of the virus closest and furthest away from the viewer) of the virus were not seen. What this observation suggested is that the clustering of structural units into fives and sixes as suggested by models that Caspar and Klug built might not represent all spherical viruses.

My theory is that, in fact, the 180 protein sub-units are organized – in a kind of tensegrity structure if you will – giving a mesh like appearance, but on the edges of the particle various edges and vertices of the mesh will superimpose, giving a bump-like appearance.\textsuperscript{158}

Klug wrote in haste to Caspar so that this caveat might be included in the close to completed manuscript. Caspar replied that he agreed that the clustering into groups of five and six was not a physical necessity.

On April 9, after an ordeal of writing, Caspar sent Klug the majority of the manuscript. Using a result of Pawley, who showed only two types of plane nets – cubic and hexagonal—can be folded onto the surface of a convex polyhedron while maintaining the same nearest neighbor contact pattern, Caspar “deduced” the possible quasi-equivalent shells by cutting and folding a hexagonal net (Pawley 1962). Caspar still needed to write the sections on deltahedra, tensegrity, clustering patterns, and the thermodynamics of shell

\textsuperscript{156} Letter Klug to Caspar, 7 April 1962, Norman Archive.
\textsuperscript{157} Letter Caspar to Klug, 11 April 1962, Norman Archive.
\textsuperscript{158} Letter Klug to Caspar, 6 April 1962, Norman Archive.
assembly. Over the next few days, Caspar was able to show that under plausible assumptions the energy of a closed shell would be lower than of a flat sheet.

Klug’s response to the first half of the draft was generally favorable. He streamlined some of the argument. Klug reiterated that they should not use the word “tensegrity” because it had a number of associations and could be interpreted either broadly or narrowly. By the beginning of May, Caspar had written the remaining sections. Klug left for Boston on May 20 ready for two more weeks of intensive writing and editing of the Cold Spring Harbor Symposium paper.

The final 24-page paper was the first paper in the published proceedings, the 27th volume of the Cold Spring Harbor Symposia. Only two papers are of comparable length: Peter Wildy and Douglas Watson’s paper on the Electron Microscopy of Viruses and Richard Franklin and David Baltimore’s paper on patterns of macromolecule synthesis in virus infected cells. Caspar and Klug’s paper consists of seven sections (See table 3.1).

---

159 Letter Caspar to Klug, 9 April 1962, Norman Archive.
160 Letter Caspar to Klug, 11 April 1962, , Norman Archive.
161 This important meeting is described in Baltimore’s biography (Crotty, 2001).
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>THE FUNCTIONAL ORGANIZATION OF VIRUS PARTICLES</td>
<td>4.5 pages</td>
</tr>
<tr>
<td>1.1</td>
<td>Simple or Minimal Viruses</td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td>Grades of Structural Organization</td>
<td></td>
</tr>
<tr>
<td>1.3</td>
<td>Sub-assembly and Self Assembly</td>
<td></td>
</tr>
<tr>
<td>1.4</td>
<td>Structural Studies</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>HELICAL VIRUSES</td>
<td>3 pages</td>
</tr>
<tr>
<td>2.0</td>
<td>Tobacco Mosaic Virus</td>
<td></td>
</tr>
<tr>
<td>2.1</td>
<td>Helical Symmetry</td>
<td></td>
</tr>
<tr>
<td>2.2</td>
<td>Physical Considerations</td>
<td></td>
</tr>
<tr>
<td>2.3</td>
<td>Flexible Rods</td>
<td></td>
</tr>
<tr>
<td>3.0</td>
<td>ICOSAHEDRAL VIRUSES</td>
<td>2.5 pages</td>
</tr>
<tr>
<td>3.1</td>
<td>Cubic Symmetry</td>
<td></td>
</tr>
<tr>
<td>3.2</td>
<td>Icosahedral Symmetry</td>
<td></td>
</tr>
<tr>
<td>3.3</td>
<td>Experimental Background</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>THE GEOMETRY OF ICOSAHEDRAL VIRUSES</td>
<td>7 pages</td>
</tr>
<tr>
<td>4.0</td>
<td>The Problem</td>
<td></td>
</tr>
<tr>
<td>4.1</td>
<td>Quasi-equivalence</td>
<td></td>
</tr>
<tr>
<td>4.2</td>
<td>Folding of Plane Nets</td>
<td></td>
</tr>
<tr>
<td>4.3</td>
<td>Deltahedra</td>
<td></td>
</tr>
<tr>
<td>4.4</td>
<td>Morphological Units and Icosahedral Classes</td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>COMMENTS ON INTERPRETATIONS OF SOME ELECTRON MICROSCOPE OBSERVATIONS</td>
<td>1 page</td>
</tr>
<tr>
<td>6.0</td>
<td>CONSTRUCTION OF SHELLS</td>
<td>2.5 pages</td>
</tr>
<tr>
<td>6.1</td>
<td>The Mechanics of Shell Formation</td>
<td></td>
</tr>
<tr>
<td>6.2</td>
<td>The Question of the Independent Existence of Morphological Units</td>
<td></td>
</tr>
<tr>
<td>6.3</td>
<td>Multi-shell and Statistical Arrangements</td>
<td></td>
</tr>
<tr>
<td>6.4</td>
<td>Shells and Membranes</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>SELF-ASSEMBLY OF ICOSAHEDRAL VIRUSES?</td>
<td>2.5 pages</td>
</tr>
<tr>
<td>7.0</td>
<td>Empty Shells</td>
<td></td>
</tr>
<tr>
<td>7.1</td>
<td>Role of Nucleic Acid in Icosahedral Virus Structure</td>
<td></td>
</tr>
<tr>
<td>7.2</td>
<td>Comparison of Icosahedral and Helical Frameworks</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.1 The structure of Caspar and Klug’s 1962 paper.
Caspar and Klug begin their paper with two facts: that nucleic acid is the “essential infective agent” and the nucleic acid is contained in a “protective package.” The “main thesis” of the paper is that there are only a limited number of efficient designs for the “coat or framework” if it is made out of a large number of identical protein molecules.

Borrowing an idea from J. D. Bernal, they suggest that biological structures do not form a continuous series in size. Rather there is a discontinuous biological order: each structure consists of units of a similar size that come together to form a structure at a higher level (Bernal 1959). Dick Crane’s ideas about efficient design give further support for this hierarchical organization (Crane 1950). If a mistake is made at a lower level it need not be incorporated into the higher level, much like a factory production line.

The dis-analogy with the production line is that there is no oversight or external direction of the viral assembly process. Viruses “self-assemble,” as had been shown for TMV by Fraenkel-Conrat and Williams in Berkeley. The self-assembly can be explained (in principle) through statistical mechanics: “Self-assembly is a process akin to crystallization and is governed by the laws of statistical mechanics.”(Caspar and Klug 1962, p. 3) That the laws of statistical mechanics do not prescribe that a set of identical subunits will self-assemble into a structure in which each subunit is situated in an identical environment is the deepest insight of the paper.

The important point is the lowest energy structure will have the maximum number of most stable bonds formed—and this may be physically realized, as in icosahedral virus shells, by quasi-equivalent bonding of identical units. These physical considerations have lead to an extension of the traditional concepts of symmetry more specifically applicable to highly organized biological structure (Caspar and Klug 1962, p.3).

Although Caspar and Klug summarize their insight in the first section, they do not explain quasi-equivalence further until the fourth section. In the second and third sections, Caspar and Klug explain the motivation for the problem that quasi-equivalence solves and how their
insights are an extension and modification of the earlier insights of Crick and Watson and of their knowledge of a rod shaped virus, TMV.

In the second section, Caspar and Klug discuss the helical virus TMV, the most well studied virus at the time. Two things are relevant about TMV for Caspar and Klug’s project. First, it self-assembles, thus providing evidence that viruses can self-assemble. Second, it is made from identical protein subunits that typically occupy identical environments (if we ignore the finite length of the rod). Furthermore, some atypical strains of TMV, such as Dahlemense strain, seem to have a periodic perturbation of the subunits that leads to the subunits laying in quasi-equivalent environments. In fact, the idea of quasi-equivalence suggests an explanation of flexible rod (or filamentous) viruses: they are helical but not rigid enough to maintain a straight form in which all the identical subunits lie in identical environments.

In the third section of the paper, Caspar and Klug motivate the problem by summarizing the work of Crick and Watson (See section 2.3.2) and the theoretical advances made by Caspar and Klug since then. They presented the theory of quasi-equivalence as superceding the earlier speculations about economy and efficiency. Exactly why this is the case is left to the second proposed paper that was never published.

The fourth section, “The Geometry of Icosahedral Viruses,” contains the heart of the Caspar-Klug theory. The basic assumption is that “the shell is held together by the same type of bonds throughout but that these bonds may be deformed in slightly different ways in the different, non-symmetry related environments.” (Caspar and Klug 1962, p. 10) Caspar and Klug then appeal to Pauling’s study of antibody-antigen interactions to quantify how much deformation a bond is allowed before the bond would be broken. Pauling’s considerations allow up to 5 degrees of bending from the average value (Pauling 1953). This number then roughly defines the degree of “quasi” in their conception of quasi-equivalent bonding. From the general equation derived by Caspar, Caspar and Klug classify all possible icosahedral deltahedra. A deltahedron is a polyhedron whose faces are
equilateral triangles. They replace the “N” of the earlier equation with “T” which stands for “triangulation number”:

\[ T = Pf^2 \]

(where \( P = h^2 + hk + k^2 \), \( h \) and \( k \) any pair of integers with no common factor)

Every known type of virus has a unique triangulation number \( T \) and is made up of \( 60T \) subunits. They note that for deltahedra where \( P \geq 7 \), there can be skew cases. For example, both \( h=1, k=2 \) and \( h=2, k=1 \), lead to a \( T=7 \) structure. Because the same structure unit can build both cases, there is an increased probability of assembly “mistakes” and for this reason Caspar and Klug consider the occurrence of skew cases in nature unlikely (Caspar and Klug 1962, p.15).\(^{162} \) The built-in-curvature of a subunit would, in theory, limit the particular structure unit to self-assemble into a unique \( T \) number, but not a particular skew-type or handedness. Using these principles, Caspar and Klug argue that reports from electron microscopists of octahedral viruses are mistaken.

Caspar and Klug end their seminal article comparing helical and icosahedral designs. Since both can be constructed by quasi-equivalent bonding, further reasons are needed if nature might prefer one design to the other. A rod-shaped virus allows for more contact between the nucleic acid and protein and may result in a more stable particle. Furthermore the assembly process is relatively straightforward. On the other hand an icosahedral virus exposes a minimum amount of surface to the environment and does not require the shell to be completely disassembled to release the nucleic acid in the infection process.

\[^{162} \text{Later skew types would be discovered in nature, nonetheless.}\]
3.8 Michael Goldberg’s Rediscovered Insight

On October 29, 1962, with news that the Cuban missile crisis may have ended, Caspar wrote to Klug. Caspar mentioned that earlier in October, in Washington DC, he had met Ted Mattern who told him about a paper that had been sent to him by an amateur mathematician by the name of Goldberg.

Michael Goldberg (1902-1990) was raised in New York City on Monroe Street on the lower east side, the first American born child (third of eight siblings) of Sarah and Harry Goldberg, a polish immigrant. He died 9 March 1990 in Washington, DC at his home, 5823 Potomac Ave., NW.

Michael Goldberg was the only one of his siblings to finish high school, and he went on to University of Pennsylvania where he obtained his BS in Electrical Engineering in 1925. He had to take a year off during his studies to earn money to continue to pay for his education. He worked for a year at the Philadelphia Electric Company, doing such things as designing a high-tension electric transmission line. In 1926 he went to Washington, DC to take a “temporary” job with the Bureau of Ordnance, Navy Department. He wanted to work at the Bureau of Standards or the Naval Research Laboratory, but no position was immediately available. He remained with Bureau of Ordnance for 37 years, working on such things as gunfire control system design.

Mathematics was Goldberg’s enduring love and he pursued it throughout his life as a hobby (Goldberg 1980). He studied at George Washington University and obtained an MA in mathematics. The summer meetings of mathematics associations were regularly attended by Goldberg and his family.

Much of his work in mathematics involved geometrical visualization. He was interested in dissection and re-assembly problems such as cutting apart 2-D shapes into the minimum number of pieces needed to reconstruct different 2-D shapes with the resulting
parts. He was also interested in “packing problems” which involved efficient, closely spaced geometric structures. It was his work on efficiency problems that now piqued the interest of Caspar.

In 1937 Goldberg had published a paper entitled “A class of Multisymmetric Polyhedra” in the obscure Japanese Tôhoku Mathematical Journal. In this paper, Goldberg effectively derived the Triangulation number equation: \( T = h^2 + hk + k^2 \) (Goldberg 1937). He also displayed the triangulation numbers on a hexagonal grid, a figure that Caspar would adapt for his future publications (See figure 3.11).

---

163 Letter Caspar to Klug, 29 October 1962, Norman Archive.
Figure 3.11 Number of capsomere clusters represented on a hexagonal grid. Only 2.5% of each polyhedron represented. The numbers in the circles depicted in Goldberg’s grid are related by to Caspar and Klug’s T-numbers by the formula $60T = 6(n-12) + 60$, where $n$ is a number in a circle. From Goldberg (1937).

Following Steinitz’s work in 1927, Goldberg had become interested in the “Isoperimetric Problem” for polyhedra: determining which convex polyhedra has the greatest volume for a given number $n$ faces (Steinitz 1927; Goldberg 1934; See Tarnai (1993) for the mathematical connections with geodesic domes). Goldberg presented this problem at the American Mathematical Society meeting in Louisiana in 1931. He discovered that one can make progress on this problem by considering the “medial polyhedra,” those polyhedra that possess trihedral vertices and only pentagonal or hexagonal faces. (By triangulating the hexagons and pentagons, one creates the figures that Caspar and Klug consider.) In his 1937 article, he considers the degeneracy of the triangulation numbers. Consider a figure described by $T=49$. This figure arises either when $h=7, k=0$ or when $h=5, k=3$. Goldberg was able to show that for any number $n$, there is a T number that has at least $n$ different
solutions. This multiple realizability has little biological significance, but Goldberg had explored features of triangulation numbers not considered by Caspar and Klug.

3.9 Historical Summary

In the 1950s and early 1960s the field of structural virology blossomed. Theoretical speculation was driven by new data generated by biochemistry, electron microscopy, and x-ray crystallography. In the early to mid 1950s, Francis Crick and James Watson speculated that “spherical viruses” possess cubic symmetry, probably icosahedral symmetry. Using single crystal x-ray diffraction, Donald Caspar confirmed that Bushy Stunt Virus possesses icosahedral symmetry in Cambridge in 1955. Watson showed the rod-shaped TMV virus to be helical and Rosalind Franklin later determined its exact helical parameters. Aaron Klug showed that turnip yellow mosaic virus and poliovirus also possessed icosahedral symmetry. The Crick-Watson framework appeared adequate until biochemical results and better electron micrographs showed that many viruses contained more than the 60 identical subunits suggested by the Crick-Watson theory. Aaron Klug and Don Caspar attempted to solve the problem. An analogy with geodesic domes allowed them to conceive an innovative new type of symmetry they called “quasi-equivalence.” The insight was to see that identical subunits could be bonded together in quasi-equivalent ways to build structures with more than 60 subunits. They generalized Buckminster Fuller’s results to include “skew” structures and derived a single formula that describes all possible structures. Later, they discovered that an amateur mathematician named Michael Goldberg, had effectively derived the same formula in the 1930s.

3.10 Some Preliminary Philosophical Lessons

This extended case history holds a number of lessons. These will not be fully explicit until later chapters when more philosophical argument is in place, but let me quickly draw some lessons that are already apparent relevant. First, there is nothing in the historical
record that suggests that Caspar and Klug adopted their theory of virus structure because of its beauty. Rather they had experimental evidence that supported the number and arrangement of protein subunits that their theory proposed. Further there is nothing in the historical record that suggests that they thought it more likely because it was beautiful.

Second, the history shows how there can be alternative explanations of the beauty of a given theory. Let me remind you that I selected Plato’s theory of elements, Kepler’s theory of planetary distances, Crick and Watson’s and Caspar and Klug’s theories of virus structure because they were beautiful in virtue of the symmetries they postulated. I stand behind the claim that the theories’ postulated symmetry is the most plausible cause of these theories’ beauty. However, the theories selected also exhibit other aesthetic properties. For example, consider Kepler’s theory of planetary distances unifies otherwise disconnected facts. The relative distances of the 6 then known planets would seem to be independent facts. On Kepler’s accounts these facts are unified and flow from his model. A Whewellian would view this as a good reason for believing that Kepler’s theory is beautiful (and true). A similar picture can be painted for the Caspar-Klug theory. Caspar’s formula specifies the exact number of subunits that any sized spherical virus has. Seemingly unconnected facts, such as that turnip yellow mosaic virus has 180 subunits and that herpes virus has 960, are seen as two different applications of the same principle.
Chapter 4  Does Beauty Render a Theory Likely?

Structure:

4.1  Beauty as a Sign, of Truth

4.2  The Form of the Aesthete’s Argument

   4.2.1  Assessing the Aesthete’s First Argument

   4.2.2  Whewell on Coherence, Beauty, and Truth

4.3  The Problem of Beautiful Rival Theories

4.4  The Aesthete’s Short-Run Rejoinder

   4.4.1  Does Beauty Influence the Likelihood of a Theory?

   4.4.2  Does Beauty Influence the Prior Probability of a Theory?

4.5  Beauty and Belief

4.1  Beauty as a Sign, of Truth

The primary conclusion that I have drawn from Chapter 3 is that Caspar and Klug did not appeal to aesthetic reasons to justify their theory. They did not believe that the beauty of the Caspar-Klug theory rendered the theory likely or increased the probability of the theory. However, for a philosopher of science, it is not enough to show that in fact Caspar and Klug did not appeal to aesthetic reasons such as the beauty of the posited icosahedral and quasi-equivalence symmetries. I need to supplement these descriptive observations with prescriptive arguments against any alleged justification of the Caspar-Klug theory in virtue of its beauty. I will attempt to make these arguments general enough to apply to other cases of theoretical beauty. I will counter the objection that it still would have been rational for Caspar and Klug to believe their theory in virtue its beauty. As I suggested in Chapter 1, let us call someone who adheres to this type of position an “aesthete”. The most ardent aesthete claims that beauty is a sign, of truth, a necessary
condition of which is that $p(T/ T \text{ is beautiful}) > k$, where $k \geq 0.5$. If beauty were a sign of truth, it would be more rational to believe a beautiful theory was true than to believe it was false. In the case of the Caspar-Klug theory, which is beautiful (at least) in virtue of the beautiful symmetries it posits and its lack of ugly properties, the aesthete believes that beautiful symmetry in the absence of countervailing ugly properties is a sign of truth. In this section I will argue that the beauty cannot be an a priori sign of truth. I will illustrate my argument with the Caspar-Klug theory. (In the next chapter, I consider whether beauty is sign of truth, i.e., whether $p(T/T \text{ is beautiful}) > p(T)$ is the case.) If the beauty were a sign of truth and a theory’s beauty were determinable without empirical inquiry, then it would be possible to justify a belief in a beautiful theory without testing it experimentally. This would be a significant achievement. Alas, the aesthete is mistaken or so I shall argue in this chapter.

In Chapter 1, I defined a necessary condition for a sign of truth as $p(T/T \text{ is beautiful}) > 0.5$. Clearly this is quite a tall order to satisfy. Consider a competing theory $T^*$, such that $\neg(T \& T^*)$. If $T^*$ is also beautiful, then it cannot be the case that $p(T/T \text{ is beautiful}) > 0.5$ and $p(T^*/T \text{ is beautiful}) > 0.5$ since the probability of mutually exclusive theories cannot sum to greater than 1. There are two types of response the aesthete can make to this argument. First, the aesthete might claim that there must be an independent reason to think that an alternate beautiful theory can be formulated. Skeptically motivated theories need justification before they are considered genuine competitors of existing theories. Good scientists do not seriously consider a “theory” such as: Descartes’ evil demon created the world to be as if $T$ were true. On the other hand, the history of science gives us some reason to think that often there will be more than one “scientifically respectable” beautiful theory of a given phenomenon. For example, the particle and wave theories of light in the 19th century competed for the correct explanation of optical phenomena. Second, the aesthete might argue that useful judgments of probability of a theory are not made in a vacuum, but rather are made in the context of additional evidence.
and consequently we should consider the combined epistemic role of beauty and the additional evidence. This complaint voices a legitimate point. The goal of using signs of truth is to learn what one should believe; any such judgment would likely be made on the basis of more evidence than merely the beauty of T.

4.2 The Form of the Aesthete’s Argument

No one, not even the arch-aesthete physicist Paul Dirac, advocates that we use beauty as the sole criterion of belief acceptance. To make the aesthete’s position as cogent as possible, let us consider the role of beauty in the belief in empirically adequate theories. Let us construe empirical adequacy logically: theory T is empirically adequate with respect to evidence E if and only if T entails E. I use the adjective “empirically adequate” for theories that are empirically adequate with respect to all known relevant evidence that falls within the scope of T.\(^{164}\) It follows that, for any T, if T is empirically adequate with respect to E, then \(p(T/E & T \text{ is beautiful}) \geq p(T/T \text{ is beautiful})\). If I can show that even in cases of empirically adequate theories, \(p(T/E & T \text{ is beautiful}) < 0.5\) for some T and E, then it follows that there exists a T such that \(p(T/T \text{ is beautiful}) < 0.5\), i.e., that beauty is not a sign of truth. This result should not be too surprising, since if a theory is not rationally believable when both empirical evidence and aesthetic considerations pull in the same positive direction, then it is unlikely that the theory will be rationally believable when there is no relevant empirical evidence or evidence that pulls against the positive aesthetic evidence for the theory. If beauty does not render an empirically adequate theory rationally believable, then it is implausible that it renders an empirically inadequate theory rationally believable. In this section, I will argue that the aesthete’s position is misguided for a

\(^{164}\) It is difficult to give a simple technical definition of the scope of T. For the case at hand, the scope of the Caspar-Klug theory of spherical virus structure is not surprisingly phenomena about the structure of spherical viruses.
number of reasons. My strategy is to charitably make explicit the aesthete’s argument and then identify the faulty reasoning.

Before examining the aesthete’s argument in detail. Let me distinguish between two different ways that the aesthete can proceed. For want of better terms, let us call the two approaches the broad approach and the narrow approach. On the broad approach, the aesthete argues that theoretical beauty more generally is a sign of truth. So far I have framed the aesthete’s position broadly. In contrast, on the narrow approach, a particular aesthetic property, such as coherence or symmetry, is a sign of truth. The two approaches collapse into one for aesthetic monists who think that there is only one aesthetic property that determines whether a theory is beautiful. For those that reject aesthetic monism, there is a distinction between the two approaches—one could succeed even if the other fails. Furthermore, there are potentially as many variations on the narrow approach as there are aesthetic properties, i.e., each variation considers a different aesthetic property as a sign of truth. I will not have space to examine every possible variation on the narrow approach. Nonetheless, many arguments against the broad approach will have narrow analogs. For the remainder of the chapter, I will consider both the broad approach and two narrow approaches: high degree of symmetry and theoretical coherence.

Consider the following argument schema, which makes explicit the broad form of the aesthete’s epistemic appeal to beauty and embeds it in an inferential context:

1. Empirically adequate theory Ta has aesthetic property P.
2. All empirically adequate theories with property P are beautiful.
∴ 3. Empirically adequate theory Ta is beautiful.
4. (∀T)p(T/T is empirically adequate & T is beautiful) > 0.5
∴ 5. p(Ta / Ta is empirically adequate & Ta is beautiful) > 0.5
6. It is rational to believe a likely theory.
∴ 7. It is rational to believe Ta.
The aesthete’s position is really three separate, but related arguments: Argument A consists of (1),(2) \(\therefore\) (3); Argument B, which represents the broad approach, consists of (3),(4) \(\therefore\) (5); and Argument C consists of (5),(6) \(\therefore\) (7). Overall, beginning from premises about an aesthetic property of a theory, the aesthete draws conclusions about what it is rational to believe. Put simply, the aesthete connects a property of a theory to probability of truth, via an appeal to the beauty of the theory in virtue of the property and an alleged connection between beauty and truth. Premise (2) connects the property \(P\) to beauty and premise (4) connects beauty to truth. To turn this argument schema into an actual aesthete’s position, one has to instantiate for property \(P\). As mentioned in Chapters 1 and 2, different thinkers have put forward different candidates for property \(P\). Plato took property \(P\) to be high degree of symmetry, Newton took it to be simplicity, and Whewell took it to be coherence. I will argue against both of these purported connections (premises 2 and 4) by looking at two aesthetic properties in more detail: symmetry and coherence. The arguments will be largely generalizable to other aesthetic properties. Before I begin the more concrete investigation, let me address an obvious objection to premise 2. That a given aesthetic property would render a theory beautiful depends upon what other properties that the theory possesses. For example, a theory might also possess ugly properties that swamp the effects of the positive aesthetic property \(P\). For this reason, let us build into premises 2 and 3, the requirement that the theory in question does not possess any ugly properties. With this modification in mind the argument A for coherence is:

(1c) Empirically adequate theory \(T_a\) is coherent and lacks ugly properties.

(2c) All empirically adequate \textit{coherent} theories that lack ugly properties are beautiful.

\(\therefore\) (3) Empirically adequate theory \(T_a\) is beautiful.

And for Plato and others who focus on symmetry as the aesthetic property \textit{par excellence}: 

\begin{itemize}
  \item The aesthete’s position is really three separate, but related arguments: Argument A consists of (1),(2) \(\therefore\) (3); Argument B, which represents the broad approach, consists of (3),(4) \(\therefore\) (5); and Argument C consists of (5),(6) \(\therefore\) (7). Overall, beginning from premises about an aesthetic property of a theory, the aesthete draws conclusions about what it is rational to believe. Put simply, the aesthete connects a property of a theory to probability of truth, via an appeal to the beauty of the theory in virtue of the property and an alleged connection between beauty and truth. Premise (2) connects the property \(P\) to beauty and premise (4) connects beauty to truth. To turn this argument schema into an actual aesthete’s position, one has to instantiate for property \(P\). As mentioned in Chapters 1 and 2, different thinkers have put forward different candidates for property \(P\). Plato took property \(P\) to be high degree of symmetry, Newton took it to be simplicity, and Whewell took it to be coherence. I will argue against both of these purported connections (premises 2 and 4) by looking at two aesthetic properties in more detail: symmetry and coherence. The arguments will be largely generalizable to other aesthetic properties. Before I begin the more concrete investigation, let me address an obvious objection to premise 2. That a given aesthetic property would render a theory beautiful depends upon what other properties that the theory possesses. For example, a theory might also possess ugly properties that swamp the effects of the positive aesthetic property \(P\). For this reason, let us build into premises 2 and 3, the requirement that the theory in question does not possess any ugly properties. With this modification in mind the argument A for coherence is:

(1c) Empirically adequate theory \(T_a\) is coherent and lacks ugly properties.

(2c) All empirically adequate \textit{coherent} theories that lack ugly properties are beautiful.

\(\therefore\) (3) Empirically adequate theory \(T_a\) is beautiful.

And for Plato and others who focus on symmetry as the aesthetic property \textit{par excellence}:
(1s) Empirically adequate theory Ta postulates symmetry and lacks ugly properties.
(2s) All empirically adequate theories *that postulate symmetry* and lack ugly properties are beautiful.
∴ (3) Empirically adequate theory Ta is beautiful.

For the remainder of this chapter, I will consider the Platonic and Whewellian broad positions separately. It will helpful to have a specific theory in mind. I will consider the Caspar-Klug theory whose beauty arguably could be due to symmetry or coherence (or both). First consider the view that the Caspar-Klug theory is beautiful due to the high degree of symmetry it postulates. The aesthete’s position appears to be the following:

(1s\textsubscript{CK}) The empirically adequate Caspar-Klug theory postulates particles with a high degree of symmetry and lacks ugly properties.

(2s) All theories that postulate entities with a high degree of symmetry and lack ugly properties are beautiful.
∴ (3\textsubscript{CK}) The empirically adequate Caspar-Klug theory is beautiful.

(4) (\forall T) p(T/T is empirically adequate & T is beautiful) > 0.5
∴ (5\textsubscript{CK}) p(Caspar-Klug theory is true / It is empirically adequate & It is beautiful) > 0.5

(6) It is rational to believe a likely theory.
∴ (7\textsubscript{CK}) It is rational to believe the Caspar-Klug theory.

4.2.1 Assessing the Aesthete’s Argument

Let us consider each of the aesthete’s premises one by one to see which are defensible. Consider (1s\textsubscript{CK}). It is true that the Caspar-Klug theory posits highly symmetrical theoretical entities. The viral structures posited by the Caspar-Klug theory have 532 point symmetry. Of all point symmetries, only spherical symmetry is more symmetrical than 532 symmetry. It also posits a subtle previously-unrecorded form of symmetry that Caspar and Klug call quasi-equivalence. Further, the theory has no significant ugly properties. If the theory did possess significant ugly properties, then
luminaries such as Monod would not call it beautiful. However, in the early 1980s, a counter-example, polyoma virus, was discovered, suggesting that the Caspar-Klug theory is not completely empirically adequate (Rayment 1982). Nonetheless, in response to the polyoma virus case, most virologists have not rejected the theory as false, but merely restricted its scope to exclude such cases.\footnote{Although see Twarock (forthcoming) for an attempt to broaden the Caspar-Klug framework.} Let us ignore these complications for the time being and accept premise (1s\textsuperscript{CK}).

Before considering premise (2s) directly, let us examine some related claims. A proponent of premise (2s) might further suggest (or be motivated by the claim) that there is a monotonic relationship between beauty and symmetry. Before you reject this claim as absurd, consider that if beauty and symmetry are synonymous in some domain then they will stand in a strict-monotonic relationship in that domain. Some philosophers have gone so far as to claim that symmetry is synonymous with beauty in certain domains. The philosopher of science, Peter Kosso, for example, while discussing the use of symmetry arguments in physics, claims that, “it is safe to say that beauty essentially means symmetry in [physics]” (Kosso 1999, p. 482). He offers little argument for this claim, but gives several quotations from eminent physicists. If beauty and symmetry were synonymous, then an object with more symmetry would be more beautiful than one with less. Furthermore, an object with the most symmetry would have the most beauty. However \textit{perfect} symmetry in science, or even outside of science for that matter, is not commonly judged to be the most beautiful. For example, some degree of asymmetry often is required for a beautiful theory or a great work of art. Likewise, Marilyn Monroe’s asymmetrically located beauty spot was said to have enhanced her beauty, not diminished it. The most geometrically symmetrical object I can imagine is an empty infinite Euclidean space. This space has every type of geometrical symmetry at every point and plane. In other words, any translation, mirror reflection, or rotation of the space leaves every feature of the space
invariant. The addition of any finite object into the space creates a combined entity that has infinitely less symmetry. Is an empty space more beautiful than one with objects? I don’t believe so. Indeed, you might say that empty space is too boring to have the maximal amount of beauty. Furthermore, this case illustrates another difference between beauty and symmetry: while there can be a maximal amount of symmetry in any given system, it is not even clear that the notion of a maximal amount of beauty makes sense. Beauty may be like the natural numbers: just as there is no largest number, there may be no largest amount of beauty that an object can have. Even Dirac who openly advocates we search for beauty in physics does not advocate that we maximize the number of symmetries. Indeed, for a theory to be interesting and beautiful requires that there be asymmetries as well as symmetries. Even our most beautiful theories in particle physics describe asymmetries called “broken symmetries” as well as the symmetries that Kosso admires\(^6\) (Close 2000).

It is too simplistic to say that symmetry is synonymous with beauty. It is also too simplistic to say that all theories whose posits are highly symmetrical are beautiful. The clearest exception to this generalization is a theory with highly symmetrical posits that also possesses ugly additional properties such as ugly laws governing the symmetrical posits. Premise (2s) explicitly excludes such a case, but nonetheless, it is still false. It is possible that a theory may possess a beautiful property, such as positing beautiful symmetries, possess no ugly properties, but fail to be beautiful. It may not be ugly either, but there is a middle ground between beauty and ugliness. A man may have beautiful hair, no ugly features, but nonetheless not be beautiful. The same is true of theories. Thus, I reject premise (2s).

\(^6\) To be fair, Kosso discusses broken symmetries in another paper, although he does not discuss their aesthetic appeal or lack of it (Kosso 2000). See Morrison (1978) for a discussion of the beauty of broken symmetries. On a related point, Martin (1989) argues that theories are beautiful and valuable insofar as they balance between potentially conflicting criteria. Presumably symmetry and asymmetry would be
Even if beauty and symmetry are not synonymous, they are closely related. Rather than consider the degree of symmetry, a more defensible claim is that particular symmetries are beautiful. This claim makes sense in light of the array of positions outlined in Chapter 2. Plato and Kepler judged cubic point symmetry as beautiful. Aristotle, Ptolemy, and Copernicus judged spherical point symmetry to be beautiful, while contemporary physicists judge more abstract symmetries, such as the Lorentz transformation, to be beautiful. In the case study at hand, icosahedral point symmetry and quasi-equivalence are judged to be beautiful. This diversity of opinion need not prompt a debate over who is right and who is wrong. If spherical symmetry is beautiful it does not follow that icosahedral symmetry is not. It is possible that all the above thinkers are correct; different epochs may have discovered new particular beautiful symmetries.

Given the rejection of premise (2s), the aesthete’s argument for \(3^{CK}\) is unsound. Premise 3 says: “The empirically adequate Caspar-Klug theory is beautiful.” While the argument for \(3^{CK}\) is unsound, I think the conclusion is nonetheless true since the Caspar-Klug theory possess two particular beautiful symmetries that are sufficient to render the entire theory beautiful. Indeed, the reason I chose to focus my dissertation on the Caspar-Klug theory was because it is one of the most beautiful biological theories. Furthermore, in addition to symmetry, it possesses other aesthetic properties such as coherence.

e.g. examples of such criteria. Lipscomb (1980) makes the same point.
4.2.2 Whewell on Beauty and Truth

Let us now turn to a Whewellian argument for the beauty of the Caspar-Klug theory. To reiterate, what I am calling the Whewellian position says:

(1c<sub>CK</sub>) The empirically adequate Caspar-Klug theory is coherent and lacks ugly properties.

(2c) All empirically adequate theories that are coherent and lack ugly properties are beautiful.

∴ (3<sub>CK</sub>) The empirically adequate Caspar-Klug theory is beautiful.

The first premise (1c<sub>CK</sub>) says that the Caspar-Klug theory is coherent. Undoubtedly, Caspar and Klug’s formula (T = h^2 + hk + k^2) applies to a large variety of viruses: it explains why the shells of BSV and polio virus have 180 protein subunits, why herpes virus has 960 protein subunits, and why canine parvovirus has 60 protein subunits, to give but three examples. To be precise, the Caspar-Klug rule does not explain why a given virus has one allowable T-number rather than another allowable T-number, but merely why it does not have a non-allowed number of subunits, for example why no virus has 120 quasi-equivalent protein subunits (T=2 is not a solution to the above equation). The Caspar-Klug then appears to unify a number of disparate facts about viruses.

However, there are two potential problems with the claim that the Caspar–Klug theory is coherent in Whewell’s sense: (1) that the facts unified do not appear to be from different classes, and (2) none of the facts unified appear to be facts that Caspar and Klug would have considered irrelevant to their theory when the created it. To see why these problems are relevant, it will be useful to explicate Whewell’s notion of coherence in more detail. What Whewell calls coherence is what he calls consilience over time. We must begin then by first understanding what Whewell means by consilience. Consider the following quotation:

We have here spoken of the prediction of facts of the same kind as those from which our rule was collected. But the evidence in favour of our induction is of a much higher and
more forcible character when it enables us to explain and determine cases of a kind different from those which were contemplated in the formation of the hypothesis. The instances in which this has occurred, indeed, impress us with a conviction that the truth of our hypothesis is certain. No accident could give rise to such an extraordinary coincidence …

Accordingly the cases in which inductions from class of facts altogether different have thus jumped together belong only to the best established theories which the history of science contains. … I will take the liberty of describing it by a particular phrase; and will term it Consilience of Inductions. (Whewell 1847, p. 230; italics added)

Whewell’s idea is that a consilient theory correctly predicts unobserved phenomena of a kind different than it was designed to explain. In so doing, a consilient theory unifies different classes of phenomena. Let me make his idea more precise. Let there be two classes of observable phenomena, O₁, … Oₙ and Oₙ₊₁, … Oₙ₊ᵐ. Imagine scientist S₁ constructs a theory T₁ that entails O₁, … Oₙ and Oₙ₊₁, … Oₙ₊ᵐ, but did not contemplate Oₙ₊₁, … Oₙ₊ᵐ in formulating her theory. In this case, T₁ would be a consilient theory.

Whewell’s notion of consilience has both an objective dimension and a subjective dimension. If observable phenomena O₁, … Oₙ and Oₙ₊₁, … Oₙ₊ᵐ are different kinds of phenomena, then this fact does not depend on whether anyone believes that this is the case or not.¹⁶⁷ Whewell thinks that it is an objective fact whether two phenomena are of different classes or not. On the other hand, whether S₁ considers or contemplates certain phenomena in formulating her theory depends trivially on the propositional attitudes of S₁ and thus Whewell incorporates a subjective condition into his notion of consilience. The notion of “not contemplating” phenomena Oₙ₊₁, … Oₙ₊ᵐ is vague and I think Whewell means something like that S₁ did not believe that phenomena Oₙ₊₁, … Oₙ₊ᵐ were relevant to the theory being formulated at the time of its formulation. To say that members of the second
class have to be *unknown* to $S_1$, not merely that $S_1$ would not have contemplated them, is too strong. It, for example, rules out one of Whewell’s favorite examples of consilience: Newton’s theory of gravity which unifies Kepler’s 2nd and 3rd laws, both of which were known by Newton. Most observations unknown to $S_1$ at time $t$, would not be contemplated by $S_1$ at $t$. Whewell does not discuss the possibility of contemplating future (as yet unknown) observations in the construction of a theory, but presumably this sort of contemplation would count against the consilience of a theory if the theory turned out to entail the observations contemplated. One might also wonder about a case in which $S_1$ contemplates a proper subset of the second class of phenomena, $O_{n+1}, \ldots O_{n+m}$. I think Whewell would still say that this case would still be a case of consilience, but perhaps a case that has less epistemic force.

Notice that whether a theory is consilient is relative to its discoverer. To see why, consider a theory $T_1$ that has been multiply discovered. As above, suppose that $S_1$ discovered $T_1$ but did not contemplate a class of phenomena, $O_{n+1}, \ldots O_{n+m}$, that $T_1$ predicts. Thus it would appear that $T_1$ is consilient. However, if a second researcher $S_2$ also independently hit upon $T_1$, but *did* contemplate both $O_1,\ldots O_n$ and $O_{n+1}, \ldots O_{n+m}$ in her formulation of $T_1$, then relative to $S_2$ this would *not* be a case of consilience. Thus, as presented, we cannot ask whether a theory is consilient *per se*, but must ask whether a theory is consilient given a particular case of its construction. We should ask, “is $T_1$ consilient relative to $S$?” Is such relativity a problem for Whewell? I think it is. This relativity raises serious problems in using Whewell’s concept. For example, suppose a third researcher $S_3$ is considering how to evaluate theory $T_1$. Remember, for Whewell, a theory’s consilience strengthens the evidence in its favor. Two facts appear relevant to the evidence for $T_1$: (1) $T_1$ is consilient relative to $S_1$ and (2) $T_1$ is not consilient relative to $S_2$. The question is, which of these facts should a third person, $S_3$, use when we calculate the

---

167 Unless $O_n$ or $O_{n+1}$ are beliefs themselves, but this complication need not concern us here.
epistemic status of the theory? Should $S_3$ judge that $T_1$ has strong evidence in its favor or not? Whewell might be willing to say that $S_1$ has more evidence for $T_1$ than $S_2$, but Whewell says little about how $S_3$ should evaluate his theory. Which of the two discoveries should $S_3$ weigh more heavily? Should they be given equal weight or can we ignore one discovery? If they get differential weight, how do we justify the inequality? If another person $S_4$ were also to come up with $T_1$ after contemplating all the phenomena, would this degrade the evidence in favor of $T_1$? Whewell’s apparatus does not help us make these decisions. Most of us will never discover a consilient theory, so are in the uneasy position of $S_3$. To make an interpretative move on behalf of Whewell, let us say that a theory $T$ is consilient if there exists at least one person $S$ who discovered $T$ but did not contemplate a class of phenomena that $T$ predicts. This move makes Whewell’s notion more useful and removes that implicit relativization to a person. I will have much more to say about the epistemic role of consilience below.

With the notion of consilience further qualified, let me define Whewellian coherence as merely continued consilience over time. A theory is coherent if, as new kinds of phenomena are observed, the theory explains them without having to introduce ad hoc assumptions, i.e., if it repetitively unifies new kinds of phenomena with little or no modification. Thus correct application of this notion, like consilience, depends on the historical development of the theory in question. If a theory has a history of successful consilience, then it is coherent. As Whewell puts it:

… we have to notice a distinction which is found to prevail in the progress of true and false theories. In the former class all the additional suppositions tend to simplicity and harmony; the new suppositions resolve themselves into the old ones, or at least require only some easy modification of the hypothesis first assumed: the system becomes more coherent as it is further extended. The elements which we require for explaining a new class of facts are already contained in our system. Different members of the theory run
together, and we have thus a constant convergence to unity. In false theories, the contrary is the case. (Whewell 1847, p. 233; italics added)

As the quotation suggests, Whewell allows for coherence to come in degrees. At least in certain cases, one can compare theories and judge them to be more or less coherent (once we know their history).

Before we examine whether coherence is linked to truth as Whewell suggests, let us see how well Whewell’s position fits into the aesthete’s argument schema discussed above. Does Whewell think coherence is a beautiful property of theories? As mentioned in Chapter 1, there is only circumstantial evidence for the view that Whewell himself thought that coherence was necessarily a beautiful property or that coherent theories were necessarily beautiful. He does think coherence is closely related to harmony: a more coherent theory tends to harmony. He also argues that we should clearly separate the practical arts from the sciences and that incoherent phenomena “disturbs [Art] not” implying that a good scientist should find incoherence disturbing and presumably something to be avoided (Whewell 1847, p. 110). On the other hand, for Whewell, whether coherent theories are beautiful would be a question for the science of the perception of beauty (in fine arts), or what he calls “Callaesthetics,” which he distinguishes from the practical arts (Whewell 1847, p. 568). Beyond these tantalizing suggestions, Whewell does not explicitly discuss the relation of coherent theories and beauty anywhere in his corpus to my knowledge. Nonetheless, the general tenor of his great work suggests that he links coherence to harmony, progress, and beauty in science—an optimistic view not uncommon in the mid to late 19th century. And, as I said earlier, I think it is no accident that “beauty” is the last word of Philosophy of the Inductive Sciences.

Let us extend Whewell’s position and suggest that a plausible Whewellian view would involve the claim that T is beautiful if T is coherent. As Whewell also holds that if T is coherent, then T is true as suggested by his quotations, it is highly likely that Whewell holds the view that beauty is a sign, of truth. Indeed to be charitable to Whewell, let us
relax the assumption that any coherent theory is true and replace it with the more plausible view that any coherent theory is likely. This relaxation allows for the possibility of false coherent theories. Even with this relaxation, the Whewellian still holds that beauty is a sign of truth. The important point is that for Whewell the connections between coherence and truth is a priori and if we can judge the connection between coherence and beauty a priori also, which seems plausible given Whewell’s view that science (including Callaesthetics) will eventually produce self-evident truths, then the Whewellian believes that beauty is an a priori sign of truth. Whewell’s reasoning might go something like this:

Premise 1: Theory $T_1$ is consilient at $t = 0$
Premise 2: Theory $T_1$ is consilient at $t = 1$

\[ \vdots \]
Premise $n+1$: Theory $T_1$ is consilient at $t = n$, where $n$ is some finite number
\[ \therefore \text{(by definition of coherence)} \text{ Theory } T_1 \text{ is coherent at } t = n \]
\[ \therefore \text{(inductively) Theory } T_1 \text{ will continue to be coherent for all } t \]
\[ \therefore \text{(a priori) Theory } T_1 \text{ is likely} \]

The overall argument mixes empirical and allegedly a priori inferences. The most interesting inference is the last one.

With Whewell’s notion of coherence clarified, let us return to the question of whether the Caspar-Klug theory is coherent. I have argued that Whewell thinks the general connection between beauty and coherence is a priori. However, for Whewell, I suggest that whether a particular theory is coherent depends upon its history and therefore is empirical. We must consider the history of the Caspar-Klug theory to determine first if it is consilient and then whether it is coherent. The Caspar-Klug theory does indeed foretell new phenomena, arguably even phenomena not known or contemplated by Caspar and Klug. They knew of a handful of spherical virus structures, notably tomato busy stunt virus and turnip yellow mosaic virus, and subsequently many more viruses have been shown to satisfy
the Caspar-Klug selection rule. (There have also been some exceptions, but let us ignore these for now.) On the other hand, one could argue that Caspar and Klug contemplated the application of their theory to other spherical viruses even if the relevant data did not yet exist when they created their theory. One could argue that these types of phenomena were expected and would have been considered relevant to Caspar and Klug. Nonetheless, the Caspar-Klug theory has been very successful in explaining new phenomena and they certainly did not contemplate the particular details of future data since they were not known.

It is not enough for a consilient theory to foretell phenomena, however, for Whewell. It must also apply to different kinds of phenomena. Are data from different virus species different kinds of phenomena? Prima facie, we might say no, since it appears to apply only to one “class” of virus, the spherical viruses. However, in the late 1950s many expected major structural differences between viruses whose hosts are very different kinds of organism. For example, one might argue that the structure of animal viruses and plant viruses are different kinds of phenomena given they produce very different effects—mosaic tobacco leaves appear quite unlike symptoms of polio. Ironically given the success of the Caspar-Klug theory, major structural differences between plant and animal viruses are no longer thought to exist. This is one curious feature of Whewell framework. If we accept a consilient theory, we accept, among other things, that phenomena that we previously thought were from different classes are now thought to be merely different manifestations of the same type of phenomenon. Those of us with the intuition that epistemic markers of success should bear logical and not historical relations to the evidence will find this feature of Whewell’s framework implausible. Why should we base our epistemic judgments of T upon a consequence—that the phenomena form different classes—of an antiquated theory or antiquated common sense that we have rejected in accepting T? Nonetheless, on the definition of consilience provided here, the Caspar-Klug theory is an example of a consilient theory, if any biological theory is. Furthermore, with simple modifications, the Caspar-Klug theory could be applied to the class of carbon molecules called Fullerenes discovered in
1985.\textsuperscript{168} By and large, the Caspar-Klug theory has been consilient up until now. Thus we are permitted to make the tentative judgment that the Caspar-Klug theory is coherent. It is always possible that the Caspar-Klug theory might not continue to be consilient and so lose its coherent status, but at present we have good evidence to believe that it is.

Is Whewell correct? Is a coherent theory beautiful? This question is vital to the viability of a Whewellian aesthete’s broad argument for the epistemic relevance of beauty. (Below I will consider the narrow argument that coherence, and not necessarily beauty, is a sign of truth) To mirror the approach I took with symmetry above, we might ask the question relatively: is a more coherent theory always more beautiful than a less coherent theory? This question goes beyond the Whewellian framework so far described that does not include an account of more or less coherence (or consilience). Indeed, additional problems arise when we attempt an ordering of theories in terms of their degree of coherence. Is a theory that unifies fewer phenomena but from more classes more coherent than one that unifies more phenomena from fewer classes? Whewell’s position is not rich enough to answer this question.

We can still make progress without an answer to this question, however. A weaker partial ordering can be achieved if we assume that of two theories, $T_1$ and $T_2$, each intended to capture the same sets of phenomena, theory $T_1$ is more coherent than coherent theory $T_2$ if $T_1$ entails all the phenomena that $T_2$ does and some additional phenomena of a different kind. Notice that on this assumption we sanitize Whewell’s notion by removing any reference to the history or discoverer of the theory. Sanitized coherence, for want of a better name, is an objective notion. Roughly the idea is that if the scope of $T_2$ is a proper subset of the scope of $T_1$, accounts for less kinds of phenomena than $T_1$, and $T_2$ is coherent, then $T_1$ is more coherent that $T_2$. On this principle, an empirically adequate “Theory of Everything,”

\textsuperscript{168} Although Caspar foreshadows this discovery in a letter (Letter Caspar to Klug, 7 February 1961, Norman Archive).
as envisioned by theoretical physicists would be the most coherent theory. Would such a
theory be the most beautiful theory? Many physicists seem to think so. However, if other
properties beyond coherence are relevant to a theory’s beauty then it is possible that a
theory of everything possess ugly properties allowing a less coherent theory to be more
beautiful. To avoid this possibility the aesthete might build in a ceteris paribus clause:
ceteris paribus, a more coherent theory is more beautiful. Nonetheless, the nature of
coherence does not require that there be a unique most coherent theory. It is conceivable
that two or more theories could have the same final scope and intended scope. On my view,
the exact number of empirically adequate theories for a given domain is an empirical
question that cannot be decided a priori either for or against uniqueness. This fact will be
relevant below. It is hubris to think that the first coherent theory of a given domain we
formulate will be the only one.

Perhaps the most contentious claim of the Whewellian aesthete is that coherence as
an aesthetic property is an empirical, indeed an historical, property of a theory. It is
certainly counterintuitive to say that to judge the beauty of a theory we have to look at the
intentions of its creator and what she considered relevant when constructing the theory.
Indeed, scientists with refined aesthetic taste ought to be able to judge the beauty of a theory
once they know the content of the theory. Questions of its historical development are
irrelevant. Otherwise, we would be forced to say that one could not know whether a given
theory is beautiful without first knowing the history of science in sufficient detail to know
which facts the creator of the theory thought was relevant to its construction. This
consequence runs counter to the way scientists judge theoretical beauty. Many scientists
make judgments about the beauty of theories while remaining ignorant of the intentions of
those that formulated the theory. These judgments are not merely provisional versions of a
more secure judgments made in light of a theory’s history. This is the reason for
considering a sanitized notion of coherence by removing the condition that the intentions of
theorist are relevant. A theory that exhibits sanitized coherence repeatedly unifies different
kinds of phenomena regardless of whether the formulator of the theory contemplated them or not.

Nonetheless, the Caspar-Klug theory is a beautiful theory. One reason it is beautiful is that with a simple formula it covers a wide variety of cases. Another reason that it is beautiful is that it posits two beautiful particular symmetries in nature. With a few caveats mentioned above, let us grant the truth of premise (3\(^{\text{CK}}\)): The empirically adequate Caspar-Klug theory is beautiful. Let us continue looking at the aesthete’s argument, keeping in mind the two ways in which an aesthete may justify the claim that the Caspar-Klug theory is beautiful.

The most significant problems for the aesthete’s argument arise in justifying premise (4) \(\forall T)p(T/T \text{ is empirically adequate} \& T \text{ is beautiful}) > 0.5\). As described in Chapter 1, \(\forall T)p(T/T \text{ is beautiful}) > 0.5\) is a necessary condition on an absolute conception of a sign or truth, what I called a sign\(_1\) of truth. The sign\(_1\) is not infallible – an infallible sign of truth would render a theory possessing a sign of truth not merely likely, but true. Rather, the sign is \textit{reliable} in the sense that it renders what is signified (truth) likely. Repeated use of the sign would lead to truth more often then not.

4.3 The Problem of Beautiful Rival Theories

The aesthete faces an uphill battle to show that beauty is a sign\(_1\) of truth. Finding signs of truth are notoriously difficult, even in the non-aesthetic realm. John Earman has presented what might be called the \textit{problem of rival theories} (Earman, [1985] 1994). He shows that if we assume that there exists a rival theory \(T^*\) that has at least as high prior probability as \(T\), \(p(T^*/B) \geq p(T/B)\), then no matter how many more confirming observations we make (in the sense that both \(T\) and \(T^*\) entail \(O_1 \& O_2, \ldots, O_n\)), then \(p(T/ O_1 \& O_2, \ldots, O_n \& B) \leq 0.5\) (Where \(B\) stands for “background knowledge,” and \(E\) stands for “evidence.”). Thus, we cannot infer a likely theory from the observable phenomena unless we “load the dice” against rival theories, to use Earman’s phrase. Actually, the situation is
even worse. One can generalize Earman’s result: if we assume in addition to $T$, $m-1$ mutually incompatible theories that exhaust all the remaining possibilities, and if these theories also save all the phenomena, then $p(T/ O_1 \& O_2, \ldots, O_n \& B) \leq 1/m$.\footnote{169 Proof: let a theory $T$, which in conjunction with background knowledge $B$ entail $O_1$, $O_2$, $\ldots$, $O_n$. Claim: Assume that $p(T/B) < 1/m$. If (i) $T_1, \ldots, T_{m-1}$ also entail $O_1, \ldots, O_n$, and (ii) $B$ entails $\neg (T_i \& T_j)$ where $i \neq j$, and (iii) $\forall i (p(T_i/B) \geq p(T/B))$ and (iv) $p(T_1 \lor T_2 \lor \ldots \lor T_{m-1}/B) = 1$, then for any $n$, no matter how large $p(T/ O_1 \& O_2, \ldots \& O_n \& B) \leq 1/m$. From Bayes’ theorem and (i) one can derive the following equality
\[
p(T_1 \lor T_2 \lor \ldots \lor T_{m-1}/O_1 \& \ldots \& O_n \& B) = p(T_1 \lor T_2 \lor \ldots \lor T_{m-1}/B)
\]
Now we know $p(T/B) < 1/m$, by assumption, and by (iv) $p(T_1 \lor T_2 \lor \ldots \lor T_{m-1}/B) \geq (m-1)/m$. Assume that the claim is false, i.e., $p(T_1 \lor O_2, \ldots \lor O_n \& B) > 1/m$, then $p(T_1 \lor T_2 \lor \ldots \lor T_{m-1}/O_1 \& O_2, \ldots \& O_n \& B) < (m-1)/m$, and the above equality cannot hold.}

Are the dice loaded against ugly theories? It is a popular proposal to claim that if non-empirical factors play a role in theory confirmation, they influence the prior probability. Wesley Salmon, for example, claims that simplicity and symmetry have a “significant bearing” on the prior probability of a theory (Salmon [1990] 1996, p. 283). Let us turn to whether beauty can influence the prior probability of a theory. A proponent of the idea that beauty is a guide to truth might suggest that an empirically adequate theory with the most beauty has a higher probability than its less beautiful rivals, for all $T_i \neq T_1$, $p(T_1/B) > p(T_i/B)$. For example, if the Caspar-Klug theory has the most beauty of any possible empirically adequate theory of virus structure and consequently has a higher probability than its rivals, then Earman’s theorem is not applicable and it is possible for $p(T/E&B)$ to converge to a number greater than 0.5. Does this suggestion work? Unfortunately, even if we accept the dubious idea that a theory with the most beauty has a higher prior probability than less beautiful rivals, one can construct a variant of Earman’s argument that I call the problem of beautiful rival theories. There is no reason to think that a theory with the most beauty will be unique. There could be multiple theories with the maximal amount of beauty. If there is a rival theory $T^*$ that also saves the phenomena and is equally beautiful, then by analogous argument, $p(T_1/B) = p(T^*/B)$ and $p(T_1/B) \leq 0.5$, since $T_1$ and $T^*$ are incompatible.
Is there any way to exclude the possibility of an equally beautiful equally empirically adequate theory of virus structure? Can the aesthete show that there are some features of theoretical beauty that rule out the possibility of genuine alternatives? The aesthete could attempt to improve her argument. One obvious approach is to modify some of the premises. Let us consider symmetry and coherence separately, beginning with symmetry. First modify the fourth premise to (4’) \((\forall T)p(T/T \text{ is empirically adequate & } T \text{ is the most beautiful}) > 0.5\). Of course, the definite article will not carry the weight needed to guarantee that a unique theory has the highest degree of beauty so further modifications are needed. Second, modify the second premise to (2’) “A theory that postulates entities with the highest degree of symmetry and lacks ugly properties is the most beautiful.” The aesthete’s hope is that she can make use of features of symmetry to guarantee a unique most beautiful theory. For example, the reasoning might go something like this: order all possible theories of virus structure according to the degree of symmetry they postulate. Take this ordering also to order possible theories according to their degree of beauty. The theory at the head of the ordered list will be the most beautiful theory, or so the aesthete might argue. Is it possible to order theories of virus structure according to the postulated degree of symmetry? Perhaps possible theories of virus structure could be classified by the point group symmetry of the postulated virus structure (bracketing quasi-equivalence). In some cases, judgment of relative symmetry of point groups is fairly straightforward. When point group B is a subgroup of point group A, then A has the symmetry elements of B and more besides. For example, a cube (point group 432) has all the symmetry of a tetrahedron (point group 32) and also possesses 4-fold axes that a tetrahedron does not. Thus, a cube is more symmetrical than a tetrahedron. However, it is not always straightforward to determine which point group has the most symmetry. To continue with the example, which is more symmetrical an icosahedron or a cube? Both have the symmetry of a tetrahedron, but a cube has three additional 4-fold axes and an icosahedron has six additional 5-fold axes. To say that the icosahedron is more symmetrical, one must assume that six 5-fold
axes have more symmetry than three 4-fold axes. Thus, to completely order the point
groups in order of degree of symmetry, one must adopt a convention to allow comparison
of different types of symmetry elements. Similar considerations apply to more abstract
symmetries. For the sake of discussion, I will make the natural assumption that it is simply
the total number of symmetry elements that determines the degree of symmetry. With this
measure of the degree of symmetry, the icosahedral symmetry (point group 532) with 60
symmetry elements is the most symmetrical. If we make the assumption that a more
symmetrical point group is more beautiful, then we can order the point groups into a scheme
of relative beauty and hopefully exclude any cases of equal degrees of beauty.

There are many problems with this line of argument. I have already touched on one
problem: the assumption that more symmetry means more beauty. Is spherical symmetry
more beautiful than icosahedral symmetry? Essentially this is a question of comparing the
discrete with the continuous and is like comparing apples and oranges. A second problem
is that there can be alternative sources of beauty other than symmetry. A theory may be
beautiful because it posits few axioms or because it elegantly derives a surprising result.
These alternative sources of beauty allow for further ways in which two theories might be
judged to be equally beautiful or even the possibility that the most beautiful theory may not
posit beautiful symmetry. Third, even if we are able to order the point groups to produce a
unique most beautiful point group, we have not necessarily ordered possible theories of
virus structure. Say we judge icosahedral symmetry to be the most beautiful, it is possible
that there are multiple incompatible theories that posit icosahedral particles. These
problems suggest that a focus on degree of symmetry does not circumvent the problem of
rival beautiful theories.

Can we use some features of coherence to circumvent the problem of rival theories?
Again many of the strategies open to the aesthete mirror those employed using symmetry
and are prone to analogous problems. One approach is to argue that a more coherent theory
should have a higher probability than a less coherent theory, ceteris paribus. If we could
enumerate all possible theories of a given phenomena and the most coherent theory had a higher probability than the others, then Earman’s theorem is no applicable and the most coherent theory’s probability can converge to a number greater than 0.5. Unfortunately, even if we could enumerate all the theories, it is implausible to give the most coherent theory a higher probability. First, the most coherent theory might not be the most beautiful once we take into account additional aesthetic factors. The most coherent theory might possess highly asymmetrical posits, for example. This objection is blunted somewhat by the ceteris paribus clause, but it is still possible that the relation between the higher degree of coherence and other aesthetic factors leads to a less than beautiful theory or a theory that is not the most beautiful of the range of possibilities. Alternatively all the possible theories might be ugly. This reply will not convince the Whewellian who holds that coherence is the only epistemically relevant factor and there is at least one coherent theory. On his view, an ordered list of theories in terms of their degree of coherence and another ordered list of the same theories in terms of their beauty will be collinear.

A more effective reply to the Whewellian attacks the view that more coherent theories are more probable. In other words, attack what I have called the aesthete’s narrow strategy of linking an aesthetic property, which in this case is coherence, to truth. One might think that we could look to the development of science since Whewell’s time and argue that since his best examples of coherent theories have been refuted, his approach was flawed. We have had over a century of science to see how well the Whewellian model has worked. Newtonian physics, Whewell’s exemplary case, has been shown to break down at speeds close to the speed of light, for example. Prima facie, the failure of Newtonian physics would appear to undercut Whewell’s approach. However, this historical approach has limited force. Presumably the Whewellian would argue that the reason that Newtonian science has been rejected is because a more consilient theory was found. Einsteinian mechanics applies to more phenomena than Newtonian dynamics. Thus, the history of science cannot be used in this simple way to refute the Whewellian principle. What such
cases show is merely that the inference from “T is coherent until now” to “T will continue to be coherent” is fallible.

It is more fruitful to attack Whewell on logical grounds. Consider then a situation in which we have a coherent theory $T_c$ that entails the two types of phenomena $O_1, \ldots O_n$ and $O_{n+1}, \ldots O_{n+m}$ and two other theories such that $T_2$ entails $O_1, \ldots O_n$ and $T_3$ entails $O_{n+1}, \ldots O_{n+m}$. Further let us assume that $(T_c \land T_2)$ and $(T_c \land T_3)$. There are two cases: (1) where the competing less coherent theories are incompatible, i.e., $(T_2 \land T_3)$, and (2) where they are not. What reason is there to think that $p(T_c) > p(T_2)$, $p(T_c) > p(T_3)$ or even $p(T_c) > p(T_2 \land T_3)$? Clearly on the first case, if $T_2$ and $T_3$ are incompatible, then $p(T_2 \land T_3) = 0$, so we would be irrational to believe this conjunction. Let us turn to case (2) and return to the question of incompatible theories. If using some principle of indifference we assigned equal probabilities, then $p(T_c) = p(T_2) = p(T_3)$ and $p(T_c) > p(T_2 \land T_3)$. Thus, it would appear that if we had to choose between two competing explanations of a set of phenomena $(O_1, \ldots O_{n+m})$, then it is more rational to believe the coherent theory ($T_c$). However, resting one’s argument on a principle of indifference, always a little shaky, is especially so when there are differences in scope. A more reasonable assignment is to say that a theory’s prior probability should be proportional to the inverse of its scope. Karl Popper shares this intuition when he says that a less falsifiable theory is more probable (Popper 1959, pp. 212-5). For Popper, a theory’s degree of falsifiability is proportional to its scope, ceteris paribus. Unfortunately for the aesthete, if prior probability is inversely proportional to scope then we have no reason to assign $T_c$ with more probability than the conjunction of $T_2$ and $T_3$, since by stipulation the two competing theoretical structures have the same scope. In general nothing about the nature of coherence requires that coherent theories have higher probabilities than (consistent) incoherent alternatives. Indeed why should we think that the world is more likely described by coherent theories? One could argue that the history of science is littered with beautiful theory after beautiful theory, each slain by an ugly fact. Indeed, one could just as well argue that history of science provides some reason to think
that the prior probability of a coherent theory is lower than less coherent alternatives. Nonetheless, prior probability offers the best hope for the aesthete and I will return to whether beauty can influence prior probability in section 4.4.2 and also in section 5.2 where I will reconsider the question of whether coherence bears upon prior probability when I consider Achinstein’s probabilistic rendition of coherence, which although is richer than the notion described here, suffers from similar problems.

4.4 The Aesthete’s Short-Run Rejoinder

At this point the Aesthete might shift gears. Perhaps she concedes that Earman’s theorem suggests that in the long run the beauty of a theory cannot guarantee that its probability will not converge to a number above 0.5. However, echoing John Maynard Keynes, the Aesthete might suggest that what happens in the long run is irrelevant to what one should believe now, given current evidence. It still could be the case that current evidence is such that it is now rational to believe that beautiful theories are likely. It is to this question that I now turn. Earman’s argument makes use of Bayes’s Theorem. Like Earman, one does not have to be card-carrying Bayesian to use Bayes’s Theorem to analyze how beauty might affect a given theory’s probability. Indeed, let us use Bayes’s Theorem to consider the three components of a theory’s probability. In its simplest form Bayes’s Theorem says:

\[ p(T/E) = \frac{p(T) \times p(E/T)}{p(E)} = \text{prior probability} \times \text{likelihood} / \text{expectedness} \]

E stands for evidence. A more complicated version of the theorem includes a term for background information B. To consider the role of the beauty of T, consider background knowledge B such that B entails “T is beautiful.”

\[ p(T/E&B) = \frac{p(T/B) \times p(E/T&B)}{p(E/B)} \]

The probability of a theory depends on three things: the prior probability, the probability of the evidence given the theory, i.e., the likelihood, and the probability of the evidence.
The denominator, the probability of the evidence, \( p(E/B) \), sometimes called the expectedness of the evidence, depends upon \( E \) and \( B \) and is the same for all theories competing to account for the same phenomena whether they possess beautiful attributes or not. If the beauty of a theory has any affect upon the theory’s plausibility, it must either affect the prior probability or the likelihood.

4.4.1 Does Beauty Influence the Likelihood of a Theory?

What then of likelihood, \( p(E/T&B) \)? This expression is the probability of the evidence given background knowledge and that the theory is true. Often this number is taken as one measure of the explanatory power of the theory. In other words, on the assumption that the theory is true, the evidence becomes likely and is no longer surprising. The problem with linking beauty to likelihood is that we can conceive of very ugly theories having a high likelihood. An ugly theory may even entail the evidence \( E \), in which case the likelihood \( p(E/T&B) = 1 \). Indeed, if we frame the problem consulting the theoretician to be a choice among empirically adequate theories as we did at the beginning of the chapter, then the likelihood is set to 1 for all candidates, beautiful or not. Even if we relax this assumption there are additional problems for the Aesthete. Perhaps the Aesthete believes that a theory possessing a beautiful property has a higher likelihood than one that does not. This belief is implausible. Why should the beauty of a feature of a theory make any experimental results less surprising? I see no compelling reason. Perhaps one could define a beautiful theory to be one that made the experimental results probable. This avenue leads to the difficulty that ugly theories can also have high likelihoods. The “theory” that says an evil demon created the world such that \( E \) is not beautiful, but has a high likelihood, indeed the likelihood is 1. To say that this is a beautiful theory just because it has a high likelihood, exceeds the mandate of even a revisionary theory of beauty. Additionally the aesthete must argue that the differences in likelihoods between a beautiful theory and its less
beautiful rivals must be great enough to render the theory likely. A further problem with this avenue is that beauty would then be relative to what experimental results were in question. Since a theory might make results \( E_1 \) probable and others \( E_2 \) improbable, a theory would be beautiful relative to \( E_1 \) and not beautiful relative to \( E_2 \). Peter Lipton’s notion of “Loveliness” has this relational feature (Lipton 1991; Lipton 2002, p. 18). A theory is lovely relative to an explanandum, according to Lipton, only if it provides a lovely explanation of that explanandum. Beauty, it seems to me, is not relational in this manner. Beauty is like truth. A theory is true or false, not true or false relative to some evidence. Likewise, a theory is beautiful or not, period. Sometimes it is said that beauty is relative to the observer, i.e., “beauty is in the eye of the beholder”, but even this sort of relativism is not of the right type to make beautiful theories have high likelihoods. Furthermore, to argue that beauty is relative and truth is not makes the aesthete’s position less defensible not more. Since I am attempting to be charitable to the aesthete’s position, I will continue to assume that beauty is not relativized to a set of evidence.

4.4.2 Does Beauty Influence the Prior Probability of a Theory?

This leaves the prior probability of a theory as the most relevant place to examine whether beauty or an alleged sign of truth does render a theory likely. Given their name, one might think that prior probabilities are determined a priori. Let me distinguish between two types of priors: (1) empirical priors and (2) a priori priors. A priori priors, symbolized \( p(T) \), are not conditional on any contingent background knowledge. Because of the lack of empirical constraint, some with good reason argue that a priori priors are often undefined. The aesthete could argue that beauty provides the needed constraint. As we shall see, I am suspicious of this argument. Empirical prior probabilities, symbolized \( p(T/B) \), are typically conditional upon the background knowledge \( B \) and consequently will depend upon the nature of the background knowledge. As the saying goes, today’s priors are yesterday’s posteriors. Given the empirical nature of these prior probabilities, it is difficult to assign
values from the philosopher’s armchair. Here I think the burden is on the Aesthete to show that there are features of the background knowledge that would determine that the prior probabilities of beautiful theories are high enough to guarantee that the posterior probabilities are greater than 0.5. It is difficult to find anyone defending this radical position. Peter Lipton comes close. He defends the idea that “loveliness is a guide to likeliness.” (Lipton 2002) He does not mean by “loveliness” exactly what I have been calling “beauty,” but rather something like explanatory power. Nonetheless he claims that his notion of loveliness aims to make sense of the “broadly aesthetic considerations” that allegedly guide inference and in this regard his claims are relevant here (Lipton 1991, p. 68). Unfortunately, Lipton does not give good arguments as to why more lovely theories should be given higher prior probabilities, let alone why the loveliest theory should be given a probability greater than 0.5. He does claim that prior probabilities are “generated in part with the help of explanatory considerations,” but provides little argument for it (Lipton 2002, p. 24). He says for example, “Given our inferential methods are successful, why should they depend upon explanatory considerations? Well they have to depend upon something, so why not explanation [i.e., loveliness]?” (Lipton 1991, p. 128) To recast his point in my language: given the under-determination of theory by the evidence, we are forced to depend upon a criterion to weed out the skeptical possibilities. Why not use beauty? As I hope is apparent, Lipton and this approach conflates a necessary condition for a sufficient one. He claims that we should adopt explanatory considerations to infer probabilities. Even if we concede the dubious point that explanatory or aesthetic considerations allow us weed out allegedly less probable candidate theories (the ugly or unexplanatory ones), it does not follow that beauty is a necessary or even a justified foundation for inference. Before he can establish this claim he needs to consider competing foundations for inference and show that explanatory considerations have advantages over completing considerations. Indeed, in the end, Lipton, at best, gives us a description of our inferential practices, not a justification of them.
Let us make the Aesthete’s position more precise. Suppose that $T_1, \ldots, T_n$ are all the known theories for a set of phenomena and suppose that $T_1$ is beautiful and the remaining theories are ugly. The Aesthete might claim that $p(T_1) > 0.5$. Furthermore if $T_1$ entails $E$, then it follows from the Aesthete’s claim that $p(T_1/E) > 0.5$. To make the aesthete’s position as cogent as possible let us make a “disregarding” assumption that as yet unknown theories are not being considered in determining the probability of known theories (Achinstein 2001, p. 107-8). Under this disregarding assumption the aesthete’s claims are that $p_{(d(u)\cap T_1)} > 0.5$ and $p_{(d(u)\cap T_1/E)} > 0.5$. One problem with Achinstein’s disregarding condition is that it threatens to disconnect probability assignments from belief. For example, the probability of an unknown theory, given the disregarding condition, is either undefined or 0, i.e., $p_{(d(u)\cap T_u)}$ is undefined or $p_{(d(u)\cap T_u)} = 0$. However, one would think that it would be rational to make a bet at some odds that an unknown theory might be true. If you would be willing to make the bet at the odds of $q: 1-q$, it suggests that you actually believe that an unknown theory has a probability of $q$ being true where $q > 0$. If the value of $q$ is significant, then use of the disregarding condition interferes with the relation between probability and belief. On my view, whether the disregarding condition is justified is an empirical question. For example, say that there is only one known theory $T_a$ of a given phenomenon and it is empirically adequate. Given that probability sums to one it follows that the probability of $T_a$ disregarding unknown theories is 1, i.e., $p_{(d(u)\cap T_a)} = 1$.\textsuperscript{170} But whether one should believe $T_a$ (and whether it is justified in assuming the disregarding condition) depends on other facts such as whether there has been a search for competing theories or not. If the phenomena that $T_a$ purports to explain have very recently been discovered and $T_a$ is the first of an anticipated number of theories that will attempt to explain

\textsuperscript{170} This inference assumes that $\neg T$ is an unknown theory. Achinstein might reply that $\neg T$ is a known theory even if it is composed of a disjunction of unknown theories which each have undefined probability values. This reply saves his position but at the cost of significant contortion.
the phenomena, we might be reluctant to adopt the disregarding condition. The aesthete then needs to justify the disregarding condition before we adopt it.

Nevertheless, the underlying intuition behind the aesthete’s assignment of probability is that a beautiful theory has a higher probability than an ugly theory. I have serious reservations about this intuition. Consider the observable consequences of $T_1, O_1, \ldots, O_m$. One can construct a new theory $T_o$ which consists in the conjunction of all of these observable consequences, $O_1 \& O_2, \& \ldots, O_m$. It is highly likely that this conjunction $T_o$ is not beautiful, or at the very least, less beautiful than $T_1$, since theories consisting of mere conjunctions of facts are almost never beautiful. If less beautiful theories have a lower probability, then $p(T_o) < p(T_1)$. However, since $T_1$ entails $T_o$, it follows from an axiom of the probability calculus that $p(T_o) \geq p(T_1)$ and $p(T_o/E) \geq p(T_1/E)$. While this does not refute the Aesthete’s position, I think it begins to undermine the intuition that lies behind it.

Let us consider some possible replies that the aesthete could make to this argument. The aesthete might respond that the conjunction of observable consequences, $O_1 \& O_2 \ldots, \& O_m$, is not a theory as it will consist in an infinite number of conjuncts and infinite conjunction is not defined. Whether the conjunction of the observable consequences is infinite or not depends upon how we describe the consequences. The observation “Inoculating tobacco leaves with tobacco mosaic virus (TMV) causes a mottling pattern” also consists of the conjunction “The first time tobacco leaves were inoculated caused a mottling pattern” and “The second time tobacco leaves were inoculated caused a mottling pattern” and “The third time tobacco leaves were inoculated caused a mottling pattern” etc. Thus depending upon how the observable consequences of a theory are described they may or may not be infinite number of them. Furthermore, even if there were an infinite number of observable consequences, we could consider a less beautiful but more likely “theory” constituted by the finite conjunction of observed consequences of the theory and avoid the aesthete’s rebuttal. However, the aesthete might argue that a finite or infinite set of observable consequences of a theory is not a theory since theories require some degree of
unity. Notice that being a conjunction alone does bar a hypothesis from being a theory. Newton’s laws when conjoined form the core of Newtonian theory, an exemplary theory if ever there was one.

Nonetheless, for the sake of argument, let us grant the aesthete that a conjunction of observable consequences of a theory is not a theory proper and that the use of beauty in ranking prior probabilities is legitimate only in ranking genuine theories. The aesthete still has to show how any theory that is entailed by another is of equal or lesser beauty, no easy task. To begin this task she will have to exclude tautologies, which are entailed by any theory, but clearly are not the most beautiful theories! But let us suppose that the aesthete can achieve his task or at least restricts the set of theories he wishes to consider to include only theories that have no deductive relations to any other member in the set or deductive relations that respect the principle that a more beautiful theory cannot entail a less beautiful one. As Kruse (2000) points out such a move is appears ad hoc. Is the procedure of assigning more beautiful theories (or more coherent theories) higher prior probabilities rational? It is not clear that there are any standards of rationality that apply here. The normal ways of constraining rational belief, such as Dutch book arguments do not apply, as they only require one to obey the probability calculus. Any assignment of prior probability, as long as the probabilities of mutually exclusive theories sum to one or less, is allowed. As the Bayesians Howson and Urbach put it:

There is nothing in logic of the probability calculus which precludes the assignment of even probability 1 to any statement, however strong, as long as it is not a contradiction. The only other way in which probabilities depend upon logic is in their decreasing monotonically from entailed to entailing statements. But this again does not preclude anybody from assigning any consistent statement as large a probability as they wish. (Howson and Urbach 1989, p. 259)

As we have granted that the aesthete’s strategy is rational in the sense that it violates neither deductive logic nor the probability calculus. Of course there is no violation because there is
very few standards to violate. In fact, the lack of rational principles to guide the assignment of prior probabilities is often seen as a weakness of a Bayesian account of conformation. Furthermore, there is no good sense in which the aesthete’s strategy of assigning prior probability is more rational than an innumerable number of alternative assignments. To see this point, consider a competing strategy, that of the anti-aesthete. The anti-aesthete thinks that more ugly theories should have higher prior probabilities than less ugly. He repudiates every relative assignment of prior probability that the aesthete makes. If \( T_1 \) is more beautiful than \( T_2 \), the aesthete asserts that \( p(T_1) > p(T_2) \), whereas the anti-aesthete asserts that \( p(T_2) > p(T_1) \). I claim that the anti-aesthete’s strategy is just as rational as the aesthete’s with respect to the a priori prior probability assignments. There is no a priori reason to privilege one approach over the other. The ardent aesthete disagrees, but on what basis can she argue for a difference between the two symmetrical positions. It cannot be from experience, since we are considering a priori priors. This leaves some non-empirical means. For the aesthete to be convincing she must provide an a priori proof of her principle, but no proof is forthcoming. To bias our a priori probabilities in favor of beauty is at best a metaphysical article of faith, which either unjustifiably biases our beliefs in favor of beauty or divorces probability assignments from belief. That we must privilege the a priori priors of beautiful theories is not a methodological necessity either. Science progresses as efficiently, or perhaps even more efficiently without a bias toward beauty. If I am correct, this leaves the aesthetes defense of the epistemic role of beauty in a precarious position and no more rational than competing approaches. Whether the aesthete strategy works for empirical prior probabilities is an open question which I will address in the next chapter when I consider James McAllister’s account, which although is inductivist rather than explanationist, shares some features with Lipton’s approach.
4.5 Beauty and Belief

Let us finally consider premise (6): “It is rational to believe a likely theory.” I previously defined “likely” to mean a probability greater than 0.5. The numerical value must be at least 0.5, otherwise it may be rational to believe $T$ while also rational to believe $\neg T$ given the same evidence. Perhaps surprisingly, premise (6) is a contentious claim. For example, subjective Bayesians reject the notion of ‘belief’ in favor of ‘degree of belief’. If a proposition has probability $r > 0.5$, a subjective Bayesian would say that we should believe it with degree $r$. The significance of the threshold of 0.5, for the Bayesian, is merely that the degree of belief for the proposition is higher than for its negation. The aesthete might be willing to say that this Bayesian rendering of the conclusion is a significant enough finding however. Indeed if the beauty of a theory was powerful enough to render a one’s degree of belief, to use the Bayesian parlance, in the theory greater than the theory’s negation the aesthete should declare a victory.

The aesthete’s principal problem is that she has not shown that beauty renders the Caspar-Klug theory—or any other beautiful theory for that matter—likely. The back of the aesthete’s argument is broken at premise (4), which claimed that beauty was a sign of truth. I have argued that in the long run there a serious difficulties with thinking that a beautiful theory’s probability will converge to a number greater than 0.5. Furthermore, the situation is no better for the aesthete if we restrict our attention to the short run. However, perhaps there are less radical, more defensible, weaker positions where the aesthete might find refuge. While beauty might not render a theory likely, perhaps it increases its probability. While justifiable belief in a beautiful theory may not accompany an increase in probability, perhaps the aesthete can argue that a theory’s beauty warrants stronger belief in the theory or weaker belief in its negation. I examine this question in the next chapter.
Chapter 5  Does Beauty Increase a Theory’s Probability?

Structure:

5.1 Beauty as a Sign₂ of Truth
   5.1.1 Four Ways Beauty could be a Sign₂ of Truth

5.2 William Whewell’s Notion of Coherence: An A Priori Sign₂ of Truth?
   5.2.1 Achinstein’s Rendition of Whewellian Coherence
   5.2.2 A Priori Signs of Truth: The Dilemma

5.3 Is Beauty an Empirical Sign₂ of Truth?

5.4 James McAllister’s Defense of Beauty as a Sign₂ of Truth
   5.4.1 Consequences of McAllister’s Model
   5.4.2 General Objections to McAllister’s Model of Aesthetics in Science
   5.4.3 Comparing McAllister’s Model with the Case Study

5.5 Is the View that Beauty is a Sign₂ of Truth Tenable?

5.6 Is the Argument against an epistemic role for Beauty Too Good?

5.7 Summary and Conclusion of Chapter 5

5.1 Beauty as a Sign₂ of Truth

If the argument in Chapter 4 was successful, then the Aesthete cannot successfully defend the view that the beauty of a theory is a sign₁ of truth, a sign that renders a theory likely without appeal to empirical evidence, which philosophers are ill-suited to garner. At this point in the debate, the Aesthete might retreat to the weaker position that beauty increases the probability that a theory is true, but not necessarily to a level above a given threshold. In other words, the Aesthete may argue that beauty is a sign₂ of truth. In this chapter, I will argue that the Aesthete cannot successfully defend even this weaker position.

The claim that a theory’s beauty might increase its probability of being true invites the following question: increase its probability relative to what? An intuitive response is to
say the theory increases its own probability relative to what its probability would be if it did not possess beauty. It is this intuition that I drew from when in Chapter 1, when I defined $\text{sign}_2$ of truth this way:

$$\text{Beauty is a sign}_2 \text{ of truth if and only if, for any } T, p(T/T \text{ is beautiful}) > p(T).$$

As mentioned in the introduction, there is a difficulty with applying this definition: the property of being beautiful can be *essential* to the theory in question. The Caspar-Klug theory would not be the same theory if “it” did not postulate quasi-equivalence and consequently failed to be beautiful. Thus, it is sometimes impossible to compare a beautiful theory with a non-beautiful version of itself. A more general way of expressing the idea of a sign$_2$ of beauty is:

$$\text{Beauty is a sign}_2^* \text{ of truth if and only if, for any two theories } T_1 \text{ and } T_2, \text{ if } T_1 \text{ and } T_2 \text{ have all the same epistemic virtues in the same degree (except beauty), but } T_1 \text{ is beautiful and } T_2 \text{ is not, then } p(T_1) > p(T_2).$$

This definition introduces the notion of an epistemic virtue. The most basic virtue is that the theories must be internally consistent and not tautologous—that is, they have to have contingent empirical content. Additionally, they should “match” the observational data. How exactly to explicate the idea of the theory matching the data or being “empirically adequate” is controversial. A simple approach takes matching the data $E$ to mean that $T$ (conjoined perhaps with some auxiliary hypotheses) entails $E$. This view, although simplistic, has the virtue of clarity and I will assume this interpretation of empirical adequacy unless stated otherwise. Van Fraassen proposed an influential semantic conception of empirical adequacy (van Fraassen 1980, p. 45). For him a theory is empirically adequate when all relevant phenomena are isomorphic with a model of the theory. He means by model a structure that satisfies the axioms of the theory. For example, a model of a $T=3$ structure on the Caspar-Klug theory is isomorphic with a TYMV particle. In this case, the isomorphism consists of mappings between 60T model subunits and 180 TYMV protein molecules that conserve nearest neighbor relations. Some philosophers of
science add further epistemic virtues: explanatory power, falsifiability, extendibility etc. Whether these additional alleged epistemic virtues are unrelated to beauty and so avoid the parenthetical clause in the definition of sign2, is a complicated question that I will address in passing below.

5.1.1 Four Ways Beauty is not a Sign2 of Truth

Let us begin with a very quick argument against the idea that beauty is a sign2 of truth. Using Bayes theorem it follows that a theory’s beauty depends upon the probability of the theory. To see why consider Bayes theorem:

\[ p(T/T \text{ is beautiful}) = p(T) p(T \text{ is beautiful}/T) / p(T \text{ is beautiful}), \]
which rearranges as

\[ p(T/T \text{ is beautiful}) \] / p(T) = p(T is beautiful/T) / p(T is beautiful) ……… BT2

If beauty is a sign2 of truth, then \( p(T/T \text{ is beautiful}) > p(T) \) and the left hand side of the equation BT2 is greater than 1. For this equality to be true, the right hand side of equation BT2 has also to be greater than 1 and thus \( p(T \text{ is beautiful}/T) > p(T \text{ is beautiful}) \). In other words, if beauty is a sign2 of truth, then the probability that a theory is beautiful is increased if the theory is true. However, this result is counterintuitive. Surely whether a theory is true, false, probable, or improbable should be irrelevant to whether it is beautiful. Suppose we are attempting to determine if a theory is beautiful. The addition of the fact that the theory is likely, or unlikely for that matter, should have no bearing on our task. Watson, Jacob, and others could judge Crick and Watson’s DNA model to be beautiful before it was tested and Watson judged his 1955 TMV model “very very pretty” before testing demonstrated it was false. On the other hand, in the past, some thinkers have linked truth to an aesthetic property. The ontological argument for the existence of God supposes that the truth of “God exists” increases God’s perfection, for example. The problems with the ontological argument notwithstanding, there are significant differences between the beauty of God and the beauty of a theory. Nonetheless, the Aesthete would say that this argument begs the
question, since what is at issue is whether a probabilistic connection between beauty and truth exists.

Consider two competing mutually exclusive and empirically adequate (in the sense that T entails E) theories T₁ and T₂ and background knowledge B. Suppose they are comparable on all epistemic respects except that T₁ is beautiful and T₂ is not. If beauty is a sign of truth, then \( p(T₁/E&B) > p(T₂/E&B) \). If one can show that the inequality does not hold then by modus tollens, beauty is not a sign of truth. One advantage of this comparative approach is that we no longer need to consider the expectedness of the evidence \( p(E/B) \). From Bayes theorem it follows that

\[
p(T₁/B) \frac{p(E/T₁&B)/p(E/B)}{p(E/T₂&B)/p(E/B)} > p(T₂/B)
\]

Thus it is the relative value of the ratios of priors and likelihoods of the two theories that determine the truth of the Aesthete’s claim. Let us consider four ways in which this inequality could be true.

1. Driven by the priors: possession of a beauty increases the prior probability of T₁ over T₂ to the degree that differences in the likelihood are irrelevant.
2. Driven by the likelihoods: possession of a beauty increases the likelihood of T₁ over T₂ to the degree that differences in the priors are irrelevant.
3. Driven by the priors and likelihoods: the beauty of the theory increases both the prior and likelihoods. Possession of a beauty increases both the priors and the likelihoods.
4. Beauty has no systematic effect on the priors or likelihoods individually, but nonetheless

\[
p(T₁/B) \frac{p(E/T₁&B)}{p(E/T₂&B)} > p(T₂/B) \frac{p(E/T₂&B)}{p(E/T₁&B)}.
\]

I will deal with the four cases in reverse order ending with the most plausible case.

The fourth case is the least likely to occur. I mention it principally for the sake of completeness. For it to obtain, when the prior probability T₁ is less than T₂, the likelihood
of $T_1$ must be greater than $T_2$, by a greater amount than $T_1$ is less than $T_2$. Furthermore, where the prior probability of $T_1$ is greater than $T_2$, the likelihood of $T_1$ cannot be even greater than the likelihood of $T_2$. That this is the case is implausible and I can think of no mechanism that would guarantee these relations exist. The mechanism would have to increase a beautiful theory’s priors when it decreases the likelihoods and vice versa. I know of no one who defends this position.

Arguments against case 3 will be similar to arguments against cases 1 and 2. If any of my arguments against case 1 and 2 are successful, they will also be successful arguments against case 3.

Case 2 concerns signs of truth that influence the posterior probability of a theory by increasing the likelihood of the theory, $p(E|T_1&B)$, over a rival $p(E|T_2&B)$ without the alleged sign of truth. Similar remarks apply to consideration of likelihoods as those mentioned in the previous chapter. First, if we are considering empirically adequate theories, $T_1$ and $T_2$, then the likelihoods are set at 1 and consequently they are equal for the theory with the alleged sign of truth and the theory without. Second, if we have a conception of beauty that necessarily involves making the evidence less surprising, then the notion of beauty risks becoming relativized to a set of evidence. Theories can be beautiful without accounting for the evidence and judgments of a theory’s beauty do not depend upon which evidence one considers, at least it seems to me. Plato’s theory of the four elements and Kepler’s theory regarding the spacing of the planets remain beautiful even though we have discovered that there are more than 100 elements and more than 6 planets. Third, consider the ugly theory that consists in the conjunction of the observable consequences of a theory. This theory will have just as high likelihood as the complete beautiful theory from which it is deduced. Even if we bar a conjunction of observable consequences from being a theory, it is possible to modify some theories to make them less beautiful without decreasing their likelihoods. That beauty increases the likelihood of a theory is by no means obvious. I see no reason to assume that beauty systematically
increases the likelihoods of a theory. On the other hand, the philosopher of science Wayne Myrvold correctly argues that if we have two theories \( T_u \) and \( T^* \) and evidence \( \{e_1, e_2\} \) it is possible that \( p(e_1 / T_u) = p(e_1 / T^*) \) and \( p(e_2 / T_u) = p(e_2 / T^*) \) and yet \( p(e_1 \& e_2 / T_u) > p(e_1 \& e_2 / T^*) \). In this case, Myrvold says that \( T_u \) unifies the evidence by making one part of evidence probabilistically relevant to another, i.e., \( p(e_1/e_2 \& T_u) > p(e_1/e_2 \& T^*) \) (Myrvold 2003). In fact he defines a measure of informational relevance: 


\[
I(e_1 e_2 / T_u) = \log_2 \left[ \frac{p(e_1/e_2 \& T_u)}{p(e_1/T_u)} \right]
\]

that is positive when \( T_u \) unifies \( e_1 \) and \( e_2 \). However, despite his information-theoretic sophistication, Myrvold cannot escape the fact that if \( T^* \) entails \( e_1 \& e_2 \), then the likelihood \( p(e_1 \& e_2 / T^*) = 1 \geq p(e_1 \& e_2 / T_u) \). An awkward consequence of Myrvold’s view is that if a theory entails the evidence it cannot unify the evidence, i.e., \( \log_2 [p(e_1/e_2 \& T_u)/p(e_1/T_u)] = \log_2 [1/1] = 0 \). He thus restricts his discussions to “bare-bones” theories that do not entail the relevant evidence. This restriction severely limits the utility of his approach. It also renders his account useless when choosing between two empirically adequate theories. (See Lange 2004, for additional criticisms of Myrvold.)

Case 1 is the most interesting of the four cases. As mentioned in the previous chapter, the most popular place where beauty is supposed to have an effect on the probability of a theory is by influencing the prior probability of the theory. Does a beautiful theory have a higher prior probability than a non-beautiful one? I began to address this issue in the last chapter, but let us take another bite at the apple. If so, this fact is either a priori or empirical. Let us spend some time on each sub-case in turn.

An argument for beauty being an a priori sign of truth can be found in theology. That the world is beautiful is alleged to follow from the nature of God.\(^{171}\) For example, Richard Swinburne claims, “if there is a God there is more reason to expect a basically beautiful world than a basically ugly one...” (Swinburne 1991, p. 150). If we assume that a

\(^{171}\) For a related discussion of how (allegedly) science and theology are converging on beauty, see Dubay (1999).
beautiful world requires a probable theory of that world also to be beautiful, then a beautiful theory, i.e., a theory with an alleged sign of truth, would be more probable than a non-beautiful theory. Perhaps this motivation lies behind Kepler’s theorizing mentioned in Chapter 1. Paul Dirac admitted to Carl Hempel that the reason for his famous pronouncements about the epistemic role of beauty was “Because the world itself is basically beautiful – God made it so.” (Hempel 2000, p. 81) Of course, this approach depends on assuming the existence of God and the contentious claim that if the world is beautiful, then so a true theory of it would be also. If Swinburne had an a priori argument for the existence of a creating God, then this approach might be fruitful. However, all known a priori arguments for the existence of God have been long thought to be fallacious. Indeed, Swinburne uses his purported connection between God and a beautiful world to show the existence of God rather than the opposite direction I am considering. To hold that true theories must be beautiful because God made it so is a metaphysically laden article of faith and not a rationally compelling reason.

5.2 William Whewell’s Notion of Coherence: an A Priori Sign of Truth?

Surprisingly we do not have to venture into theology to find someone who argues that it is a priori that beautiful theories of a certain type have higher prior probabilities. As mentioned in previous chapters, a plausible interpretation of William Whewell’s position claims that it is a priori that coherent theories have a higher probability. As he puts it:

In [true theories] all the additional suppositions tend to simplicity and harmony; the new suppositions resolve themselves into the old ones or at least require only some easy modification of the hypothesis first assumed: the system becomes more coherent as it is further extended. … In false theories, the contrary is the case (Whewell 1847, p. 68).

In other words, if only true theories tend to coherence, the a coherent theory’s probability is 1. On one sanitized version of Whewell, given the phenomena, we can judge whether a theory or an “additional supposition” tends to simplicity and harmony without
experimental testing. Given the relevant phenomena, it can be known a priori whether it is coherent. As I discussed in the previous chapter, Whewell himself links coherence to his historical notion of consilience and thus claims that it is an empirical question whether a particular theory is coherent. Nonetheless, as Whewell’s language suggests, for him the general connection between coherence and truth is a priori.

Is it true that beautiful coherence is a sign of truth? I think that coherence is too weak to be a sign of truth. In the previous chapter, I considered an interpretation of coherence that focused on the relative widths of scope of rival theories. I argued that wide scope alone was not a good reason to privilege the prior probability of a theory. Here I will consider Peter Achinstein’s interpretation of Whewell’s notion of coherence that focuses on the mutual probabilistic dependences of parts of a theory. To keep the different interpretations clear, I will call this position Achinstein’s Whewell. Not surprisingly, I will argue that Achinstein’s Whewell also fails to show that beauty (in the form of coherence defined below) is a sign of truth. Achinstein draws a similar conclusion in his chapter.

5.2.1 Achinstein’s Rendition of Whewellian Coherence

Consider Achinstein’s probabilistic rendition of Whewell’s notion of coherence. Let T be the set of hypotheses \{h_1, ..., h_m\}. A hypothesis can be said to be coherent with respect to other hypotheses. A set of hypotheses, what I am calling a theory, can also be said to be coherent. On Achinstein’s definition:

\[ h_1 \text{ is coherent with } h_2, ..., h_m \text{ on } B, \text{ if and only if } p(h_1/h_2, ..., h_m & B) > k, \text{ and } p(h_1/h_2, ..., h_m & B) > p(h_1 / B). \]

Then Achinstein introduces the notion of coherence for a set of hypotheses:

A set of hypotheses \{i.e., a theory\} h_1, ..., h_2 is coherent, on B, if and only if each hypothesis is coherent with the other members of the set on B (Achinstein 1991, p. 130).

We can derive consequences from these definitions that cast doubt upon whether this probabilistic form of coherence is strong enough to be a sign of truth. Let us assume that
T is coherent. From the first part of the definition it follows that \( h_i \) is coherent with \( h_2, \ldots, h_m \) on \( B \). It follows by the above definition that \( p(h_i/h_2, \ldots, h_m & B) > p(h_i / B) \). Whewell claims that falsity cannot exhibit coherence. However, consider the theory \( T^* \) that consists of \( \{\neg h_1, \neg(h_2, \ldots, h_m)\} \). Is it also coherent? Part of the proof is straightforward: we can show that \( \neg h_1 \) is coherent with \( \neg(h_2, \ldots, h_m) \) on \( B \). Here is the proof:

\[
p(h_1/h_2, \ldots, h_m & B) > p(h_i / B) \quad \text{by assumption}
\]
\[
p(h_2, \ldots, & h_m / h_i & B) > p(h_2, \ldots, & h_m / B) \quad \text{by Bayes’s theorem}
\]
\[
1 - p(h_2, \ldots, & h_m / h_i & B) < 1 - p(h_2, \ldots, & h_m / B)
\]
\[
p(\neg(h_2, \ldots, & h_m) / h_i & B) < p(\neg(h_2, \ldots, & h_m) / B)
\]
\[
p(h_i/ \neg(h_2, \ldots, & h_m) & B) < p (h_i / B) \quad \text{by Bayes’s theorem}
\]
\[
1 - p(h_i/ \neg(h_2, \ldots, & h_m) & B) > 1 - p (h_i / B)
\]
\[
p(\neg h_i/ \neg(h_2, \ldots, & h_m) & B) > p (\neg h_i / B)
\]

To complete the proof of \( \neg h_1 \) is coherent with \( \neg(h_2, \ldots, & h_m) \) on \( B \), we need to show that \( p(\neg h_i/ \neg(h_2, \ldots, & h_m) & B) > k \). I cannot offer a proof of this claim, but using Bayes’ theorem again one can show that \( p(\neg h_i/ \neg(h_2, \ldots, & h_m) & B) > [p(\neg h_i/B) - p(\neg(h_2, \ldots, & h_m) /B)/p(\neg h_i/B)] \) follows from the coherence of \( T \). As \( k \to 1 \), then \([p(\neg h_i/B) - p(\neg(h_2, \ldots, & h_m) /B)/p(\neg h_i/B)] \to 1 \) and \( p(\neg h_i/ \neg(h_2, \ldots, & h_m) &B) \to 1 \). In other words, if we assume that \( T \) is coherent, then it is possible that \( \neg h_1 \) is coherent with \( \neg(h_2, \ldots, & h_m) \) on \( B \). Next we must consider the second part of Achinstein’s definition. One needs to show that \( \neg(h_2, \ldots, & h_m) \) is coherent with \( \neg h_1 \) on \( B \). Using an analogous argument it is possible that \( p(\neg(h_2, \ldots, & h_m)/\neg h_1) > k \). Finally, one must show that \( p(\neg(h_2, \ldots, & h_m)/ h_1 & B) > p(\neg(h_2, \ldots, h_m)/ B) \) and thus show that \( \neg(h_2, \ldots, & h_m) \) is coherent with \( \neg h_1 \) satisfying the second part of Achinstein’s definition. The last step follows deductively:

\[
p(\neg h_i/ \neg(h_2, \ldots, & h_m) & B) > p(\neg h_i/B)
\]
\[
p(\neg h_i/B)p(\neg(h_2, \ldots, & h_m)/ h_1 & B)p(\neg(h_2, \ldots, h_m/B) > p(\neg h_i / B) \text{ by Bayes theorem}
\]
\[
p(\neg(h_2, \ldots, & h_m)/ h_1 & B) > p(\neg(h_2, \ldots, h_m)/ B)
\]
What are we to take from this result? At the theory level, if T* is the set of hypotheses \{\neg h_1, \neg(h_2 \& \ldots \& h_m)\}, then if T is coherent, then with plausible assumptions T* also is. However, at least one of T and T* must be false. In particular, even if T is as coherent as possible, i.e., k = 1, then an incompatible competing theory is coherent also. On Achinstein’s rendition of coherence, Whewell is mistaken when he claims that false theories cannot be coherent. At the hypothesis level, there is the same problem. It is not inconsistent to say that \(h_1\) is coherent with \(h_2, \ldots \& h_m\) on B and also that \(\neg h\) is coherent with \(\neg(h_2, \ldots \& h_m)\) on B, but it is inconsistent to claim that theory \(h\) and theory \(\neg h\) each possess a sign of truth, since they cannot both increase in probability relative to the same background knowledge. Herein lies the problem with thinking that coherence, even beautiful coherence, is an a priori sign of truth.

A number of options for Achinstein’s Whewell present themselves now. He could argue that we should abandon Achinstein’s probabilistic rendition of the notion of coherence because it diverges significantly from what Whewell himself intended. Alternatively, he could accept that Whewell exaggerated when he claimed that false theories cannot be coherent and that we should retain the probabilistic formulation of coherence since it is much more precise than Whewell’s actual words. As the former option amounts to an outright defeat, I assume Achinstein’s Whewell would opt for the latter option.

Achinstein’s Whewell can respond thus: while coherence does not guarantee truth, it nonetheless makes it more likely. In other words, other things being equal, he might argue a coherent theory is more likely than a theory that is not coherent. If this were true then coherence would be a sign of truth and the Aesthete would still have a significant victory. Alas, even this weaker claim does not hold up or so I shall argue. Consider the following thought experiment. Ignoring background knowledge for the sake of exposition, let \(T = \{h_1, h_2, h_3\}\), \(p(h_1) = p(h_2) = p(h_3) = c < k < 1\), and \(p(T) = p(h_1 \& h_2 \& h_3) = c^3\). Let
us assume that \( h_1, h_2 \) and \( h_3 \) are statistically independent of each other and consequently \( T \) is not coherent. Consider another theory \( T^* = \{h_1^*, h_2^*, h_3^*\} \) where \( h_1^* = h_1 \& h_2^* = h_2 \& h_3^* = h_3 \& h_1 \). In this case \( p(h_1^*/h_2^* \& h_3^*) = p(h_2^*/h_1^* \& h_3^*) = p(h_3^*/h_1^* \& h_2^*) = 1 > k \). Therefore, \( T^* \) is coherent. Is \( p(T^*) > p(T) \)? No, the two theories actually express the same propositions—the only real difference is between them is that \( T^* \) is coherent and \( T \) is not. Coherence does not necessarily increase the probability of a theory.

This type of counterexample can be extended to more complicated cases in which each hypothesis is not entailed by the conjunction of the rest. Continuing with the example. Consider \( T^\bullet = \{h_1^\bullet, h_2^\bullet, h_3^\bullet\} \) where \( p(h_i^\bullet) < k \). Let firstly \( h_1^\bullet = [h_1 \& p(h_2) > k] \), secondly \( h_2^\bullet = [h_2 \& p(h_3) > k] \), and lastly \( h_3^\bullet = [h_3 \& p(h_1) > k] \). Now it follows that \( p(h_1^\bullet/h_2^\bullet \& h_3^\bullet) > k \), \( p(h_2^\bullet/h_1^\bullet \& h_3^\bullet) > k \), and \( p(h_3^\bullet/h_2^\bullet \& h_3^\bullet) > k \). The idea is that \( T^\bullet \) is similar to \( T \) but consists of mutually coherent hypotheses and so is coherent. Is the probability of \( T^\bullet \) necessarily greater than \( T \)? No, \( p(T^\bullet) = p(h_1^\bullet \& h_2^\bullet \& h_3^\bullet) = p(h_1 \& h_2 \& h_3 \& p(h_i) > k \& p(h_2) > k \& p(h_3) > k) = p(T) = p(h_1 \& h_2 \& h_3) \) since \( h_i \) entails \( p(h_i) > k \), where \( k < 1 \).

Thus in this case making a theory coherent does not increase a theories probability.

Achinstein’s Whewell cannot show that probabilistic coherence is a sign of truth. To generalize my strategy here: coherence appears to be a property of the way the theory is represented, not a property of the content of a theory and so can vary when we vary the way of representing the same facts. The probability of a theory depends on its content and not the way it is represented. This representation dependence undercuts the possibility of coherence being an an a priori sign of truth.

---

172 I thank Palle Yourgrau for pushing me to develop this point.
5.2.2 A Priori Signs of Truth: The Dilemma

In general, I propose that proponents of a priori signs of truth will face the dilemma illustrated by Swinburne and Achinstein’s Whewell. Alleged a priori signs of truth are either too weak or are laden with excess metaphysical baggage. To demonstrate this generalization conclusively would require a more general argument than I have given here. Nonetheless, the form of the reply to any proponent of an a priori sign of truth would mirror that given above. Not all is lost for the aesthete, however. A more plausible position is to argue that whether a sign of truth exists is an a posteriori question to be settled with empirical evidence.

5.3 Is Beauty an Empirical Sign of truth?

What evidence is there for the general claim beauty is a sign of truth or the more specific claim that in the case of theory Tn, beauty is a sign of truth? An obvious way to proceed would be to appeal to induction. Suppose we could list all the beautiful theories that have been proposed or are about to be proposed. From this list we randomly sample n-1 theories and discover that each is true. In this case the aesthete might claim there is inductive support for the next sampled beautiful theory being true. In other words, the argument might go as follows:

Beautiful theory T₁ is true
Beautiful theory T₂ is true
...
Beautiful theory Tₙ₋₁ is true

Therefore, (probably) Beautiful theory Tₙ is true.

Let us call this optimistic induction. We might imagine a somewhat less optimistic view that says that not all beautiful theories are true, but a majority of them have are. In such a situation, assuming there has been no bias in how the beautiful theories have been selected, the aesthete says we still have a posteriori evidence for beautiful theory Tₙ being true. The
pertinent question is how well does optimistic induction describe our experience in science. Clearly the history of science is not a random sampling process. However, bracketing this problem, there are additional worries. If we look at the history of science, it is better described by a process that is often called pessimistic induction:

Beautiful theory $T_{n+m}$ is false
Beautiful theory $T_{n+m-1}$ is false
...
Beautiful theory $T_{n+1}$ is false
Therefore, (probably) Beautiful theory $T_n$ is false.

As with optimistic induction, a less pessimistic induction ensues if only a majority of past beautiful theories have been false. Clearly if there is an empirical basis for pessimistic induction, then it undermines the Aesthete’s hope for beauty being an empirically discoverable sign of truth. A third possibility that neither optimistic induction nor pessimistic induction is justifiable, e.g., that beauty and truth are neither positively nor negatively correlated, also undermines the Aesthete’s position. What seems needed at this point is an empirical survey of the history of science to see whether there is more support for pessimistic or optimistic induction. It would seem that there is little for the philosopher to do than wait for the results of such a survey.

5.4 James McAllister’s Defense of Beauty as a Sign of Truth

Rather than wait for the results of an historical survey, the philosopher James McAllister, a modern Aesthete, argues that the meaning of beauty in science provides a reason for thinking that beauty might be linked to epistemic (or what he calls empirical) properties of theories such as probability of truth. He argues that what constitutes beauty in science changes over time making a probabilistic connection with truth more likely. Furthermore, because he thinks that many types of property are potentially aesthetic he is optimistic about our ability to discover links between aesthetic and epistemic properties of
theories if they exist. As McAllister gives a contemporary defense of the aesthete’s position, let us consider his position in detail. First he considers a list of properties of scientific theories, “to which aesthetic value could conceivably be attributed.” (McAllister 1996, p. 35) He acknowledges that this list would be very long and perhaps even infinite, but nonetheless says very little about how one would come up with such a list or even if it is possible for someone to do so. What he calls an aesthetic canon is enumerated by giving each property on the list a numerical weighting that reflects its aesthetic value. McAllister points out that most properties in the list will have zero weighting. He then gives an account of the dynamics of the aesthetic canon that can produce conceptions of beauty (or aesthetic value) that are signs of truth:

I propose the following model of the mechanism by which scientific communities formulate their aesthetic canons for theory evaluation. A community compiles its aesthetic canon at a certain date by attaching to each property a weighting proportional to the degree of empirical adequacy then attributed to the set of current and recent theories that have exhibited that property. The degree of empirical adequacy of a theory is, of course, judged by applying the community’s empirical criteria of theory evaluation. I name this procedure aesthetic induction (McAllister 1996, p. 78).

One might complain about the vagueness of some of McAllister’s machinery. For example, what does it mean for a theory to “exhibit” a property. Consider three toy theories: (1) All spherical viruses possess quasi-equivalent symmetry; (2) Some spherical viruses possess quasi-equivalent symmetry; (3) No spherical viruses possesses quasi-equivalent symmetry. Does McAllister want to say that all three “exhibit” the property of predicting the frequency of quasi-equivalent symmetry in nature? It is not clear. This problem becomes more acute if a property is defined in part by the theories in which it features. What exactly constitutes quasi-equivalence depends in part upon which of the above theories we endorse.

Nonetheless, let us consider an attempt to McAllister’s position more concrete. Presumably, the set of recent and current theories is determined by a scientific community.
To determine the weighting of a particular property $P$, say positing 532 symmetry, we take the subset of all the current and recent theories whose members posit 532 symmetry. We then consider the observed empirical adequacy of the members of the set. Although McAllister does not say anything as precise, he seems to assume that each theory $T_i$ in the set has a single measure of it empirical adequacy $E_i$. Let $E_p$ be the average value of $E_i$ of the members of the set of theories that possess $P$. The weighting of property $P$, which McAllister calls $W_p$, is proportional to $E_p$. A beautiful theory for McAllister is one in which scientists are “moved to project beauty into” as a consequence of the theory possessing one or more properties with positive aesthetic weight (McAllister 1996, p. 34). In other words, if a theory possesses one or more properties that are weighted highly in the aesthetic canon (and none weighted negatively?) and in virtue of these properties scientists call the theory beautiful, then it is a beautiful.

You might wonder if McAllister is giving an objective or subjective account of theoretical beauty in science. His account has both objective and subjective dimensions. Aesthetic properties of theories, for McAllister, are intrinsic properties of theories. As properties of theories they are objective. It is objective that a given theory postulates a certain type of symmetry, for example. (This objectivity is perhaps more difficult to defend in the case of aesthetic properties such as simplicity.) However, what makes these objective properties aesthetic depends on how scientists respond to them. Roughly the idea is that if a property of a theory (properly?) invokes an aesthetic response from a scientist it is an aesthetic property. Furthermore, theories that possess aesthetic properties and that scientists call beautiful in virtue of these properties are the beautiful ones. For McAllister the beauty of a theory is not an “intrinsic property” of the theory but is rather projected

---

173 I think that once McAllister’s account is laid out, there is a tension between defining an aesthetic property in terms of aesthetic responses or in terms of its correlation with empirical success. While McAllister argues that these two definitions will be co-extensive in the long run, they are not necessarily co-extensive in the short run even if his account is true.
Aesthetic values, such as beauty, are not located in the world, but rather projected into the objects by observers. An object of perception, such as a scientific theory, may have, among its intrinsic properties, some which evoke aesthetic responses in observers, for example inducing them to project the value of beauty into the object. I deem such properties aesthetic properties. A scientist is moved to project beauty into a theory by virtue of holding one or more aesthetic criteria, which attach aesthetic value to properties that the theory has. (McAllister 1996, p. 34)

One consequence of projectivism, at least according to John Mackie the champion of the view in ethics, is that value judgments are all literally false—they are not about objective features of the world as they purport to be. McAllister does not address this potential problem for his view.

Why does McAllister think that beauty may be a sign of truth? To put it simply, McAllister proposes that our notion of beauty is jerry-rigged so that it tracks (possibly with some lag) what we take to be the epistemic virtues of our best past science. He assumes that possession of these epistemic virtues or what he calls “good empirical performance” is a sign of truth:

If an aesthetic property were a sign of truth, a theory that exhibited that property would necessarily be true, and would therefore show the best empirical performance conceivable. In contrast, a property that has no link with truth may be found in theories of all degrees of empirical adequacy. Thus, we may recognize a property that is a sign of truth from the fact that its presence in a theory is correlated with good empirical performance.

(McAllister 1998, p. 177)

There is much to disagree with here. Beginning with the first sentence, it is possible that if an aesthetic property is found only in true theories, then this is a contingent fact. True theories of other possible worlds may not share the property. Thus the conditional
expressed in the first sentence is false. The final sentence of the quotation assumes some inductive principle since we have only access to observed empirical performance and a sign of truth, for McAllister, correlates with all empirical performance. Furthermore, an additional problem is that a false theory, even one far from the truth, can have good empirical performance. It would seem that McAllister overreaches. All his argument shows is that there might be a correlation between beautiful theories and empirical success (not necessarily probability of truth) and thus there may be some inductive support for thinking that the next beautiful theory will be empirically successful. However, there is an unbridged gap between empirical success and probability of truth.

Although he does not put it this way, let me make McAllister’s position more precise by interpreting him probabilistically. If there are positive correlations between scientists’ aesthetic evaluations of theories and their evaluations of empirical success, and scientists are reliable in both their evaluations, then it follows that discovering some theory has an aesthetic property would increase the probability of empirical success. As he puts it, “aesthetic judgment can yield an estimate of the empirical adequacy of theories that does not rely on empirical judgment.” (McAllister 1996, p. 92) Discovering that a theory has property P with a high aesthetic weighting, amounts to discovering that it has a property that has been possessed by past successful theories. As noted above, a weak link in McAllister’s argument is his link between good empirical performance and truth, or in my weaker probabilistic reading, the link between good empirical performance and probability of truth. McAllister addresses this issue in his conclusion:

It may be that some aesthetic properties exist that are correlated with high degrees of empirical adequacy. A scientific realist would describe this eventuality as one in which all theories that are within a certain distance from the true theory of the universe have particular aesthetic properties. (McAllister 1996, p. 203)

This response does not successfully rebut the charge that radically false theories might nonetheless possess a high degree of empirical adequacy (even if there is a correlation
between an aesthetic property and empirical adequacy). Even a realist would be willing to allow this possibility. Furthermore, McAllister relies on the vague notion of closeness to truth. His account would be much clearer and more complete if he had a metric to measure closeness to truth. Verisimilitude, however, proves to be a very difficult concept to explicate. Instead, I propose that McAllister’s notion can be given a probabilistic interpretation. As well as gaining clarity, a probabilistic reading allows McAllister a way of responding to the objection that radically false theories might be empirically successful. Let us say that theories that are correlated with high degrees of empirical adequacy and are consequently for McAllister’s realist “within a certain distance from the true theory” have on average a greater probability than theories that have lower degrees of empirical adequacy. This assumption requires justification but at least it is more defensible than McAllister’s original formulation as it allows that some ugly theories have high probability. However, if beautiful theories are disproportionately those with higher probability then picking a random theory and learning it is beautiful arguably increases the probability that it is true. On this reading, McAllister endorses the view that beauty is a sign of truth, if any property exists that empirically successful theories disproportionately share. Remember this property might not be considered aesthetic now, but McAllister claims that, given enough time, scientists will eventually discover its correlation with empirical success and deem it beautiful.

5.4.1 Consequences of McAllister’s Model

Before I begin with my objections to his account, let me sketch some of the consequences of McAllister’s position. McAllister is at pains to resist what he calls reductionism: that aesthetic and empirical evaluations are merely aspects of one another (McAllister 1996, p. 61). As he points out, scientists often treat them as independent. As Watson said of the multiple RNA-strand TMV model, “it is very very pretty but does nature always like to be pretty?” (See Chapter 2). No doubt McAllister is correct here: the question, “Is the theory true?” and “Is the theory beautiful?” are distinct questions with
different truth conditions. However, McAllister claims that “… in the longer term a correlation between scientists’ empirical and aesthetic evaluations tends to emerge.” (McAllister 1996, p. 65) Scientists’ answers to the question “is theory T beautiful” will be correlated with the answer to the question “were theories similar to T successful?” Notice that the latter question is ambiguous between whether the theories similar to T were judged to be beautiful by past scientists or by current scientists.

McAllister turns the usual way of looking at the role of aesthetics in science on its head. Typically aesthetic considerations are thought to operate in the context of discovery by fostering novel ideas in a progressive way. On the traditional account, the scientist is often compared to the artist who creates something new and importantly different from past creations (Engler 1994). For McAllister, science’s aesthetic dimension is conservative and backward looking. Aesthetic considerations constrain how different a new theory can be from the tradition. Assuming, as McAllister does, that at least some past theories have been successful, a theory radically different from all past theories will be ugly or at least certainly not beautiful. A scientist driven by aesthetic considerations cannot be a true revolutionary.

One of the most interesting claims of McAllister’s book is that genuinely revolutionary science is driven by empirical and not aesthetic considerations. For McAllister it is Kepler and Bohr, not Copernicus and Einstein that were the true revolutionaries.\footnote{De Regt (1998) argues that Einstein was motivated by aesthetic reasons and this fact undercuts McAllister’s claim.} Copernicus was motivated by his dislike of the Ptolemaic devices of the equant and eccentric that contravened the Aristotelian tradition of perfect spherical motion.\footnote{One would think that over a millennium of “successful” use of Ptolemaic astronomical devices would be plenty long enough for them to become part of the positive aesthetic canon. McAllister does not seem to} Kepler is forced to abandon spheres and adopt ellipses because of the data. McAllister argues that in the 20th century quantum mechanics represents revolutionary science whereas the special theory of relativity does not, being based on backward looking Galilean intuitions.

\begin{footnotesize}
\begin{itemize}
\item \footnote{De Regt (1998) argues that Einstein was motivated by aesthetic reasons and this fact undercuts McAllister’s claim.}
\item \footnote{One would think that over a millennium of “successful” use of Ptolemaic astronomical devices would be plenty long enough for them to become part of the positive aesthetic canon. McAllister does not seem to}
\end{itemize}
\end{footnotesize}
5.4.2 General Objections to McAllister’s model of Aesthetics in Science

What are we to make of McAllister’s proposal? Unsurprisingly, I think there are serious problems with McAllister’s account. I will begin with the more general problems and move to the more specific problems that arise when we apply McAllister’s ideas to the case study. I think there are admirable features of his account. He squarely addresses the problem of the rationality of aesthetic judgment or aesthetic “evaluation” to use his vocabulary. He bites the bullet and boldly claims that epistemic uses of beauty can be rational and warranted. He claims, correctly I believe, that if there is a linkage between beauty and empirical adequacy (or any aesthetic property and any empirical property) this cannot be known a priori but depends on the actual history of science. Part of the problem interpreting McAllister’s position is the vagueness of his term “theory appraisal.” McAllister thinks that we (inductively) generate an aesthetic canon for use in “theory appraisal.” The only sense I can make of his term “theory appraisal” is that we use the aesthetic canon along with more mundane empirical criteria such as consistency with extant data, consistency with other accepted theories, predictive power, explanatory power, etc., to appraise theories to have higher or lower weight and it is more rational to believe a theory with higher weight than one with lower. In other words, ceteris paribus, theories that possess properties positively weighted in the canon—postulate certain symmetries, possess easily visualizable models, minimal types of forces, etc.—”should be expected to reproduce” the empirical success of past theories that had these properties. This comparative conclusion—more beautiful theories are rationally believable—does not necessarily mean any theory should be believed, but merely that we can order theories in terms of believability. For example, it is more rational to believe that a small number of Yeti exist than to believe one billion Yeti exist, even though it is not rational to believe any Yeti appreciate this problem with his view.
exist. For McAllister’s notion of theory appraisal to have any teeth, he must be committed to at least the claim that if empirical criteria are equal, it is more rational to believe theories that are correctly judged to be more beautiful.

Before I mention specific criticisms let me make an important distinction. McAllister’s project has both a descriptive and a prescriptive dimension. First, as the above quotation suggests, McAllister aims to describe how scientists’ aesthetic sensibilities are actually shaped. Briefly, properties of theories are judged beautiful roughly in proportion to the degree they are possessed by past successful theories. This is a descriptive claim. Second, there is the prescriptive claim that because aesthetic sensibilities are formed this way (by “aesthetic induction”), it is rationally justifiable to use current aesthetic judgment in appraising a theory epistemically. Actually, in the end, McAllister hedges a little on his prescriptive aspirations, since he only gives a pragmatic justification of induction. As he puts it, “… it is rationally justifiable to allow our appraisals of theories to be shaped by aesthetic criteria. There may exist certain aesthetic properties—certain simplicity properties, certain symmetry properties, an so on—that are conducive to theories having high degrees of empirical adequacy.” (McAllister 1996, p. 204) Since the payoff is great if we discover such correlations we should continue to use the aesthetic induction argues McAllister: “[A]s long we cannot rule out that aesthetic properties exist which are conducive to theories’ having high degrees of empirical adequacy, it would be foolish to deny our chances of discovering them. We should therefore continue to perform the aesthetic induction.” (McAllister 1996, p. 205) The two dimensions of McAllister’s position—descriptive and prescriptive—can be detached. McAllister could be right about the way in which scientists’ aesthetic sensibilities are shaped and wrong about their warrant and vice versa. I am more concerned with the prescriptive dimension of McAllister’s claims, although I will have something to say about some of his descriptive claims also.

Let me begin my criticism with McAllister’s argument sketched in the preceding paragraph whose conclusion is that we should continue to use the aesthetic induction. It
seems to be a *non sequitur*. McAllister has not shown that performing the aesthetic induction is the only way to discover correlations between aesthetic and empirical properties of theories, if they exist. Perhaps there are other more efficient methods for discovering these purported correlations. Therefore it does not follow that we *should* continue to perform the aesthetic induction. Presumably McAllister means something a little weaker such as we would *not be irrational* to or it is *defensible* to continue to use the aesthetic induction. Even so, I think he underestimates the costs of performing the aesthetic induction if there are no projectible correlations between potentially aesthetic properties and truth. One probable consequence is that we will sometimes accept more beautiful theories that are less likely to be true. McAllister argues that we should find beautiful theories more believable than ugly ones. In believing a beautiful theory, or even believing it more strongly, because of its beauty, there is some risk that it will turn out to be false. Say there are two theories, one beautiful and one ugly that meet the non-aesthetic criteria equally. McAllister says we should find the beautiful one more believable. Adopting this attitude as a general strategy increases the chances of accepting beautiful *false* theories and rejecting true ugly theories. I maintain that if two theories meet the non-aesthetic criteria equally we should not prefer one over the other, in the sense of either believing one with more confidence or systematically betting on the truth of the beautiful one.

We should also be aware of the slippery meaning of what it is to be “using aesthetic induction.” There is a narrow and a broad interpretation. On the narrow interpretation, “using aesthetic induction” means merely *judging* a theory to be beautiful only if it possesses properties in common with empirically successful theories. Narrowly construed aesthetic induction is merely a model of the dynamics of aesthetic preference in science. In fact, calling it “induction” rather than “aesthetic change” is a little misleading. McAllister, I take it, calls it induction since he induces from the past empirical performance of science to a current conception of beauty. On a broad interpretation and the interpretation I think McAllister intends, “using aesthetic induction” means that additionally these aesthetic
judgments should shape how scientists act towards the world. Other things being equal, McAllister believes scientists should work on more beautiful theories. Additionally, when things are not equal, he believes scientists should sometimes work on more beautiful, but otherwise less epistemically valuable theories. However, working on beautiful theories at the expense of ugly theories or even biasing one’s beliefs in favor of beautiful theories has grave costs if beauty is not a sign of truth. In that case, using aesthetic induction broadly construed would impede scientists’ discovery of the truth—the goal of science according to McAllister—by wasting resources that could be more efficiently deployed. (I will have much more to say about the role of beauty in the logic of pursuit in Chapter 6.) Moreover, one could discover these purported empirical-aesthetic connections without allowing aesthetics to infect what theory we believe or what theory we work on during the process of discovery. In cases where the only relevant difference between theories is aesthetic we could suspend judgment and thus minimize the risk of bias and error. Why should we take the risk that there might be projectible correlations between aesthetic and epistemic properties of theories before we discover them? At the very least we are not rationally required to take on these risks as McAllister suggests.

Let me turn to another problem with McAllister’s account. He does not sufficiently explain why scientists consider so few properties aesthetic. He admits that the list of properties conceivably could be attributed aesthetic value or weight is potentially infinite, but only “a few” are given a non-zero weight. His notion of aesthetic induction suggests there should be many more. Consider the following contrived example. In a hypothetical scientific community there have been 10 theories, of which 5 have been empirically successful and 5 have not. Consider a property P that occurs randomly in half of all the theories. By random I mean that knowing that T_i has property P and T_k does not, is of know help predicting whether T_k has P. The probability that P occurs in only the empirically successful theories and not in the unsuccessful theories, i.e., the probability that P is as strongly correlated with empirical success as possible, is 5! 5!/10! = 0.003968 or about four
chances in 1000. If $P$ is as strongly correlated with empirical success as possible, then it receives the highest weighting in the aesthetic canon, according to McAllister. While this number might appear to be very small, if there are 250 properties or more like $P$, then we would expect at least one randomly occurring property to be the maximally correlated with empirical success. By randomly occurring I mean that by learning that one theory which has property $P$, we cannot improve our ability to predict which of the remaining theories also has $P$. To McAllister’s credit, in general, if the number of theories increases, then the chance that a randomly assigned property is maximally correlated with the empirically successful theories decreases. However, for a finite number of theories $N$, if there are a potentially infinite number of properties $P_1, \ldots, P_m$, as McAllister concedes there might be, and they each occur randomly in a fraction $(a/b)$ of the theories, then for any $N$ we should expect an infinite number of properties to be maximally correlated with success. Consequently, on McAllister’s view, we should expect an infinite number of aesthetic properties, clearly an absurd result. To see this, consider the expected number of properties maximally correlated with empirical success ($E$) which equals the number of properties ($m$) multiplied by the probability of being maximally correlated with empirical success:

$$E = m \times \frac{(aN/b)!((b-a)N/b)!}{N!}$$

As $m \to \infty$, $E \to \infty$, for any finite $N$.

In other words, if we assume that all potentially aesthetic properties are assigned to theories without regard to the empirical success of a theory (i.e., randomly), then we would expect some properties to be correlated with empirical success simply in virtue of the number of properties assigned. In this case, McAllister’s model would say that we should expect an infinite number of signs of truth. McAllister’s account falls apart if he allows an infinite number of properties and a finite number of theories. For McAllister, the unexplained mystery remains: why do scientists consider so few properties to be aesthetic? Furthermore, this hypothetical case also suggests that McAllister’s account is prone to predict signs of truth even when they don’t exist. Consider the $N+1$ theory, and any given
P, which is maximally correlated with success in the first N theories. P gives no information about the empirical success of theory N+1 since it occurs randomly (by stipulation), but P would be viewed as a beautiful according to McAllister.

When there is empirical support for a property that McAllister would call aesthetic, the role of aesthetics becomes epiphenomenal. Consider an example from evolutionary biology. It is quite common in constructing a phylogenetic tree to assume the most “parsimonious” tree. To give a slightly contrived example, if we are considering the evolutionary tree of blue jays, ravens, and humans, it is more likely that blue jays and ravens had a more recent common ancestor than ravens and humans or blue jays and humans. To make this claim we assume that the three species are related and the common ancestor of all three did not have wings. We make the assumption that wings only evolved (or devolved) once, i.e., the least amount or most parsimonious amount of evolutionary change needed to account for the distribution of wings. If humans had a more recent common ancestor with a species of bird than the most recent ancestor of oriel and ravens, then wings would need to have evolved (or devolved) more than once. The reason that we assume that wings only evolved once (and thus the simplest tree) is neither because simplicity is beautiful and beauty is likely nor that in general simplicity is likely, but rather because we have learned that evolving a wing is a rather improbable event in the history of evolution (Sober 1989). Evolutionary theories that have recognized that wings evolve infrequently have been successful. Thus, on McAllister’s model, assuming a parsimonious phylogenetic tree most likely would be considered an aesthetic property of the model. However, it seems to me, that aesthetics has nothing to do with the rationality of the assumption of a simple or parsimonious tree. Parsimony of wing-morphogenesis is an ordinary descriptive fact that has been discovered by evolutionary biologists and is partially empirically justified even before we test the above tree. Calling it an aesthetic property does not appear to make any difference to scientific practice.
5.4.3 Comparing McAllister’s Model with the Case Study

To examine McAllister’s proposal further, let us consider how well it applies to the Caspar-Klug theory case study. As I argued earlier, the Caspar-Klug theory has (at least) two properties that render it beautiful: (1) that it posits icosahedral symmetry and (2) that it posits quasi-equivalence symmetry. So far I think McAllister would agree with me. He calls symmetry a “class of aesthetic property”, one of four such classes that he discusses in some detail (McAllister 1996, pp. 40-44). He correctly points out that it is particular symmetries that we should examine, since scientists’ aesthetic preferences are specified more finely than the general class of symmetries. Let us consider the two symmetries one by one.

First, the Caspar-Klug theory postulates virus particles that possess icosahedral symmetry, also known as 532 symmetry. According to McAllister, if we judge a theory to be more probable on the basis of beauty, there must have been consciously or unconsciously an application of the aesthetic induction. To apply the aesthetic induction, we must look to find empirically successful theories that also have postulated icosahedral symmetry or quasi-equivalence symmetry. McAllister tells us little about how one would go making a selection of relevant theories. Does one look at other virological theories, other structural biological, other biological, other biophysical theories, other microscopic theories, other scientific theories, etc? Presumably one would get different results from the aesthetic induction depending how broadly we cast the net. Another way of putting the same problem is that Caspar and Klug are members of numerous communities: the x-ray crystallography community, the plant pathology community, the Anglo-American scientific community, the Cambridge community, among others. For McAllister “the aesthetic canon” is relativized to a community. Which community is the relevant community for making the aesthetic induction? McAllister says very little about this problem. Unless McAllister is willing to allow that warrant is also relativized to a community, then for his account to be complete he has to describe which community or discipline is relevant and
why. Without this extension to his account, it is possible that a particular symmetry may be warranted on aesthetic grounds from the point of view of physics and not from the point of view of biology. On my probabilistic interpretation of his claim, McAllister seems to allow the possibility that a symmetry increases the probability of a theory from the point of view of biology, but not from physics. I want to say that a beautiful symmetry either increases the probability of a theory or not, it does not matter if you are wearing your physicist’s hat or your biologist’s cap. On the other hand, thinkers such as Carnap and Kuhn relativize epistemic evaluation to broad conceptual frameworks. McAllister might argue that he too wants to relativize probability assignments to different disciplines. Unfortunately, the conceptual frameworks of Kuhn and Carnap are incompatible in the sense that we are forced to adopt at most only one at any one time. Physics and biology are consistent with each other—we don’t have to choose which to believe, and we can believe both at once. Indeed if our best physics told us some fact were probable, a biologist has to accept this probability assignment. The same is true of our theories of the world. In the end, we want to know which theories we should believe about the world and to do that we require only one probability value per theory. Science as a whole provides society with our best picture of the world. It is hopeless to say we should believe theory T when we are in a physics lab, but not believe it in a biology lab. It also conflicts with scientific practice.

Examining icosahedral symmetry brings out a further weakness in McAllister’s account. Theories that posit icosahedral symmetry are rare. In the entire history of science, there are only a handful of theories that have postulated icosahedral symmetry. I mentioned the two most famous in Chapter 2: Plato postulating that the universe and water exhibit icosahedral symmetry; and Kepler postulating that platonic solids lie nested within the orbits of the planets. There have been a few other occurrences of icosahedral symmetry in pre-1962 science. In the 1880s, Ernst Haeckel (1834-1919) aboard the HMS Challenger drew famous pictures of microscopic single celled creatures called radiolaria (Haeckel 1974). Two of them he named *circorrhegma dodecahedra* and *circogonia icosahedra*
because he thought they were shaped like dodecahedra and icosahedra. More recently the
icosahedral boron molecules such as $B_{12}H_{12}^{-2}$ were discovered (Lipscomb 1977). This
sample of theories seems too small to carry the inductive weight that McAllister’s account
demands. Also the theories are not part of any one discipline, let alone one community, but
are scattered around history of science. Plato’s theory is arguably not even part of science.
This scarcity raises the question, why do we find icosahedral symmetry aesthetically
appealing if in fact we do? McAllister’s account might be called an internalist account. For
him, the aesthetic weights we attribute to properties are determined by considerations
internal to the history of science. An alternative account might allow for external sources of
aesthetic value in science. For example, in addition to internal sources of aesthetic value, we
might transpose the aesthetic value that we attribute to properties of people, art, or
architecture into analogous properties in science. Two types of alternatives exist: purely
external accounts and mixed accounts. On the purely external account, sources of aesthetic
value in science always come from the culture external to science. A mixed view allows that
sources for aesthetic value from both inside and outside of science. McAllister’s internalist
account is overly restrictive. External explanations of why icosahedral symmetry fare better
than McAllister internal competitor. Consider the following speculative externalist
explanations of why icosahedral symmetry is beautiful. I suggest that the use prominent
use of platonic solids in geometry, beginning with Euclid’s Elements, partially explains why
we find them beautiful in empirical science (See Cromwell (1997) for a history of
polyhedra). As for why we find certain symmetries beautiful, recent work in evolutionary
psychology suggests that preference for bilateral symmetry has evolved as part of mate
preference.\footnote{As Richmond (1997) points out, bilateral symmetry probably evolved as a solution to locomotion in a
straight line. Each side of the bilaterally symmetric organism has effectively the same resistance due to

\textit{...}

\textit{...}
preference for icosahedral symmetry might be what Gould and Vrba (1982) call an exaptation—a trait that evolved for one purpose and becomes used for another. These explanations are speculative, but are just as plausible, if not more so, than McAllister’s model.

There are additional problems with applying McAllister’s account to the case study. Theories that posited 5-fold symmetries in crystallography have been notoriously unsuccessful. Of the few theories that postulate icosahedral symmetry more generally, most of these theories have been empirical failures. The universe is not a dodecahedron, there are no platonic solids interspersed between the planets, and Haeckel’s drawings reflect as much his own aesthetic taste as what actually exists in nature. The icosahedral boron theory has more empirical support, but it was, for the most part, unknown to virologists when Caspar and Klug postulated their beautiful theory in the 1960s. In the 1980s a new form of carbon, C\textsubscript{60}, called Buckminsterfullerene was discovered. This is a case of a successful theory that postulates icosahedral symmetry. However, given its discovery 20 or more years after the Caspar-Klug theory, it is irrelevant to judgments of the beauty of the Caspar-Klug theory at the time. To put the problem pointedly, McAllister’s scheme cannot account for the judgment that icosahedral symmetry as found in the Caspar-Klug theory is beautiful, since there have been too few successful scientific theories which have posited icosahedral symmetry. In general, McAllister’s backward looking account has difficulty allowing for the beauty of novel features of theories. Because there is little precedent for icosahedral symmetry, McAllister’s claim that there is some sort of inductive warrant for the beautiful theory rings hollow. That Boron forms icosahedral structures is no reason to think viruses will and does not make the Caspar-Klug theory more likely.

---

friction and the same locomotive power.  
\textsuperscript{177} There is only a crude empirical regularity known as Bode’s law.
Let us turn to the other symmetry postulated by the Caspar-Klug theory: quasi-equivalence. This property is difficult to define exactly. The idea is that two or more otherwise identical protein subunits, making the same contacts, can occupy slightly different positions in the structure. To visualize the idea, think of the Rubik’s cube. There are nine 2-dimensional squares on each face of the cube. Each square borders four other squares. A square in the center of the face borders four squares on the same face; a corner square borders two squares on the same face and two squares on different faces. Using the Caspar-Klug terminology, corner squares and center squares share quasi-equivalent positions: they both border 4 other squares but the angles that they subtend at the borders is variable. (Let us ignore the fact that a 90° angle is probably too large to be considered “quasi.”) Rubik’s cube is a tessellated cube. Caspar and Klug consider tessellated icosahedra. The idea of quasi-equivalence led to a novel solution on how to build closed viral structures out of more than 60 identical protein subunits. Some similar remarks apply to quasi-equivalence as they did to icosahedral symmetry. Before Caspar and Klug’s work, there had been few, if any, theories that had posited quasi-equivalence. On the other hand, the guiding assumption of quasi-equivalence can be found in early theorizing about protein bonding. It was not new to propose that proteins have specific bonding capabilities and will bond in the same way to a substrate. The genius of Caspar and Klug was to see that a regular pattern of bonding could be maintained in large structures if there was some degree of flexibility in the bond lengths and angles.

McAllister could argue that these previous conceptions of protein bonding could be material for the aesthetic induction in this case. No doubt one could argue that the Caspar-Klug theory was more likely because it assumes properties of proteins that were consistent with the then current view of protein-protein interactions. However to argue that this is an aesthetic reason for believing the Caspar-Klug theory would be a mistake. It looks like a straightforward case of induction rather than the more exotic aesthetic induction. Indeed what I think this example shows is that the distinction between an aesthetic and an empirical
property of a theory breaks down when you push McAllister’s account. McAllister faces the following dilemma. Either there is inductive warrant for a property of a theory or there is not. If there is no inductive warrant for a property of a theory, McAllister concedes that its beauty can do no epistemic work. Indeed properties that are not possessed by some successful theories will not be beautiful properties on McAllister’s account. If there is inductive warrant for a property of a theory and the property is one that scientists’ judge to be beautiful, then the warrant gained by a theory sharing that property would seem to be due to the existence of successful analogous theories and not the alleged beauty of property. To the extent that regular patterns of bonding have been part of successful theories of protein interaction (in enzymology, for example), Caspar and Klug’s use of quasi-equivalence does yield some warrant for their theory, but why connect this warrant with aesthetics? Any aesthetic considerations are epiphenomenal. The skeptic will ask, why does one need to consider beauty, all one needs is old-fashioned induction and perhaps analogy? If successful theories have been simple or posited certain symmetries, then the analogy will preserve these properties in the target theory. If you think that by focusing on icosahedral symmetry and quasi-equivalence symmetry I am being too specific, remember that icosahedral symmetry and quasi-equivalence are the properties in virtue of which scientists are “moved to project beauty into” the Caspar-Klug theory. They are, therefore, the relevant aesthetic properties according to McAllister (McAllister 1996, p. 34).

The criticisms of McAllister’s view become more acute if you reject the atomism he presupposes. McAllister assumes potentially aesthetic properties of theories are shared among a number of theories. To give an aesthetic property a weighting of its importance, McAllister asks you to consider the average empirical success of the set of theories that possess that property. This approach ignores the possibility that the aesthetic properties of a theory are often unique to the theory in question. Perhaps theoretical beauty is more closely linked to originality than McAllister allows. Consider the possibility that the beauty of the Caspar-Klug theory is not due either 532 symmetry or quasi-equivalence individually,
but the way in which Caspar and Klug ingeniously combine them. This combination of symmetries has never been seen before in the history of science. There is no set of past theories that have postulated these two symmetries together, i.e., no set of theories for McAllister to use to calculate aesthetic value.

5.5 Is the View that Beauty is a Sign of Truth Tenable?

Let me turn from a straightforward attack on McAllister’s positive position to an argument against the view that beauty is a sign of truth. There are two ways to approach this argument: directly by considering the posterior probabilities and indirectly by considering the priors and likelihoods via Bayes theorem. I will begin with the indirect route. The argument attempts to undermine the suggestion that beauty can have any influence upon a theory’s likelihood, \( p(E/T&B) \). My insight involves the a priori /a posteriori distinction. It is close to a received view that a priori knowledge cannot change the probabilities of contingent propositions. Thus, to change the probability of a contingent proposition one needs facts about the world. Grasping the meaning of a contingent proposition is not enough. Consider the likelihood, \( p(E/T&B) \). Assume that given \( T&B \), we were to generate some a priori knowledge of fact \( F \). It is not contentious to claim that \( p(E/T&B) = p(E/T&B&F) \). One could argue that knowledge of the beauty of a theory is an example of a priori knowledge \( F \). One could argue that given \( T \), the fact that \( T \) possesses certain properties, such as positing certain symmetries, is knowable a priori. On the other hand, it is possible that an incomplete theory leaves the symmetry group of its posits undetermined. Nonetheless, limiting ourselves to a complete theory \( T \), a sufficiently competent user of \( T \) can determine the symmetry of its posits a priori. In this respect, I maintain that a theory as an abstract object is unlike a concrete artwork.\(^{178}\) To determine

---

\(^{178}\) The contrast between artworks and theories fades if we consider artworks that are not concrete. For example, Beethoven’s 5th symphony is not merely the sum of its performances, but an abstract arrangement
whether a given painting exhibits a given symmetry requires one to investigate the world. For example, one can look at the painting or ask the painter, etc. But one is not required to investigate the world to determine if a theory posits a particular symmetry. The further judgment of whether any property is beautiful and whether the theory itself is beautiful also does not require further worldly investigation. As long as one has the requisite faculty of taste, one can judge if a given property and the theory are beautiful or not—it does not require an experiment. Since these judgments are a priori, they cannot affect the value of the likelihood, p(E/T&B).

In case you are still unconvinced, let me point out that the last argument relying upon the a priori/a posteriori distinction can be modified and redirected against the broader claim that beauty affects the posterior probability of theory, p(T/E&B). One does not need to know anything about the world to determine if a theory possesses certain properties such as simplicity or symmetry and whether these properties are beautiful. All one requires is that one grasps the meanings of the sentences that represent the theory and have the requisite sense of taste. The argument might be represented deductively as follows:

________________________

of notes.
(A). A biological theory consists of contingent propositions.

(B). Facts knowable a priori cannot increase or decrease the probabilities of contingent propositions.

(C). Given an understanding of a theory T and background knowledge B, further facts regarding its beauty are knowable a priori.

\[ \therefore \] (D). Given the understanding of a theory and background knowledge B, facts about its beauty cannot increase or decrease the probability of a biological theory.

As it stands, the argument is still quite crude and open to objection. Consider an objection from Peter Kosso (personal communication). Kosso questions the truth of (B). He argues that logical consistency can be known a priori and since we do not want an inconsistent theory, we rule out inconsistent theories a priori. Ruling out these theories increases the probability of the remaining consistent theories. Thus, a priori knowledge affects the probabilities of contingent propositions, or so Kosso argues. This type of intuition can be found in the literature. Richard Swinburne argues that what he calls an analytic truth—that the Copernican theory entailed very similar observable consequences as the Ptolemaic theory—\textit{decreases} the probability of the Ptolemaic theory and increases the probability of the Copernican theory (Swinburne 1973, p. 60). Let me quickly reply to this type of worry. Both on a subjective and an objective account of probability, the probability of an inconsistent theory is zero. We mistakenly can think that an inconsistent theory has a non-

\[ \text{\footnotesize 179} \] Given my immediate audience, let me be clear about what I am not saying. I am not claiming that the evidential relation, “E is evidence for T” is empirical, as Achinstein claims, although it may well be (Achinstein 1998). My argument works if the probabilistic evidential relation is a priori or empirical. If the evidential relation is a priori, as is the received view among philosophers of science, then given a theory T, one can specify a priori what would constitute evidence E for the theory. For example, if the theory is “All Ravens are Black,” then one need not consult the world to know that the existence of a black raven would be evidence for the theory. One would, however, need to consult the world to know if any black ravens exist. While Achinstein and his opponents disagree about whether what would constitute evidence can be known a priori, they agree that once it is known what would constitute that evidence, whether the evidence exists is contingent and cannot be known a priori. To use the words of Achinstein, I am claiming that if beauty of T were evidence (construed probabilistically) for a theory T, then it would be evidence that could be known a priori. But since we do not think that evidence can be gained a priori, the beauty of T cannot be evidence for a theory.
zero probability, but we do not increase or decrease its probability by discovering that it is inconsistent. If the probability of an inconsistent theory does not change when we rule it out, then ruling it out cannot affect the probability of the remaining consistent theories. Likewise the probability of an analytic truth or a truth knowable a priori is 1. Adding this truth to the background knowledge does not affect the probability of any proposition. One way to conceive of the situation is to understand that the background knowledge consists of all its logical consequences including all analytic propositions.

One might also object to (C). For example, McAllister might argue that to know which properties contribute to a theory’s beauty you need to know the properties of the set of successful theories. Which theories have been successful is obviously an empirical question. Thus using a conception of beauty, such as McAllister’s, that depends upon the properties of successful theories would appear to require that a judgment about a current theory’s beauty also be an empirical question. Premise (C), however does not claim that it is a priori question whether a theory is beautiful, it claims that given understanding of the theory and background knowledge, one can determine if a particular theory is beautiful without worldly investigation. Now when we are considering the posterior probability of a theory $p(T/E&B)$, it does not seem unreasonable to say that one understands theory T and we know the background knowledge. It is background knowledge after all. Nonetheless, it is possible that the background knowledge does not contain what is needed to determine a theory’s beauty.

A more effective objection to (C) is to claim that theoretical beauty depends on future empirical facts. For example, a Whewellian-inspired conception of beauty might claim that a theory is beautiful if it is coherent with respect to present and future observations. On this view, whether a theory continues to be coherent depends on what the future observations will be and thus is an empirical question that depends on future observation. Future observation is not part of background knowledge. If this conception of beauty were correct, then it would blunt the force of the above argument. However, as a
conception of beauty, it runs counter to common sense. Scientists judge today’s theories to be beautiful even though they do not know what the tomorrow’s observations will be. They do not make tentative judgments revisable upon further empirical data. Moreover, scientists continue to judge some theories as beautiful even when the theories have been conclusively refuted. While one might grant that one’s sense of taste is fallible, its fallibility is not due to the possibility of further observation that shows that a theory deviates from reality in certain ways. For example, scientists judge Crick’s comma-free genetic code “most beautiful” even though it has shown to be false in the most important respects (Maynard Smith 1999).

McAllister might reply that his position amounts to the claim that $p(T/B&E) > p(T/E&B')$ where $B'$ is the set of propositions $B$ minus any propositions about the beauty of $T$. McAllister might concede that from the specification of a theory one could derive whether a theory postulates a particular property or not. However, McAllister resists the idea that one can judge whether any particular property is beautiful without implicitly assuming facts about the history of science. In opposition, I have claimed that as long as one has the appropriate sense of taste, one can make this judgment. Thus both the claim that $T$ postulates property $P$ and the claim that $P$ is beautiful do not add any empirical facts to the background knowledge and consequently $p(Ts/E&B) = p(Ts/E&B')$. McAllister, I presume, would accuse me of smuggling in empirical facts when one engages one’s sense of taste. For McAllister, in correctly judging theory $T$ beautiful because it postulates property $P$, one also implicitly claims that $P$ has been postulated by past successful theories. This fact is somehow smuggled along with aesthetic judgment. As indicated above, I am skeptical of this claim. Consider the following thought experiment: suppose theory $T$ was the very first scientific theory put forward. Could we judge that it was beautiful? Unless he supplements his account, McAllister would have to answer in the negative since there is no history upon which to build an aesthetic induction. On the contrary, I claim we would still be able to judge that the theory is beautiful and perhaps even give it a weight. There are sources of aesthetic value outside of science. As indicated earlier certain symmetries may
strike us as beautiful for evolutionary reasons. Other aesthetic properties of theories may have analogs in art or in everyday life. From these external sources, we can evaluate the beauty of even the first scientific theory. In this hypothetical case, McAllister and I would agree that there would be no extra-empirical warrant, although for different reasons. Nevertheless, I think it is more plausible to say that a theory’s beauty does not depend upon when it was proposed. Beauty is not tethered to the contingencies of the history of science.

Perhaps the Aesthete concedes that a priori knowledge cannot change the probability of a theory, but she might argue that it determines the prior probability of a theory. In the previous argument I considered the probability given a body of background knowledge. If we have no background knowledge then it would appear that a priori facts are all we have to use in determining the prior probability of a theory. That we can determine the prior probability of a theory a priori has motivated previous philosophers of science. Carnap’s project of assigning probabilities to state descriptions (possible worlds) comes to mind (Carnap 1952). His attempt is widely seen as a valiant failure. Perhaps the most famous attempt to derive probabilities a priori is the principle of indifference, the rule that assigns equal probabilities to mutually exclusive and exhaustive outcomes in cases where we are ignorant about the mechanism for generating the distribution of outcomes. Because the set of outcomes can be differently partitioned, the principle leads to contradictory results. These failures have not stopped subjective Bayesians who think that probabilities represent degrees of belief and consequently assign a theory’s probability to be proportional to one’s degree of belief in the theory. Let us consider how beauty might be used to assign prior probabilities to theories. Suppose $T_1$ is more beautiful than $T_2$. The Aesthete might claim $p(T_1) > p(T_2)$. If $T_1$ and $T_2$ each entail the observed data $e_1, e_2, \ldots, e_n$, then $p(T_1/e_1 & e_2 \ldots, & e_n) > p(T_2/e_1 & e_2 \ldots, & e_n)$. If this were true, then the Aesthete would have a significant victory in showing that beauty is a sign of truth.

Although intuitively appealing, the Aesthete’s suggestion has a near fatal flaw. I discussed this flaw in Chapter 4. Consider a logically weakened version of $T_1$, that we will
call $T_1^*$. By logically weakened I mean that $T_1$ entails $T_1^*$ but $T_1^*$ does not entail $T_1$. If $T_1$ is beautiful or more beautiful than $T_2$ it does not necessarily follow that $T_1^*$ is beautiful or more beautiful than $T_2$. For example if $T_1^*$ is the proposition that $T_1$ is true in our solar system, even if $T_1$ is beautiful, $T_1^*$ is not beautiful, or at least less beautiful, than $T_1$ because $T_1^*$ includes an ad hoc restriction on its scope. The Aesthete’s above principle says that if $T_1$ is more beautiful than $T_1^*$, then $p(T_1) > p(T_1^*)$. However, this instantiation of the aesthete’s principle violates the probability calculus since if $T_1$ entails $T_1^*$, by the axioms of probability $p(T_1) \leq p(T_1^*)$. The aesthete must do quite a lot of footwork to circumvent this objection. (See section 4.4.2.) One might think that abandoning a relative notion of beauty, where we can have more or less of it, might solve the problem. However, the problem re-emerges even if beauty has no degrees, but merely is either present or absent in each theory. Without further restrictions, it is possible that a beautiful theory entails a weaker ugly one and the problem reemerges. A drastic reply by the aesthete is to first restrict the set of theories to which beauty can be attributed. For example, if she makes the restriction that each member the set of theories bear no entailment relationship to any other member, then the objection loses its force. If Newton’s theory entails Kepler’s laws and we agree that Newton’s theory is more beautiful, then either Newton’s theory or Kepler’s laws are not genuine theories. A less drastic restriction would be to stipulate that no theory can entail a less beautiful theory. That is, stipulate that if it appears that a theory entails a less beautiful rival then it or the less beautiful consequence is not a theory that should be considered as a live possibility. This move is perhaps less ad hoc than it appears since we do not have to choose between $T_1$ and $T_2$ if $T_1$ entails $T_2$. We cannot adopt $T_1$ without also adopting $T_2$. The aesthete’s suggestion then is to prune the possible theories into deductively isolated theories; none of which entail any of the others. Once this pruning is done, one can assess the beauty of each member without violating the probability calculus even if beauty is a sign of truth. But as discussed in section 4.4.2, the aesthete’s justification of a priori prior
probabilities is no stronger than the anti-aesthete’s who thinks that ugliness is a sign of truth.

What then are we to say about the relation between beauty and truth. On my account any general connection between beauty and truth (or falsity) would be an empirical connection. So far I agree with McAllister, although I do not think theoretical beauty is jerry-mandered to match empirical success. Discovering this connection would be an empirical discovery. Because I do not think beauty is jerry-mandered, I think the probability of discovering a link between beauty and truth is lower than McAllister does. If this general fact were part of the background knowledge and if Tb is a beautiful theory, then it is relevant to the probability of the Tb. If it is knowable a priori that Tb is beautiful, then adding this particular fact to the background knowledge does not increase the probability of Tb. This is because in a sense the background knowledge already contains this (an all other) a priori facts. What changes the probability of a given theory is whether a general claim about the empirical connection between beauty and truth is part of the background knowledge. Consider a particular beautiful theory Ti. “Adding” the knowledge that Ti is beautiful to the background knowledge does not change the probabilities: \( p(Ti/B) = p(Ti/B & Ti \text{ is beautiful}) \). However, it does not follow that beauty is irrelevant to its probability since whether there is a general connection between beauty and truth is relevant. Consider the general claim that all beautiful theories are true. Whether this is part of the background knowledge does affect the probability of the theory: \( p(Ti/B & \text{most beautiful theories are true}) > p(Ti/B & \text{most beautiful theories are false}) \). So on my view, we have a surprising result: that a given theory is beautiful does not affect its probability, but a general connection between beauty and truth does. Beauty is not a sign of truth as I defined it using conditional probabilities, but it is possible that beauty is relevant to a theory’s probability. Whether it is relevant, that is, whether there is a general connection between beauty and truth, is an empirical question—one for scientists or historians of science, not philosophers, to answer.
5.6 Is the Argument against an Epistemic Role for Beauty Too Good?

Let us change direction and address a potential objection to the principle thrust of my argument so far. It might be objected that my argument proves too much. As well as attack the view that beauty and truth have an a priori connection, the same type of argument also appears to “work” against any inductive principle. This conclusion—that no inductive principle is a priori justified—however is unacceptable an objector might argue, since it (allegedly) forces us into skepticism about science, the view that no scientific theory is rationally believable. (Note that to be consistent, the skeptic must hold that if T is a theory then ~T is not.) Thus by modus tollens we should reject the central thrust of my argument. To reply to this type of objection takes us out of philosophy of science and into epistemology. It would be presumptive of me to attempt to solve deep epistemological problems en passant, but I hope the reader will allow me to gesture at a reply to the skeptic that is in the same spirit as the approach articulated in this dissertation.

My machinery can be used to illuminate the skeptical worry. Consider the definitions of sign of truth: (1) Property P of theory T is a sign, of truth only if p(T/P) > 0.5; and (2) P is a sign of truth if and only if p(T/P) > p(T). I considered the case where property P is beauty of T. I have argued that in the case of beauty these two principles are not a priori, but rather empirical principles. If they were a priori true, then there would be a convincing reply to the problem of beautiful rival theories, but they are not. Instead, I suggested that if P were an empirical property it would require empirical evidence to determine if (1) or (2) were true. If it is a priori that a given beautiful theory Tb is beautiful, then (2) is false because p(Tb/P) = p(Tb). Now consider the response which says, what if P were some sort of inductive support, for example, what if P were the property of T having observed empirical adequacy? Cannot one run the some objections that I have used in the previous chapters and show that observed empirical adequacy is not an a priori sign of truth? Indeed the argument appears to work for empirical adequacy proper as well. (A
theory is empirically adequate proper if it has observed empirical adequacy and adequacy with respect to all future observations as well.) Is observed empirical adequacy or empirical adequacy proper an a priori sign of truth, in either sense?

My answer is a qualified no. I do not think there is an a priori justification for observed empirical adequacy being a sign of truth. If there were an a priori justification, it would effectively be a solution to the problem of induction, or at least Goodman’s new riddle of induction. Such a justification would say that there is a general way to establish the high probability of a theory. However, there is no general a priori way to rule out the possibility of competing empirically adequate theories. The problem of induction is a genuine philosophical problem. On the bright side, we should also be suspicious of a priori arguments that suggest that competing empirically adequate theories must exist for any theory T. Whether a given phenomena is multiply describable by science depends upon what the phenomena in question are. On my view, it is an empirical question whether an inductive principle works in a given domain. To borrow an example from John Norton, if we examine a piece of substance X and determine its melting point, are we justified in inferring that the next piece of X we examine will have the same melting point? Well it depends upon the empirical question of what X is. If X were the element bismuth, then we would be justified given facts about elements. However, if X were a piece of wax, then we would not be justified in the inference since wax does not have a uniform composition and the melting points of different pieces of wax vary.

We can see the empirical nature of inductive principles in my sign-of-truth framework. Whether p(T/T has observed empirical adequacy) > 0.5, depends upon whether the observed phenomena that T covers allow for competing theories with observed empirical adequacy. If there are other competing theories, then learning this fact is relevant to the truth of the inequality. If there are other competing theories, however, they are relevant only if they don’t have significantly lower prior probabilities (See Section 4.3). Likewise, whether observed empirical adequacy is a sign of truth depends upon empirical facts. This
is true even if prior probabilities are determined a priori. To see this remember that \( p(T | T \text{ has observed empirical adequacy}) = p(T) \frac{p(T \text{ has observed empirical adequacy} | T)}{p(T \text{ has observed empirical adequacy})} \). Since, a true theory will be empirically adequate, \( p(T \text{ has observed empirical adequacy} | T) = 1 \), if \( p(T \text{ has observed empirical adequacy}) < 1 \), observed empirical adequacy is a sign of truth (\( p(T/T \text{ has observed empirical adequacy}) > p(T) \)). Whether \( p(T \text{ has observed empirical adequacy}) < 1 \) is an empirical question. This claim has both a moderate and a radical reading. On the moderate reading it is empirical because it depends upon what the observed empirical phenomena actually are and this is contingent. It could be that all theories about a particular phenomenon are empirically adequate. On the radical reading, even given a particular \( T \) and given particular observed phenomena, it is still empirical whether \( p(T \text{ has observed empirical adequacy}) < 1 \). Nonetheless, on either reading determining \( p(T/T \text{ has observed empirical adequacy}) \) requires worldly investigation. Whether or not observed empirical adequacy is a sign of truth is an empirical question.

One might object to my empiricism on the grounds that either a circularity or an infinite regress looms. Relatedly, recall that in Chapter 1, I accused Newton of potentially introducing circularity if he justifies his inductive simplicity principle on empirical grounds. If one justifies a principle of inference using a fact then the same fact cannot be justified by the principle. For example, if we claim that we can infer the simplest hypothesis because of the general fact that the world is simple, we cannot justify the fact that the world is simple because of simple facts inferred using the general fact. Likewise, I claim that whether observed empirical adequacy is a sign of truth in any given case depends upon empirical facts about the case. This is circular only if we inferred these facts in a like manner via assuming them. However, to infer facts that license one inductive inference we use different facts (Norton 2003). For example if I infer that the next piece of bismuth will have the same melting point as the last (271.3°C), the inference is licensed by the fact that bismuth is an element. The correct inference that concluded that bismuth is an element was also licensed by facts, but they included different facts. Does an infinite regress loom?
Potentially, but the pertinent question is whether it is a vicious regress or not. How these chains of inference terminate, I think, this is still an open question. Presently the chains of inference beginning with melting points of particular elements terminate in facts about particle physics. Furthermore, what type of termination occurs, if any, is a question for scientists not philosophers. Indeed, in an important sense what is it for scientists to make progress is to give deeper reasons (i.e., deeper facts) for why certain regularities obtain.

Is the skeptic then right? Can we never be certain that we have probable knowledge? Yes and no. He is right in that there is no a priori proof that there is no infinite regress of facts. Relatedly I agree that there is no a priori proof that all phenomena cannot be multiply describable in equally empirically adequate ways. On the other hand, there is no a priori proof that all phenomena can be multiply describable in equally empirically adequate ways either. Likewise, I have argued that there is no a priori proof that all beautiful theories are likely or even more likely than ugly ones. Both the skeptic and the aesthete make the mistake of endorsing generic, global, a priori claims about issues that are specific, local, and empirical. Further, focusing on beauty does not help us solve skeptical worries about under-determination of theories by the evidence. The difference between my rejection of the aesthete’s position and the skeptic’s rejection of all knowledge claims is that my arguments, unlike the skeptic’s, do not undermine the practice of science. Science would cease to be as effective if scientists worried about whether we are merely all brains in vats, or whether the world was created 5 minutes ago. On the contrary, having some scientists examine ugly possibilities removes what would be an unjustified bias in science: to look only at beautiful theories.\textsuperscript{180}

\textsuperscript{180} Quine (1963) identifies a related bias in science with respect to simplicity. He leaves it less clear whether he thinks it should be eliminated however.
5.7 Summary and Conclusion of Chapter 5

To conclude this chapter, I have examined whether beauty increases the probability of a theory and have produced mixed results. I suggested that we should not a priori rule out an empirical correlation between our empirically successful theories and our beautiful ones. On the other hand, I have argued against an a priori connection between beauty (either generally or in the form of symmetry or coherence) and truth. If we consider a Bayesian account of a sign of truth defined in terms of conditional probabilities, then beauty cannot be an a priori sign of truth. Of course, one philosopher’s modus ponens is another’s modus tollens – an ardent aesthete might argue that we should reject the Bayesian apparatus rather than abandon the notion of an a priori sign of truth. Nonetheless, on examination of cases of coherence and symmetry, the aesthete who argues for an a priori connection between beauty and truth has no adequate justification. Her argument either relies on illegitimate theological faith or the aesthetic property in question is too weak to guarantee high probability or even a rise in probability. For example, I argued the Whewell’s notion of coherence, either construed in terms of scope or probabilistically as Achinstein does, does not require that coherent theory have a higher probability than a rival, all other things equal. Whewell himself misconstrues the problem in two different ways. He argues that it is an empirical question whether a given theory is coherent, but an a priori one that a coherent (beautiful) theory is more likely than a set of incoherent (ugly) theories. In direct opposition, I hold that to determine the beauty of a theory does not require empirical investigation and is thus an a priori matter. Additionally the general connection between beauty and truth or beauty and probability of truth is empirical. On this latter question, I am in agreement with McAllister, who has the most comprehensive contemporary account of the epistemic role of aesthetics in theoretical science. Unlike my account where it is a purely contingent whether there is a correlation between empirical success and beauty, McAllister’s account conforms the meaning of beauty to properties common to past successful theories and while this gerrymandering does not guarantee that
there will be a connection between beauty and truth, it makes it more likely. Using the Caspar-Klug theory as a case study, I argued that McAllister’s account fails to explain why the Caspar-Klug theory is beautiful and suffers from a number of additional weaknesses that make his model unattractive. Fans of a positive role for beauty in science will probably be disappointed in my conclusion that there might be an empirical connection between beauty and truth. Their disappointment will be alleviated somewhat in the next chapter where I will turn to the question of additional non-epistemic roles for beauty in science.
Chapter 6  Beauty and the Logic of Pursuit

Structure:

6.1  Introduction
6.2  Broadening the Project: from Means to Goals and from Belief to Action
6.3  The Logic of Pursuit
6.4  A Brief Survey of Views
  6.4.1  Herbert Feigl: Beauty has no Role
  6.4.2  Pierre Duhem: Beauty is a Goal of Science
  6.4.3  Richard Feynman: Beauty as Physical Laws
  6.4.4  Henri Poincaré: Beauty as Motivation
  6.4.5  Thomas Kuhn: Beauty as Subjective Value
  6.4.6  Taking Stock
6.5  A Decision Theoretic Analysis
  6.5.1  Goals and Means in Decision Theory
  6.5.2  Beauty and Decision Theory
  6.5.3  Two Decisions: by the Society and by the Individual
  6.5.4  Inter-value Realism and Inter-value Anti-realism
  6.5.5  Goals as Constraints on Utility Values
  6.5.6  Ideal Deliberation: Individual Inter-value Realism
6.6  Poincaré, Duhem, and Kuhn Reconsidered
  6.6.1  Duhem Reconsidered
  6.6.2  Duhem Reconsidered
  6.6.3  Kuhn Reconsidered
6.7  Beauty and The Freedom of Inquiry: A Proposal
  6.7.1  Freedom of Inquiry and Decision Theory
6.8  Summary and Conclusion
6.1 Introduction

If the argument in the last two chapters has been successful, then the aesthete is wrong to claim that beauty is an a priori sign of truth in any probabilistic sense. I have argued that beauty does not play an a priori epistemic role in science and hopefully have at least convinced you that if it does play a role, then it is an empirical question that has yet to be settled. I would like now to broaden the focus of my investigation and consider whether beauty plays a non-epistemic role in science. In other words, bracketing the possible epistemic role for beauty, is there some other function, or functions, for beauty in science and if so what would they/it be? I have some trepidation proceeding in this direction since it would seem that this question has potentially many different types of answer and I will not have space to consider each in the detail that it deserves. Moreover the issue would appear to be a can of worms to be opened only by the foolish.\(^\text{181}\)

---

\(^{181}\) There are many ways to approach the non-epistemic role of beauty in science. One can imagine a multitude of possible roles for beauty in science. Consider the following list:
1. No role: Beauty plays no non-epistemic role in science.
2. A role in pursuit: Beauty provides a reason to pursue a theory over an ugly rival
3. A role in theory acceptance: While beauty might not make a theory likely, it nonetheless provides a reason for accepting a theory.
4. A goal of science: Creation of beautiful theories is what science aims to do (among other things).
5. A motivator of scientists: The reason that scientists study nature is that they are seeking beauty.
6. A means to some other goal of science: Previously I have been considering whether beauty is a sign of truth, but perhaps there are other values that can be achieved through gaining beauty. There are a number of sub-positions here depending on which goal beauty helps one achieve.
7. A promotional role: Beauty is a way of promoting science to the general public.
8. A discovery of science: Beauty is discovered by science.
9. A pedagogical role: Beauty in science is used to teach science.
10. A justification of science in general: The beauty of science justifies scientific endeavors.
11. A source of inspiration: Contemplation of beautiful structures leads to increased creativity.
12. A means of idealization: Constraints upon how good idealizations are formed.
13. A means to improve the comprehensibility of complex data:
14. A conception of beauty as part of a paradigm or research program.
15. Beauty is explained by science: Science will explain why we judge beautiful things beautiful. This is not exactly a role for beauty in science as it is usually understood, but it a position adopted by some scientists.
16. Beauty as an ideal of experimentation. (See, for example, Parsons and Rueger (2000) and Holmes (2001))
6.2 Broadening the Project: from Means to Goals and from Belief to Action

To see two related ways in which one could broaden my approach, consider what I have argued in previous chapters. I have implicitly assumed that truth is a goal of science and that if beauty were relevant to that goal it would stand in some probabilistic relation to truth. If beauty were a sign₁ or sign₂ of truth it would be relevant in assessing which theory to believe, since ceteris paribus, a beautiful would be rationally believable (sign₁) or more rationally believable (sign₂). To broaden the project, at least two avenues are open.

First, we might consider broadening the means-goal structure I have assumed. For example, one might wonder whether beauty is a means to a goal of science other than truth. On this avenue beauty gains its value to science insofar as it is a means to an important goal of science. For example, Foster and Sober (1994) argue that simplicity is valuable insofar as it increases predictive accuracy. Another possibility is that beauty is a goal of science in its own right. That is, beauty has independent value in science. The primary value of theoretical beauty in science may not depend upon how it relates to other properties of theories. A third possibility is that beauty has value in science, but, as the argument might go, neither as a goal of science nor as a means to a goal of science. It could be what van Fraassen (1980) calls a “pragmatic virtue.” It could be valuable to us, but only due to contingent properties of humans and not due to properties common to all possible inquirers. However, even for van Fraassen, pragmatic virtues are still properties of theories for which scientists should strive and as such I will consider them goals of (our) science, although perhaps not primary goals. The pertinent question then becomes, how can we compare the different types of value—aesthetic, epistemic, pragmatic, etc.—operative in science? Is there an ordering in terms of priority or importance or is the relative importance of the different types of value essentially contextual? Thus one was to broaden my project is to consider the relation of beauty to the goals of science.

Second, in the preceding chapters I have focused on whether theoretical beauty is relevant to belief. The natural progression is to consider the relevance of beauty to action in
science. In particular, does a theory’s possession of beauty relevant to how a scientist should act toward the theory? For example, instead of claiming that its beauty gives us a *reason to believe* a beautiful theory, the aesthete might claim that beauty gives a *reason to pursue* a theory. In other words, beauty might be a reason to work out the consequences of a beautiful theory, apply it to more complex systems, add or reformulate its assumptions, or, in general, devote time and money to develop the theory and its potential evidence, etc., even if beauty is not a reason to *believe* the theory. A consideration of reasons for pursuit is said to be a consideration of the “logic of pursuit” in contrast to the better-known “logic of justification.” (Curd 1980) Reasons for pursuit of a theory are not unimportant. Scientific research has significant costs and with a limited budget we as a society should promote the most worthy projects first. In the United States, this task is in large part given to federal funding agencies such as the National Science Foundation (NSF) or the National Institutes of Health (NIH). Furthermore, individual scientists want to make reasonable choices as to which projects to invest their time and energy. Reasons for action and goals are related. On various views of scientific rationality, the goals of science give scientists standards to determine good reasons for action.

6.3 The Logic of Pursuit

Before proceeding let me say a little more about the logic of pursuit. Peter Achinstein (1993) proposes that philosophers, such as N. R. Hanson and Charles Peirce, who have considered reasons for the pursuit of a theory make (at least) six claims:

1. The reasoning is weaker than that which establishes a theory or shows that it is probable.
2. The reasoning occurs before experimental tests have shown the theory to be probable or improbable.
3. The reasoning is governed by a “logic.”
4. The reasoning may involve an appeal to certain general methodological criteria in addition to explanation or deduction of phenomena, e.g., “simplicity” and “testability.” Or to put the point differently, such general constraints on the explanation or derivation may be imposed.

5. The reasoning may also involve an appeal to the practical costs of pursuit, such as time, money and energy.

6. Finally, the reasoning is different in kind from that employed to establish a theory or show that it is probable. (Achinstein 1993)

To make progress understanding how beauty plays a role in the pursuit of a theory, let me briefly consider which of these six claims is potentially relevant to the question of whether beauty plays a role in the logic of pursuit. Given the last two chapters that have argued that beauty need not influence a theory’s probability, I will accept that claim 1 is appropriate. As for claim 2, once there is sufficient experimental evidence for a theory, then the initial reason for pursuing it may lose its force. However, on the other hand, there may be reasons to continue pursuing a theory even when we know that it is probable. For example, we may want a higher degree of probability or we may want to see if the theory can be modified to explain different types of phenomena outside the scope of its original formulation.

Although claim 2 might express the typical time that reasons for pursuit are offered, I think it is overly narrow or our purposes. Exactly what claim 3 entails is not clear. While it is often not made explicit what exactly it is to be a logic of pursuit, I assume it implies that there exist some forms of reasoning whose validity does not vary from person to person. Deductive logic is the exemplary case—it is the form of an argument that determines deductive validity. Analogously, in the logic of pursuit, if all the premises are made explicit, the validity of reasoning should not depend upon where it is made, when it is made, or who is making it. At least this is the ideal to which theorists strive. Admittedly this elaboration of the logic of pursuit is still loose. In practice, claim 3 amounts to a search for general principles that govern reasoning about scientific pursuit. Claim 4 is the most obvious place
where beauty might be relevant to theoretical pursuit. For example, as discussed in Chapter 1, many scientists take simplicity to be an aesthetic notion. Furthermore, prioritizing simple theories over complex and beautiful theories over ugly appears to be the same sort of “appeal to certain methodological criteria” to which Achinstein alludes. In some sense, claim 4 is a necessary condition on the possibility of positive philosophy of science regarding theoretical pursuit and as such I will assume that one goal of the current inquiry is to make explicit reasoning principles that would constitute a logic of pursuit. Given that it is often funding agencies that worry about which theories should be pursued, claim 5 is entirely appropriate. Strangely enough, claim 5 does not mention the benefits of research, which, as well as costs, are required for a complete decision theoretic analysis. To satisfy claim 6, the goal in examining the role of beauty in the logic of pursuit is to show how beauty is involved in general reasoning schema that do not necessarily have to do with a theory’s probability but whose conclusion concern the rationality of pursuing a given theory. As we shall see, I will endorse this claim: beauty can be relevant to the rationality of theoretical pursuit without influencing the probability of any given theory.

6.4 A Brief Survey of Views

In the remainder of this chapter, I will move from considerations of belief and means to considerations of action and goals. To begin, let me sample and classify the views of several influential thinkers: Herbert Feigl, Pierre Duhem, Henri Poincaré and Thomas Kuhn. Following the survey I will recast their positions within decision theory, a framework in which the logic of pursuit can be made more precise. To make decision theory relevant to the goals of science, one needs principles which link the goals of science to the utilities associated with theoretical pursuit. I will propose some candidate principles and draw distinctions among different approaches to uncovering such principles.
6.4.1 Herbert Feigl: Beauty has no Role

That beauty plays no role in science is perhaps the easiest position to describe since it involves no positive claims but rather merely the negation of all positive positions. It is difficult to find anyone defending stark aesthetic nihilism in science. Herbert Feigl comes close. He suggests that it is “utopian” to bridge Snow’s two cultures (Snow 1959).

Aesthetic commonalities between the arts and sciences occur only in the psychological (hence non-philosophical for Feigl) aspects of scientific and artistic creation (Feigl 1970, p. 9). If we take aesthetic judgments to be value judgments, as they often are, then logical positivism can be seen as a project that banished aesthetic value from philosophy of science and relegated discussion of aesthetics in science to the psychological realm (Putman 2002; Wartofsky 1993, p. 8). However, as I will discuss roles for beauty that would fall outside the scope logical positivist philosophy of science, even a logical positivist would not necessarily be an aesthetic nihilist, as it is understood here.

6.4.2 Pierre Duhem: Beauty is a Goal of Science

Is beauty one of the goals of science? Although I have not found someone who baldly makes this claim, many people have positions that implicitly assume that creation of beauty or some species of beauty is a goal of science. Let me mention a few different writers to show how they implicitly endorse beauty being a goal of science. Pierre Duhem, in his influential The Aim and Structure of Physical Theory, has been widely discussed as promoting anti-realism: the view that science does not, and should not, aim to discover truths about reality. However, a careful reading of his book reveals that his position is subtler. He thinks that physical theory “represents,” as simply, completely, and exactly as possible, a set of experimental laws and it is impossible not to think that the resulting “natural classification” of experimental laws corresponds to “real affinities among the things themselves” (Duhem [1914] 1954, pp. 19, 26). Nonetheless, the representation of experimental laws neither explains reality nor “conforms” to it. Furthermore, in a spirit
reminiscent of Hume, Duham claims that we cannot prove that nature possesses the relations suggested by the relations in the classification although we cannot help but think such connections exist. Thus, he does not think that the aim of theoretical science is truth in any realist sense, although the structure of physical theory may “hint” at the real affinities of things (Duham [1914] 1954, p. 30). Rather, he suggests that the goal of theoretical science is a beautiful orderly classification and an economical representation of otherwise disparate experimentally discovered physical laws.

Order wherever it reigns, brings beauty with it. Theory not only renders the group of physical laws it represents easier to handle, more convenient, and more useful, but also more beautiful.

It is impossible to follow the march of one of the great theories of physics, to see it unroll majestically its regular deductions starting from initial hypotheses, to see its consequences represent a multitude of experimental laws down to the small detail, without being charmed by the beauty of such a construction, without feeling keenly that such a creation of the human mind is truly a work of art (Duham [1914] 1954, p. 24).

Thus, for Duham, the orderly representation and classification that is theoretical physics is beautiful. It does not necessarily follow that Duham thinks that creation of beautiful theories is a goal of science, however. Let us distinguish two senses of goal of science. On the first sense, if X is the aim of science and X is co-extensive with Y, it does not follow that the aim of science is also Y. For example, if my goal is to kill an animal with kidneys, it does not follow that my goal, in this sense, is also to kill an animal with a liver. On the second sense of goal, if X is the aim of science and Y is coextensive with X then Y is also an aim of science. On the second sense of aim, Duham thinks that beauty is an aim of science. On the first sense, it is less clear. If it were only a contingent connection between order and beauty, one could argue that Duham believes the aim of theoretical science is not beauty necessarily but order. On the other hand, one could argue that for Duham theoretical beauty
just is theoretical order. Nonetheless, on either reading, for Duhem a beautiful physical theory realizes the goal of science to a greater extent than an ugly theory.

6.4.3 Richard Feynman: Beauty as Physical Laws

Duhem is not unique in his view. The physicist Richard Feynman expresses a similar sentiment about the aesthetic worth of physical laws:

The artists of the Renaissance said that man said that man’s main concern should be for man, and yet there are other things of interest in the world. Even the artists appreciate sunsets, and the ocean waves, and the march of the stars across the heavens. There is then some reason to talk of other things sometimes. As we look into these things we get an aesthetic pleasure from them directly on observation. There is also a rhythm and a pattern between the phenomena of nature which is not apparent to the eye, but only to the eye of analysis; and it is these patterns and rhythms which we call Physical Laws (Feynman 1965, p. 13).

To put Feynman’s claim baldly, physical laws, like sunsets and stars, are beautiful and cause scientists and others who possess the “eye of analysis” to feel aesthetic pleasure. If we assume that a goal of science is the discovery of laws and these laws are beautiful, then there is an important sense in which, for Feynman, science can be said to aim for beauty. The contemporary philosopher of science, Peter Lipton, endorses a related view. As discussed earlier, he defends the view that “loveliness is a guide to likeliness.” He also defends the view that in science loveliness has value in its own right (Lipton, personal communication). In some cases, Lipton holds that we should accept a less likely lovely explanation over a more likely ugly rival because we value more than merely truth. Although he does not put it this way, the view loveliness is one of goals of science coheres with Lipton’s more specific views on scientific inference.
6.4.4 Henri Poincaré: Beauty as Motivation

The goals or aims of science should have practical import. One reason for articulating goals is to provide reasons for acting in one way rather than another, that is, goals we adopt provide a reason for acting in a way which realizes the goals rather than not. Thus, if beauty were a goal of science, then, other things being equal, scientists would be rational to pursue beautiful theories. Indeed, it is not difficult to find scientists who say that what motivates them as scientists is that they are pursuing beauty. Typically, philosophers of science worry more about justification than motivation, although they are closely linked in the analysis of supposedly rational scientists. If one is rational, one expects that one’s motivations for a given action would also serve as justifications of that action. Henri Poincaré articulates the view that the search for beauty motivates (and thus justifies) his scientific research:

The scientist does not study nature because it is useful; He studies it because he delights in it; and he delights in it because it is beautiful. If nature were not beautiful, it would not be worth knowing, and life would not be worth living (Poincaré [1905] 1958, p. 8).

Indeed Poincaré thinks that all true scientists study nature because it is beautiful. A related position holds that scientific theories are beautiful (as opposed to nature itself) and their beauty motivates the scientist. The two positions are almost indistinguishable for those realists who think that a true theory of a beautiful nature will also be beautiful. Interestingly, Poincaré thinks that, as it turns out, beautiful theories are also the most useful. The reason for pursuing beauty for Poincaré is over-determined – it is both useful as well as delightful. He speculates that the success of Greek culture might be explained by its “strong” conception of intellectual beauty that allowed it to exterminate barbarian alternatives!

6.4.5 Thomas Kuhn: Beauty as Subjective Value

The final thinker I would like to consider is Thomas Kuhn. In his influential *Structure of Scientific Revolutions*, Kuhn claims that “aesthetic considerations” can lead
scientists to reject an old paradigm in favor of a new one (Kuhn 1970, p. 155). He says little about what he means by “aesthetic considerations,” but in his postscript, he gestures at judgments of simplicity, consistency, and plausibility (Kuhn 1970, p. 185). In a later article, Kuhn considers aesthetics in science more directly. On the relation between art and science, Kuhn writes:

Undoubtedly, …, considerations of symmetry, of simplicity and elegance in symbolic expression, and other forms of the mathematical aesthetic play important roles in [art and science]. But in the arts, the aesthetic is itself the goal of the work. In the sciences it is, at best, … a tool: a criterion of choice between theories that are in other respects comparable, or a guide to the imagination seeking a key to the solution of an intractable puzzle (Kuhn 1969, p. 342 italics mine).

Although Kuhn probably would not have put it this way, the quotation suggests two roles for aesthetics in science: one in the context of justification and another in the context of discovery. Further, as Margolis argues, it seems overly restrictive to confine these two roles to revolutionary science as Kuhn does (Margolis 1997, p. 193). It is the latter role in theory choice that is pertinent to our current discussion—how aesthetic considerations supposedly guide us to adopt one theory over another. Kuhn’s notion of theory “adoption” or theory “choice” is ambiguous. It can be interpreted to mean adopt-as-true or to mean adopt-as worthy-of-pursuit. Kuhn is usually interpreted as intending the latter interpretation.

As you might have inferred from the quotation from Kuhn, some properties responsible for a theory’s beauty are the same properties of theories or “theoretical values” that feature in standard lists of criteria for theory choice. For example, Kuhn mentions five criteria for “evaluating the adequacy of a theory”: accuracy, consistency, scope, simplicity, and fruitfulness (Kuhn 1973, p. 323). He acknowledges that the list is incomplete and that there are differences in how we interpret each one as well as further difficulties in assigning
relative weights to each. Presumably, we could add further criteria such as consilience, explanatory power, elegance, computational tractability, visualizibility, as well as more specific properties of theories. Some consider some of these properties as aesthetic, and Kuhn himself invites that interpretation in the quotations above. Others believe that many criteria of theory choice are clearly non-aesthetic. Not surprisingly, there is not universal agreement over which, if any, of these criteria for theory choice are aesthetic. Further, it does not follow that even if a theory is beautiful because of its simplicity, that all cases of simplicity are aesthetic. Complicating the picture, the relative weights we assign to these disparate criteria seem to depend on what we wish to do with the theory in question. That is, the value we assign to these criteria depends on the practical goals or aims we have. For example, if we need to solve an engineering problem, we care more about accuracy and computational tractability than scope and fruitfulness. If we are teaching students, we might value simplicity over accuracy. Nonetheless, Kuhn’s remarks suggest that at least some of these criteria are aesthetic, and additionally, the particular balance between them is an aesthetic choice.

For Kuhn, members of a scientific community share among other things a common set of values. Some of these values are aesthetic. For example, a specific interpretation of the meaning of simplicity might involve an appeal to what is beautiful. Say a scientist accepts one theory T over a competing T* because T is simpler than T*. Further she takes simplicity to be connected with beauty. Now it does not follow that she believes T, since for Kuhn, theory acceptance need not involve belief. Rather it is a commitment to the theory that is based on values that define the scientist as being a member of a particular scientific community. Now you might worry that values within scientific communities change over time and from person to person suggesting that values, as Kuhn conceives of them, are not objective. Kuhn is receptive to this possibility:

182 See Jones and Galison (1998) for several articles that discuss visualizibility in science.
… values may be shared by men who differ in their application. Judgments of accuracy are relatively, though not entirely, stable from one time to another and from one member to another in a particular group. But judgments of simplicity, consistency, plausibility, and so on often vary greatly from individual to individual (Kuhn 1970, p. 187). Aesthetic values are more likely to be of the latter more variable type of value and thus a pernicious subjectivism looms. Kuhn replies to this potential problem with the following two points:

First, shared values can be important determinants of group behavior even though the members of the group do not apply them in the same way … Men did not paint alike during the periods where representation was a primary value, but the developmental pattern of the plastic arts changed drastically when that value was abandoned. … Second, individual variability in the application of shared values may serve essential functions in science. … If all the members of a community responded to an anomaly as a source of crisis or embraced each new theory advanced by a colleague, science would cease (Kuhn 1970, p. 188). Kuhn’s response to the problem of subjectivity invites many more questions. What determines the limits of the variation in value application? Would science really cease if there were perfect agreement on the interpretation and application of values? Surely there could be disagreement over how to interpret the theories themselves (as opposed to merely disagreement over the criteria of choice). Such disagreement would provide an alternate way in which there could be differentiation within a community. But my purpose here is not to pass judgment on Kuhn’s philosophy of science. Let me however, mention, a further wrinkle in Kuhn’s view. Kuhn has been discussing intra-community variation in values; there is additional inter-community variation. Members of different paradigms may share wildly different conceptions of beauty. This is one reason that Kuhn does not talk in terms of “reasons for belief” when discussing theory choice since it is possible for scientists from different paradigms to adopt or choose different theories given the same evidence
based on the same values but differentially interpreted and/or weighted. Indeed they could make perfectly opposite decisions, i.e., it is possible on Kuhn’s view that for any two theories of a given phenomena, $T_x$ and $T_y$, if scientist $S_1$ from paradigm 1 adopts $T_x$ over $T_y$, then scientist $S_2$ from paradigm 2 adopts $T_y$ over $T_x$. For many philosophers who desire a more robust sense of objectivity, this result renders Kuhn’s position overly liberal. Nonetheless, Kuhn does illustrate how aesthetic value might play a role in theory acceptance, even if beauty is not a sign of truth and he takes seriously the variation in how aesthetic properties are weighted.

6.4.6 Taking Stock

Let us take stock. In this chapter I am broadening my focus to consider the role of beauty as a goal or an aim of science, and as a reason for action, namely whether beauty can serve as a reason for pursuing a theory. I presented a spectrum of positions. Herbert Feigl represents the view that beauty is not a goal of science. Duhem thinks that order is a goal of theoretical science and beauty necessarily accompanies order. Poincaré thinks that beauty is the reason that scientists do what they do, the usefulness of beautiful theories notwithstanding. Kuhn argues that differences in interpreting and weighting aesthetic values, at least in part, distinguishes scientists from different paradigms. For him it is possible that each scientist uses a differing set of values when choosing which theory to pursue. How should one respond to these positions? Clearly they do not exhaust the possibilities, but they do show a different ways in which beauty might play a role in the pursuit of science. Indeed the authors surveyed would appear to be answering quite different questions: Duhem asks what is wrong with English physics; Poincaré asks why pursue science; and Kuhn asks why accept one theory over another. My strategy will be to use decision theory to analyze the role of beauty in the logic of pursuit. I will develop a suggestion of Hempel (1965) who suggests we quantify the utility of the possible outcomes of research and treat acceptance as a question of decision theory (Hempel, [1960] 1965).
With this analysis in place, the relations among the thinkers will be clearer. Hopefully there will also be some illumination of decision theory itself.

6.5 A Decision Theoretic Analysis

In decision theory, the overarching normative principle is that one should act in such a way as to maximize one’s expected utility. I think it is an appropriate framework to consider the rationality of theoretical pursuit. To see how it is applied to the question of the practical relevance of beauty, consider a case where we are considering pursuit with respect to two mutually exclusive theories, one beautiful Tb and one ugly Tu. To keep the analysis relatively simple, let us assume that for each theory one has to choose simply between pursuing it or not.\(^\text{183}\) (A more general analysis might consider deciding the degree \(r_i\) to pursue each theory Ti.) The relevant magnitudes can be summarized in the following table:

---

\(^{183}\) Rota (1997) makes the point that, in mathematics at least, one does not directly pursue beauty if one hopes to create it. If such a view were correct of all of empirical science, it would undermine the goal of chapter 6. I assume that the choice is between pursuing two theories that have been partially developed such that we can determine that one is beautiful and one is not.
Tb is true & Tu is true & Both are false

<table>
<thead>
<tr>
<th></th>
<th>Tb is true</th>
<th>Tu is true</th>
<th>Both are false</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(Tb)</td>
<td>U(Pursue Tb) and not pursue Tu/Tb</td>
<td>U(Pursue Tb and not pursue Tu/Tu)</td>
<td>U(Pursue Tb and not pursue Tu/~Tu &amp; ~Tb)</td>
</tr>
<tr>
<td>Pursue Tb and not pursue Tu</td>
<td>P(Tb)</td>
<td>P(Tu)</td>
<td>P(~Tu &amp; ~Tb)</td>
</tr>
<tr>
<td>Pursue Tu and not pursue Tb</td>
<td>U(Pursue Tu) and not pursue Tb/Tb</td>
<td>U(Pursue Tu and not pursue Tb/Tu)</td>
<td>U(Pursue Tu and not pursue Tb/~Tu &amp; ~Tb)</td>
</tr>
<tr>
<td>Pursue Neither</td>
<td>U(Pursue neither/Tb)</td>
<td>U(Pursue neither/Tu)</td>
<td>U(Pursue neither/~Tu &amp; ~Tb)</td>
</tr>
<tr>
<td>Pursue Both</td>
<td>U(Pursue both/Tb)</td>
<td>U(Pursue both/Tu)</td>
<td>U(Pursue both/~Tu &amp; ~Tb)</td>
</tr>
</tbody>
</table>

Table 6.1 A Decision Theoretic Matrix for Theory Pursuit.

Here U(A/B) symbolizes the utility of doing A given B is true. There is no column for both Tb and Tu being true since we assumed that the two theories were mutually exclusive. The probabilities sum to unity: P(Tb) + P(Tu) + P(~Tu & ~Tb) = 1. The expected utility of an action is calculated by summing the utilities weighted by the probability that the appropriate state of affairs obtains. Thus, the expected value of pursuing Tb (and not pursuing Tu) consists of three terms:

\[ E(\text{Pursue Tb and not pursue Tu}) = P(\text{Tb})U(\text{Pursue Tb and not pursue Tu/Tb}) + P(\text{Tu})U(\text{Pursue Tb and not pursue Tu/Tu}) + P(\text{~Tu & ~Tb})U(\text{Pursue Tb and not pursue Tu/~Tu & ~Tb}) \]

Analogous equations hold for the three other possible actions. The most reasonable course of action has the highest expected utility. A more reasonable course of action has a greater expected utility than a less reasonable course of action. To calculate the maximal expected utility and hence the most reasonable course of action, one needs more information than merely the probabilities that are used in determining what to believe. In theory, one calculates the expected utilities of all the possible actions (i.e., the four actions above in our
case) and the action with the highest expected utility is the most reasonable one. Decision theory then shows a way in which beauty can be relevant to action, even if it is irrelevant to the probabilities of theories. If beauty influences the values of the utilities, it can influence the most reasonable course of action.

The use of the decision theoretic framework to illuminate the role of aesthetics makes a number of contestable assumptions. Perhaps the most significant one is that decision theory assumes that the expected net benefit of an action can be adequately represented by a single number. Thus pursuing a theory might create different types of value—create knowledge, make predictions, create a beautiful object, make possible technical progress, etc.—but decision theory assumes that these different types of value can somehow be summed. An opponent of this approach might argue that different sorts of value cannot be summed in this way. They could argue that the different sorts of value are incommensurable and there exists no one-dimensional measure that captures a “total” amount of value or utility. To put this same point differently: The utilities classified in the above matrix can each be thought of as consisting of benefits and costs: $U(A/B) = \text{benefits of } A \text{ given } B – \text{costs of } A \text{ given } B$. Presumably, the costs of pursuing a theory can be measured in dollars (although even this is contentious) and some of the benefits of having pursued that theory can also be measured in dollars, but does it make sense to place a monetary amount on the beauty of a theory? How would one go about determining a figure? It would be philistine to even make the attempt or so the opponent of this approach might argue. Additionally, they would argue that talk of “utiles” rather than dollars offers little improvement. In response, let me make a couple of points. First, we often do place monetary values on beautiful things. Consider an art auction. If we can value a beautiful product in monetary terms, it does not too much of a stretch to value the production of a beautiful theory. Second, often using decision theory, we do not need to assign definite numbers to the utility values in order to work out the most reasonable course of action. Under certain conditions just the relative differences in utility are sufficient to determine
which action is most reasonable. Finally, when we choose to act one way rather than another, we have, in effect, assigned relative weights to conflicting values. Although we might not be conscious of using any rules to make the choice, this unawareness does not mean that the action can be described as following inter-value balancing principles.

The decision theory framework presupposes that the distribution of alternative outcomes is irrelevant to determining which course of action to pursue. Since only the average expected value is calculated, the variance of the expected value plays no role in decision-making. In other words, the framework does not take into account different degrees of risk. For example action A might have the highest expected utility because there is a very small chance of a extremely large utility, even though there is a large chance of a low utility. Action B might have a lower expected return, but a high chance of a moderate return. In this case, decision theory says one should choose action A over B even when the most likely outcome will be that action A will yield less utility than action B. This choice might be justified by the claim that in the long run systematically choosing A over B will lead to a greater utility. However, the skeptic might reply that these long-term considerations should not dominate a short run choice. For example, one might argue that people are rational to buy insurance even if the expected utility is negative (otherwise the insurance company would not make money). A decision theorist might take into account a risk averse person by modifying the utility values, but whether this completely captures the complaint is controversial. Even modified utility values have distributions over possible outcomes and these are not taken into account.

6.5.1 Goals and Means in Decision Theory

The final assumption of decision theory that I would like to consider relates to goals and means. The framework appears to collapse the distinction between different types of goals and, in part, the distinction between means and ends. At the beginning of the chapter, I distinguished between the position that beauty was a goal of science and the position that
beauty was a means to another goal of science such as predictive accuracy. On either of these views, possession of beauty would increase the expected utility of pursuing a theory, other things being equal. However, it is difficult and in certain cases impossible to use this framework to distinguish between the two positions. It could be that promoting another goal gives a pursuit of a beautiful theory more utility than the utility associated with the goal of beauty, or vice versa. The relative amounts of utility gained by being a goal or a means to a goal depends as much, or more, on the degree to which goals are met as the difference between goals and means. Thus, unless we supplement this framework with a further principle, we cannot use it to decide between whether beauty is a goal in its own right or merely a means to another goal. A weak principle can be formulated to distinguish between goals and means:

If of two properties G and M, one of them is a goal and the other is only a means to that goal and \( U(\text{Pursue T/Pursuing T achieves } G \text{ and } M) = U(\text{Pursue T/Pursuing T achieves } G) > U(\text{Pursue T/Pursuing T achieves } M) \), then G is the goal and M is the means to that goal.

In cases of alleged counter examples, such as \( M_1 \) is a means to goal \( G_1 \), but pursuing \( M_1 \) is more valuable than \( G_1 \), the principle dictates that the alleged means is actually a goal (in addition to being a means). On my view, a mere means cannot be more valuable than the goal it realizes. Nonetheless, the scope of this principle is limited. It is of no help when goals are only partially met— in which case what is needed is a utility function that takes purported goals and the degrees to which they are met as arguments. Unfortunately, there is little hope of realistic general utility functions being simply represented. Furthermore, we cannot use the decision theory framework alone to determine which goal beauty furthers, if any. Chapters 4 and 5 considered whether beauty was a means to truth – here beauty influences the probabilities, but when we consider other goals of science that are independent of the probabilities then beauty will raise the benefit of pursuing a theory. On the other hand, decision theorists will argue that it is a strength of decision theory that it
allows for the achievement of different goals all to count towards the utility of pursuing a theory. I will return to the question of goals and means to goals below.

6.5.2 Beauty and Decision Theory

From the framework of decision theory, whether there is more reason to pursue a beautiful theory depends on the values of the utilities arrayed in the above matrix and the various probabilities of the different theories. We have discussed the relation of probability and beauty in Chapters 4 and 5. Let us turn then to a consideration of the relative values of the various utilities. Each utility can be described as a net benefit: U(A/B) = B(A/B) – C(A/B) = benefits of A given B – costs of A given B. A natural assumption would be that the costs of pursuing a theory are independent of its truth, i.e., C(Pursuing T/ T) = C(Pursuing T/ ~T). Generalizing we might claim that the costs of pursuing a particular theory are the same across the outcome partition, i.e., across a row in the above table. Differences in the utilities across a row then largely depend upon the differences in benefit. This is not too surprising since most would expect that the utility of pursuing a theory would be greatly increased if the theory were true. What then of comparisons of the utility values in columns? This is a comparison of the different options of pursuit given that we know the truth-value of at least one of the theories in question. Before getting to the crux of the matter—the comparison between the utilities of pursuing a beautiful theory Tb and pursuing an ugly competitor Tu—let us look at whether we can say anything about the composition of utilities. Are there any constraints on U(Pursue both Tb and Tu/X) given the values of U(Pursue Tb and not pursue Tu/X) and U(Pursue Tu and not pursue Tb/X)? It is not clear, since costs and benefits appear to pull in different directions. Clearly it is very likely that C(Pursue both Tb and Tu/X) > C(Pursue Tb and not pursue Tu/X) and that C(Pursue both Tb and Tu/X) > C(Pursue Tu and not pursue Tb/X) since we would expect the cost of pursuing both theories will be higher than the cost of pursuing one of them. Given considerations about economies of scale and the idea that some of the work pursuing
Tb will make pursuing Tu easier, a case could me make for \( C(\text{Pursue both}/X) < C(\text{Pursue Tu and not pursue Tb}/X) + C(\text{Pursue Tb and not pursue Tu}/X) \). On the benefit side, an analogous argument could be made: some benefit is gained by pursuing Tb or by pursuing Tu and these benefits are not simply summed if you pursue both. Thus the following inequality is likely to hold: \( B(\text{Pursue both}/X) < B(\text{Pursue Tu and not pursue Tb}/X) + B(\text{Pursue Tb and not pursue Tu}/X) \). Unfortunately since these inequalities show that both relative benefits and relative costs decrease when we combine pursuits, but nothing about the relative amounts of decrease, little follows about the resulting utilities.

Given that it is quite likely that \( U(\text{pursue both}) > U(\text{Pursue Tb}) \) and that \( U(\text{Pursue both}) > U(\text{Pursue Tu}) \), decision theory would seem to say one should pursue both theories, ugly and beautiful, and thus have little to say about the choice between the two theories. However, this is not guaranteed to happen especially if pursuing one of the theories has a negative expected utility. Furthermore, we usually have a budget constraint and even if \( U(\text{Pursue both}) \) is maximal, since we have to pay costs before we reap benefits, pursuing both might not be a financially viable option. Let us then turn to the question of whether pursuing Tb or pursuing Tu is more reasonable. Within this framework the question of whether Tb is more reasonable than Tu becomes, is \( E(\text{Pursue Tb and not pursue Tu}) > E(\text{Pursue Tu and not pursue Tb}) \)? In other terms, the question becomes: is the following inequality true:

\[
P(\text{Tb})U(\text{Pursue Tb and not pursue Tu}/\text{Tb}) + P(\text{Tu})U(\text{Pursue Tb and not pursue Tu}/\text{Tu}) + P(\sim\text{Tu}\&\sim\text{Tb})U(\text{Pursue Tb and not pursue Tu}/\sim\text{Tu}\&\sim\text{Tb}) > P(\text{Tb})U(\text{Pursue Tu and not pursue Tb}/\text{Tb}) + P(\text{Tu})U(\text{Pursue Tu and not pursue Tb}/\text{Tu}) + P(\sim\text{Tu}\&\sim\text{Tb})U(\text{Pursue Tb and not pursue Tu}/\sim\text{Tu}\&\sim\text{Tb}).
\]

Naturally, one might wonder if this decision theoretic approach can yield much insight as one now has a cumbersome equation with 9 unknowns (or 8 if we eliminate one probability term with the knowledge that the probabilities sum to 1). The equation is, however, useful in classifying a number of positions and a number of different reasons for adopting the
aesthete’s position. The aesthete considered in Chapter 4, who argues that beauty is a sign, of truth, might argue that it is more reasonable to pursue Tb because the differences in probabilities (p(Tb) > p(Tu)) swamp opposing differences. But there are many other positions a lover of beauty might occupy. Moving to the utilities, a number of positions emerge. At the most extreme, it might be argued that the utility of pursuing a beautiful theory is always greater than pursuing an ugly one no matter which theory turns out to be true. That is, the most radical decision theory aesthete might hold the following three claims:

1. \( U(\text{Pursue } Tb \text{ and not pursue } Tu/Tb) > U(\text{Pursue } Tu \text{ and not pursue } Tb/Tb) \)
2. \( U(\text{Pursue } Tb \text{ and not pursue } Tu/Tu) > U(\text{Pursue } Tu \text{ and not pursue } Tb/Tu) \)
3. \( U(\text{Pursue } Tb \text{ and not pursue } Tu/\neg Tu \& \neg Tb) > U(\text{Pursue } Tu \text{ and not pursue } Tb/\neg Tu \& \neg Tb). \)

A radical aesthete who held that beauty was the only goal of science would defend all three claims. A more moderate aesthete might only defend one or two of them. The most defensible part of the radical decision theory aesthete’s position is the claim that \( U(\text{Pursue } Tb \text{ and not pursue } Tu/Tb) > U(\text{Pursue } Tu \text{ and not pursue } Tb/Tb). \) In this case, by pursuing Tb, we are attempting to create a true and beautiful theory instead of an ugly false theory. It is highly likely that pursuing a true and beautiful theory yields more utility than an ugly and false one. On the other hand, the true beautiful theory might be so difficult to develop and wield in practice that more utility can be gained by having pursued an ugly false approximation. Nonetheless, ceteris paribus, the utility of having pursued a true beautiful theory is greater than having pursued an ugly false one. The least defensible part of the aesthete’s position is the claim that \( U(\text{Pursue } Tb \text{ and not pursue } Tu/Tu) > U(\text{Pursue } Tu \text{ and not pursue } Tb/Tu). \) Clearly if Tu is true then there will be a significant amount of utility due to have pursued (and thus possibly created) a true theory. The only way for the extreme aesthete to defend this point is to argue that the beauty resulting from pursuing Tb outweighs the value of having pursued the true theory Tu. For example, an aesthete might
argue that it is better to pursue a beautiful ideal law (like the ideal gas law) than an ugly complex true law.\textsuperscript{184} Again, if beauty is a more important goal than truth or even if pursuing Tb rather than Tu satisfies more of the goals of science to a greater degree, the aesthete’s position would be justified. Most plausibly this occurs when the beautiful theory while false approximates truth. The pertinent questions then become, how do we determine the goals of science, their relative priority, and the degree to which pursuing a beautiful theory, which happens to be false, satisfies them? These questions become more pressing when we consider the third part of the radical aesthete’s claim: that U(Pursue Tb and not pursue Tu/~Tu&~Tb) > U(Pursue Tu and not pursue Tb/~Tu&~Tb). Here the aesthete claims that there is more utility in pursuing the more beautiful of two false alternatives. She could, for example, justify this claim by appeal to the fact that having pursued a beautiful theory yields aesthetic pleasure for the researchers and could yield a beautiful product for society’s contemplation, even though that product would be false. To refocus the issue, how do we decide whether aesthetic pleasure and value should contribute to the utilities and to what degree?

6.5.3 Two Decisions: by the Society and by the Individual

The question of inter-value balance is a difficult one. I suggest that to make some progress we separate at least two distinct projects that so far have been conflated. First we can consider the above matrix as representing the decision parameters of a single scientist considering which theory to work on. Second, we can consider the above matrix as representing the decision parameters of a larger group of people, at the extreme, society as a whole. Thus, when issues of how society should fund research are used to justify study of “the logic of pursuit,” these two distinct projects risk being conflated. I will consider the case of society from now on to illustrate the group decision problem. At the national level,

\textsuperscript{184} Cartwright (1983) argues along these lines, although she does not connect her position to aesthetics.
national funding agencies make these decisions on behalf of the society. The utility values for a single scientist and for society need not be the same. In fact, we would expect them to differ because we would expect different types of consequences of theoretical pursuit to be relevant to different degrees and therefore the utilities for society and the individual scientist to be different. Of course, this does not mean that many of the consequences valued by the scientist will not be valued by society too. For example, we would expect that the aesthetic pleasure that a scientist gains by pursuing a theory to be weighted more highly by the individual scientist pursuing the theory, than society at large. I don’t think this is too surprising. On the other hand, contemplation of the beautiful theories of science could be argued to be a social good that funding agencies should take into consideration. Nonetheless, it would be an unlikely coincidence if the degree of value attributed to pursuit of a beautiful theory were the same for the person pursuing the theory and society at large. Each project begs the question of which consequences of pursuit are relevant and should be taken into account when determining the correct utility values.

6.5.4 Inter-value Realism and Inter-value Anti-realism

On either the individualistic or social project the crux of the decision theoretic rendition turns on the relative amounts of the relevant utilities. I am not going to provide a mechanism for generating the correct utilities. Indeed I would be suspicious if anyone claimed to have one. Nonetheless, there is clearly more to be said on the issue. Philosophers might be grouped into two groups according to how they approach this problem. First, there are the inter-value realists. They hold that the correct balance among different types of values is independent of what we take the balance to be. In other words, there is a correct balance to be discovered and what ever this balance is, it does not change if we change what we take it to be. Opposed to the inter-value realists are the inter-value anti-realists. They hold that the correct balance among different types of value depends, at least in part, on what we take them to be. They would argue that the facts needed to justify the
realists’ balance do not exist. For them, there is no correct balance to be discovered; rather it has to be invented by us. The distinction between inter-value realists and anti-realists is vague insofar as the referent of “we” in the above distinction is vague. To be more precise, let “we” refer to the individual in question in the individual decision problem and refer to society in the societal decision problem. Notice that it is not inconsistent to be a realist with respect to one of these decision problems and not with the other. The instrumental rationality that is usually assumed by decision theorist is neutral on the inter-value realism /antirealism issue. Stripped down decision theory does not provide any constraints on the utility values, nor a way to determine them. Rather goals in the form of utilities are taken as given, and action is judged reasonable relative to the utilities. This is true of both the individual and the social decision problem.

6.5.5 Goals as Constraints on Utility Values

Naturally, we might want to supplement the stripped down decision theory with constraints upon utility values and thus give a more substantive view of practical reason in science. As hinted at above, I suggest that a good way to consider how build in constraints into the decision theoretic account is to look at the goals of science. Consider the following fairly weak principle:

Ceteris Paribus, if beauty is a goal of science and Tb is beautiful and Tu is not, then

\[ U(\text{pursue } Tb \text{ and not } Tu/X) > U(\text{pursue } Tu \text{ and not pursue } Tb/X), \]

where X is anything that does not violate the ceteris paribus clause.

A case where X does not violate the ceteris paribus clause is when Tb and Tu are both false (or true). As indicated above, we can consider this principle from both an individual and a societal perspective. Further progress is more controversial. Two fronts are open. First, we

---

185 It is possible that utility values for an individual scientist might be a function of the utility values of other scientists. Some scientists for example gain utility from being unconventional. This point is neutral
might consider whether beauty is in fact a goal of science and how one could determine whether it is. Second, we might consider if there are any further principles that hold when things are not equal. For example, is it ever more reasonable to work on Tb when Tu is true or we know Tu is true? These two projects are not unrelated. An aesthete who thinks that beauty is the only goal of science—let us call her a pure aesthete—may argue that, no matter what other differences exist between Tu and Tb, it is more reasonable to pursue Tb. The same could be said for the less pure aesthete who claims that beauty is the dominant goal, among many, of science. In other words, on this view, a theory with more beauty trumps those with less, no matter what non-aesthetic virtues the two theories have. On this somewhat crazy view, when two theories have equal degrees of beauty, one considers the next most prominent (non-aesthetic) goal to adjudicate the tie at the aesthetic level. A pure aesthete has a more difficult time defending her view at the individual level. Here it is more plausible that an individual’s goals and their priority change depending on context. If one is writing an introductory text, you might prioritize the goal of beauty and simplicity over accuracy. If one is making a real-life engineering calculation you probably value accuracy over beauty and simplicity.

On the face of it, you might think that the question of the goals of science could be settled by empirical methods. The actual practice of science, in so far as it has goals, has either implicit or explicit goals. If science has explicit goals, they can be known simply by consulting with the explicators. Of course, science as a whole does not have explicit goals; it is only philosophers of science who try to make them explicit. Even for them you might think that making implicit goals explicit is an empirical descriptive project. However, philosophers of science have not treated them this way. For example, van Fraassen’s construal of the realism/anti-realism debate is in terms of the aims of science. For van Fraassen the philosophical debate is over whether science aims at a literally true theory of

between what I am calling inter-value realism and inter-value antirealism.
the world or merely at empirically adequate theories. This debate is not to be settled with an
empirical survey or at least this is not the way van Fraassen and his interlocutors precede. If
it were merely an empirical question, we could design the appropriate survey and settle the
matter without the need for philosophical articles and books. On the contrary, van
Fraassen’s anti-realism provoked a large philosophical literature and renewed interest in
realism and antirealism as a philosophical issue. Perhaps the realism/antirealism debate is
better described as a debate over what the aim of science should be. A similar point could
be made over whether beauty is a goal of science: it is more a question of what the goals of
science should be rather than what they currently are.

6.5.6  Ideal Deliberation: Individual Inter-Value Anti-Realism

How then do we determine what the goals of science should be? If the reader will
allow me to some latitude to speculate, I would like to sketch how an inter-value anti-realist
could answer the question, at least from the societal perspective. I don’t advocate this
approach as the correct answer to the question, but rather provide this rather lofty sketch to
show one way in which the goals of science might be determined. Philip Kitcher’s
discussion of a similar issue can be appropriated for use in this project (Kitcher 2001). He
aims to characterize what he calls “well-ordered science.” The goal of the first phase of a
three-phase procedure realizing a well-ordered science consists of decisions “to commit
resources, such as investigators and equipment, in particular amounts to particular projects”
(Kitcher 2001, p. 118). The second phase “pursues those projects in the most efficient
way, subject to moral constraints that rule out certain physically possible options.” The
third phase involves turning the various investigations into practical consequences. One can
use Kitcher’s apparatus for a different purpose. To achieve the first phase, Kitcher
imagines that society consists of ideal deliberators whose preferences take into account the
needs of the others, the significance attached to each possible avenue of inquiry, and the
needs of long term inquiry. The process of deliberation either leads to a consensus on
which projects each deliberator thinks society should pursue or acceptance of a “fair” ordering of projects, or if consensus is not achieved, a democratic vote on which ordering to adopt. Instead of producing an ordering of projects, one could imagine that Kitcher’s society of ideal deliberators be employed to determine an ordering of goals for science and an associated ordering of the different values in science. Rather than have deliberation occur at the level of projects to pursue, they could deliberate about what society should value in the products of our science, i.e., what properties of scientific theories we should value and to what degree. I call this an anti-realist approach to the question of goals in science because if the deliberators came to a different consensus, the goals would also change.\textsuperscript{186} Whether beauty is a goal of science and its relative priority then would be one result of such a deliberation. Thus we would have constraints upon the utilities for the decision theoretic problem at least for the societal perspective. It remains a debatable question whether an individual scientist is rationally bound by these constraints or whether they have the freedom to deviate from what society says they should value as scientists.

6.6 Poincaré, Duhem, and Kuhn Reconsidered

Keeping in mind the decision theoretic approach, the distinction between an individualistic and societal perspective, and the distinction between a realist and an anti-realist view of utilities, let us return to the views of Poincaré, Duhem, and Kuhn. Can their views be captured within decision theory? Can we categorize them neatly with the distinctions I have drawn? It is to these questions that we now turn.

\textsuperscript{186} If the deliberators could not come to a consensus about which values should have priority, then a form of pluralism might emerge. Instead of a single ordering of values, the product of a dead locked deliberation might be a number of orderings that are weighted in proportion to the number of deliberators that endorse each ordering. These weightings then would be relevant in determining the proportion of beautiful projects that are supported by society.
6.6.1 Duhem and Decision Theory

Let us begin with Duhem. He thinks that beauty accompanies the order that scientists should aim for in constructing their theories. That order is a goal of theoretical science does not depend upon what we take the goals of science to be. Neither do the utilities associated with pursuing a given theory. Duhem then, on my account, is an inter-value realist. In fact, he is an inter-value realist at both the individual and the societal level. His position is most apparent when Duhem discusses the imaginative yet shallow “weak-minded” English scientist. He chastises English scientists, notably Sir William Thomson and James Clerk Maxwell, for not providing an ordered theoretical science but a series of disconnected models. He clearly thinks that if a scientist had to choose between pursuing an abstract beautiful “French” theory and a shallow ugly “English” one, she should pursue the French one. This truth is independent of whether we take the utility of the English type of theory to be superior as the English scientists he mentions did. In other words, Duhem thinks that the English scientists are wrong to think that their way of doing things is better. Science, as a whole, and individual scientists in particular should emulate the French scientist. To put Duhem’ position in the language of decision theory, he thinks that \( E(\text{pursuing } Tb \text{ and not } Tu) > E(\text{pursuing } Tu \text{ and not } Tb) \) in large part because \( Tu \) does not possess the order that characterizes good theories. Can we attribute to Duhem views about the utility values we find in the decision matrix? This attribution is made difficult because the various values are predicated on different truth-values of \( Tb \) and \( Tu \). As mentioned above, standardly interpreted, Duhem does not believe that theories are true or false, in the sense that they correspond with the world. Let us assume that where we consider the probability of certain theories, Duhem understands it to mean the probability of it containing “natural affinities,” to use his language. Unfortunately, he does not define exactly what he means by the term, “natural affinity,” but he seems to mean that natural phenomena are properly classified according to which laws mature physics says governs them. Duhem, I think, would argue that a theory without order is impossible to show
natural affinities.” If this is so, then the utility associated with the pursuit of a true ugly theory is irrelevant to practical reason, i.e., $U(\text{Pursue } Tu \text{ and not pursue } Tb/Tu)$ multiplied by $0 = 0$. Indeed, at root Duhem’s desire to pursue theories with order and beauty seems to be driven by a metaphysical belief that the world has “ontological order” and this order can only be captured by an orderly physical theory and thus $p(Tb) > p(Tu)$. Duhem is quite conservative in this regard.

Nonetheless, if we assume that $p(Tu) > 0$, and if decision theory can be applied to Duhem’s position, then we need values of pursuit given that $Tu$ is true. Duhem does consider whether the ugly English models are fruitful for discoveries (Duhem [1914] 1954, ff. 93). After a brief historical survey, he concludes that, “the share of the booty it [the use of ugly mechanical models] has poured into the bulk of our knowledge seems quite meager when we compare it with the opulent conquests of the [French-style] abstract theories.” (Duhem [1914] 1954, p. 99) In fact, although there have been some discoveries with the use of models, their use of these models may even “obscure discoveries already made.” Furthermore, the beautiful abstract theory has extra utility of being an orderly classification and the extra utility due to the “esthetic emotion” that it produces (Duhem [1914] 1954, p. 24). In the language of decision theory, Duhem claims $U(\text{Pursue } Tb \text{ and not pursue } Tu) >> U(\text{Pursue } Tu \text{ and not pursue } Tb)$ and perhaps even $U(\text{Pursue } Tb \text{ and not pursue } Tu/Tu) > U(\text{Pursue } Tu \text{ and not pursue } Tb/Tu)$.

6.6.2 Poincaré Reconsidered

Poincaré’s position is closely related to Duhem’s. He too thinks that nature is beautiful. He too thinks that a beautiful theory has value. Its value is partially due to being beautiful and partially due the coincidence that beautiful theories tend to also be the most useful. From what little he says, Poincaré is also a realist about the relative value of beauty. Recall that Poincaré claims: “The scientist does not study nature because it is useful; He studies it because he delights in it; and he delights in it because it is beautiful. If nature
were not beautiful, it would not be worth knowing, and life would not be worth living.”
(Poincaré [1905] 1958, p. 8) The subject of the first sentence “the scientist” applies to both
a given individual scientist and all scientists, and for this reason, Poincaré is a realist at both
the individual and societal levels. Furthermore, beauty is clearly the primary value. It makes
it worthwhile to study nature.

What would Poincaré say of an individual scientist who delights in studying
ugliness and who, ceteris paribus, chooses to pursue ugly theories over beautiful
alternatives? Let us call this person the anti-aesthete. Would Poincaré deny that he would
be a true scientist or that he could be acting rationally as a scientist? Duhem comes close to
this position when he excoriates the English obsession with ugly mechanical models. The
response to this question is a touchstone of inter-value realism at the individual level. The
inter-value realist who values beauty (and not ugliness) denies that the anti-aesthete can
reasonably pursue ugliness over beauty, other things being equal. On the other hand, the
inter-value anti-realist allows an individual more freedom to balance different types of value
in a way the scientist sees fit. For example, if pursuit of ugliness significantly contributes to
the scientist’s utility, then it is possible that pursuing an ugly theory is more reasonable than
a beautiful alternative, other things being equal. Poincaré rejects this inter-value anti-
realism. He would argue that the anti-aesthete does not have a good reason to pursue an
ugly one over a beautiful one. Although he does not discuss the issue in the form I have, he
does argue that if there were no simple (beautiful) facts, then science would not be possible
at all (Poincaré [1905] 1958, p. 5). It is a closely related position to say that pursuing
“dreadful” complexity is not science at all. A (true) scientist does not count the
“capricious variations” in ladybug populations, says Poincaré, echoing Tolstoy. The
essential preference for simplicity and beauty explains why scientists examine the very
small and the very large—this is where simple facts can be found. I think Poincaré’s
attitude expresses a chauvinism formed by over-emphasis on physics and mathematics. Nonetheless, Poincaré and Duhem represent inter-value realism at the individual and societal levels.

6.6.3 Kuhn Reconsidered

In the light of Duhem and Poincaré, Kuhn is a radical. He is an inter-value anti-realist, both at the individual and societal levels. He argues that the balance between different values varies across time and among scientists. This variation in values promotes efficient science as diversity and competition between different perspectives drives science forward, or so Kuhn argues. How an individual scientist (correctly) balances the different types of value is in part dependent on the importance that scientist places on the values. Kuhn argues that “features of individual personality and biography” explain much of the variation among scientists (Kuhn 1970, p. 185). I would argue that it is because we value freedom of inquiry that we as a society allow scientists a certain degree of freedom in how they balance different values. Kuhn, on the other hand, stresses how the passive process of “professional initiation” molds an individual’s particular balance of values and explains the commonalities among scientists. The scientist is constrained to some degree by her membership in a paradigm: to remain within her paradigm, her balance cannot vary significantly from the paradigm average, but nonetheless, there can be personal differences among different members of the same paradigm. At the societal level, Kuhn does not attribute science as a whole with goals beyond the values that all scientists share. Presumably these values are quite weak as any interesting value has associated with it disagreement that fractures the scientific community into more specific scientific communities. He does, however, think that paradigms are constituted by largely shared

---

187 Barrow (1999) argues that a focus on simplicity, symmetry and physics characterizes a “Platonic” approach to science and a focus on complexity, asymmetry, and biology characterizes an “Aristotelian”
values. The particular balance and interpretation of the values constituting a paradigm are not determined top-down, but rather emerge as a rough consensus among member scientists. Kuhn’s position can be captured within decision theory very well. Each scientist’s set of values applied to a particular decision is represented by a set of utility values and the paradigm values can be represented by the mean and standard deviation of the population. Typically we would expect members from a single paradigm to largely agree over which theory should be pursued.

Kuhn can accommodate the example of the anti-aesthete within his framework. In fact, there is nothing within his system that prohibits a community of anti-aesthetes constituting a paradigm. Some may argue that this is a weakness of his system—that it is too loose to rule out cases such as this. Indeed, in his 1970 postscript, Kuhn adamantly defends his position from charges of irrationality made by Lakatos and others. A charge of irrationality would be valid if scientists could give up a value like consistency perhaps, but here we are discussing how to react to aesthetic features of our theories of the world. In fact, we could make the case on behalf of Kuhn that if every scientist choose to work on the beautiful theory and not the ugly alternative that this would bias scientific practice as a whole and delay us developing a true ugly theory, if in fact the world is best described by an ugly theory. That a small percentage of scientists do concern themselves with counting ladybug populations should be commended.

6.7 Beauty and The Freedom of Inquiry: A Proposal

Let me end by sketching the beginnings of an inter-value antirealist view at the individual level. As a society we allow individual scientists the freedom to formulate and prioritize their own values. There are some constraints that are often codified by legal and ethical imperatives, but generally society does not dictate methodological values to individual approach to science. On my view, Poincaré emphasizes the former at the expense of the latter.
scientists and neither should it. Each scientist is given the freedom to choose the exact balance between different kinds of value and to use this balance to guide his or her research program. My position is clearly inter-value antirealist as what the “correct” balance between values for an individual scientist is determined by what the individual scientist takes them to be. Of course, there are incentives for an individual scientist not to choose a balance too far from their community norm if she wishes to be a successful member of that scientific community. Two ways justifying this attitude toward its scientists are open. One could argue that granting such freedom to scientists leads to efficient science or one could argue that the value of the freedom granted is its own justification. This former type of justification lies behind Paul Feyerabend’s famous crusade against fixed rules of methodology (Feyerabend 1975; Lloyd 1996). Although he is concerned more with the free choice of methodological rules, Feyerabend’s writings also support the idea freedom of choice of value. The latter option makes freedom of inquiry, in the sense of choice of value as well as choice of project, a goal of science.

6.7.1 Freedom of Inquiry and Decision Theory

Beyond being an inter-value antirealist position, how does this view fit into the decision theoretic framework I presented earlier in this chapter? There are some difficulties in representing the value of freedom with in the decision theoretic framework. On the other hand there is complementary also. For decision theory to be useful, it would have to presuppose that the decider is free to choose any of the options represented (For a related point, see Sen (2002) p. 593). If there were no freedom, then it would be futile to calculate the most rational course of action, because it could not be chosen. I am however proposing a more radical freedom, not the freedom to choose what theory to pursue, but rather to choice which properties of a theory are valuable and to what degree. More specifically I suggest that individual scientists choose the degree to which they value beautiful properties of theories. Thus theoretical pursuit is a two stage process: first choose the properties that
you value; then apply these values to select the theory to pursue. The second stage is simply an application of decision theory as I have presented it earlier. An alternative way of conceiving of the freedom of inquiry is more radical. It says that dictates of decision theory are irrelevant – one should not be bound to choose the theory whose pursuit maximizes expected utility. These constraints impede freedom and should be disregarded. On this alternative, to have freedom of inquiry we free the inquirers from the constraints of practical rationality as described here. I am not endorsing this view. I think decision theory is a useful tool in determining what to do once the relevant utilities (values) have been chosen.

This is not to say, however, that decision theory gives a complete picture. If we turn our attention to how to use decision theory to decide what properties of theories to value with high utilities, we confront the possibility of an infinite regress. In other words, to use decision theory to choose which properties of theories to value, we need utility values for adopting these utility values and we merely push the problem back one level. How do we determine these 2nd order utility values? Decision theory requires that we need 3rd order utility values. And the infinite regress takes hold.

There are additional problems applying decision theory to the choice of utility values. Consider the case where we choose between two inequalities: (1) $U(T_b/X) > U(T_u/X)$ and (2) $U(T_b/X) < U(T_u/X)$. By analogy with the first order decision, the relevant second order decision matrix might look like this:
<table>
<thead>
<tr>
<th></th>
<th>$U(T_b/X) &gt; (T_u/X)$ is true</th>
<th>$U(T_b/X) &lt; U(T_b/X)$ is true</th>
<th>Both are false</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p(U(T_b/X) &gt; U(T_b/X))$</td>
<td></td>
<td>$p(U(T_b/X) &lt; U(T_b/X))$</td>
<td>$p(Both$ $are$ $false)$</td>
</tr>
<tr>
<td>Adopt $U(T_b/X) &lt; U(T_b/X)$</td>
<td>$U(Adopt$ $U(T_b/X) &lt; U(T_b/X)$</td>
<td>$U(Adopt$ $U(T_b/X) &lt; U(T_b/X)$</td>
<td>$U(Adopt$ $U(T_b/X) &lt; U(T_b/X)$</td>
</tr>
<tr>
<td>Adopt $U(T_b/X) &gt; U(T_b/X)$</td>
<td>$U(Adopt$ $U(T_b/X) &gt; U(T_b/X)$</td>
<td>$U(Adopt$ $U(T_b/X) &gt; U(T_b/X)$</td>
<td>$U(Adopt$ $U(T_b/X) &gt; U(T_b/X)$</td>
</tr>
<tr>
<td>Adopt Neither</td>
<td>$U(Adopt$ $neither/U(T_b/X) &gt; (T_u/X))$</td>
<td>$U(Adopt$ $neither/U(T_b/X) &gt; (T_u/X))$</td>
<td>$U(Adopt$ $neither/Both$ $are$ $false)$</td>
</tr>
</tbody>
</table>

Table 6.2  A Decision Theoretic Matrix for Utility Adoption.

However, this decision matrix borders on nonsensical, especially for the inter-value antirealist. What does $p(U(T_b/X) > U(T_u/X))$ mean? If one is an inter-value realist, it means the probability that the value of utility pursuing $T_b$ is greater than pursuing $T_u$. (I have ignored the pursuit operators in the above table to make the equations less unwieldy.) For the inter-value realist this probability is defined independently of what the individual scientist takes it to be. For the inter-value antirealist, however, this probability is not independent of what the scientist takes the utilities to be. Indeed, what the scientist takes the values to be just is what the decision problem is trying to solve. Thus, for the inter-value antirealist we have a circularity in determining the probability values for the above decision problem: the decision depends on the probability values, which in turn depend upon the decision.

Thus freedom to choose utility values cannot be entirely governed by decision theory. There must also be what might be called the spontaneity of individual values. That a particular scientist chooses to pursue a beautiful theory over an ugly theory (because she
has chosen to value beauty over ugliness) is a right that we as a society grant. No doubt that there are empirical connections between certain theoretical properties that are candidates for values and practical outcomes that society values. These connections are exploited by funding agencies on our behalf to promote biases in research that hopefully lead to more valuable outcomes for future society. However, many of the ultimate theoretical values are neither fixed by society nor by reason, and we add to their value by allowing individual scientists to freely choose them (See Sen 2002, Chapters 20 and 21, for more on the value of freedom of process.) A controversial consequence of my position is that it would not be irrational for an individual scientist (the anti-aesthete) to pursue the uglier alternative, if her choice of values dictated this course of action. In fact, Kuhn might be right that many scientists’ values are fixed by their professional history and context, but scientists are not entirely passive. They have the ability to choose the value of beauty and thus how to react to it. An anti-aesthete, although rare, is not necessarily irrational.

To return to the opening question of this chapter: Is beauty a goal of research? Not necessarily. There is no global answer. However, locally many scientists choose it to be one of their implicit goals. Ceteris paribus, they would rationally pursue a beautiful theory over an ugly rival. This is true even if beauty does not lead to any more practical benefits. For them, the value of beauty exemplifies their right to the freedom of inquiry. On the flip side of the coin, I suggested that an anti-aesthete who reverses the typical aesthetic ordering also illustrates society’s willingness not to restrict the freedom of inquiry. Feyerabend argues that this pluralism is good for inquiry and society as a whole. I have suggested that it also could be justified by the value of freedom itself.

6.8 Summary and Conclusion.

If beauty were irrelevant to a theory’s probability, would beauty be irrelevant to the practice of theoretical science? No, not necessarily. Decision theory shows how aesthetic value can influence the utility of theory pursuit. For many scientists, the beauty of a theory
increases the expected utility of pursuing the theory. For them, there is an important sense in which beauty is a goal of their science. I argued that we should distinguish at least two different decisions: (1) the decision confronting an individual scientist over what theory to pursue and (2) the decision confronting a society over what research to fund and promote. The role of beauty need not be the same in both cases. Poincaré, Duhem, and Kuhn argue that aesthetic value should influence how scientific research should proceed. Their insights can largely be captured within the decision theoretic framework. Kuhn disagrees with Poincaré and Duhem over the status of the various utilities associated with pursuing a given beautiful theory. Poincaré and Duhem believe that the balance between different types of value is independent of what we take it to be. They are inter-value realists. They think that certain values are fixed by the nature of (correct) science. Kuhn, on the other hand, promotes the importance of individual variation and argues that the correct balance between different types of value depends upon one’s personal biography. Alternatively, I suggested that one could argue that the value of freedom of inquiry justifies a scientist’s right to value aesthetic properties in a manner that she sees fit, given certain constraints. On this suggestion, freedom of inquiry underwrites an inter-value anti-realism at the individual level. Furthermore, I sketched the beginnings of an inter-value anti-realist position over the social decision theory problem where the correct balance between different values is a consensus achieved by democratic ideal deliberators. One consequence—and one virtue—of my view is that some scientists can reasonably choose to pursue ugly theories over otherwise similar beautiful alternatives. Given that the world appears to be a mixture of beauty and ugliness, it is appropriate that scientists, those who society deems are the best equipped to investigate the world, as well as investigate beauty also investigate ugliness.
Appendix I  Archival Sources

Novartis Foundation Archive
41 Portland Place
London W1B 1BN

Tate Gallery Archive
Millbank
London SW1P 4RG

Wellcome Library for the History and Understanding of Medicine
History of Medicine Collections
183 Euston Road
London NW1 2BE

Royal Society Library and Archives
6-9 Carlton House Terrace
London SW1Y 5AG

Frank MacFarlane Burnet Records
University of Melbourne Archives
Melbourne, Victoria
Australia

MRC Laboratory of Molecular Biology Archive
Hills Road
Cambridge CB2 2QH
United Kingdom

Anne Sayre Collection of Rosalind Franklin Materials
Special Collections
Albin O. Kuhn Library & Gallery
University of Maryland, Baltimore County
1000 Hilltop Circle
Baltimore, MD 21250
Rosalind Franklin Papers
Churchill Archives Centre
Churchill College
Cambridge CB3 0DS
United Kingdom

Jeremy Norman Collection
Private Archive for Molecular Biology
P.O. Box 867,
Novato CA 94948-0867

R Buckminster Fuller Papers
Department of Special Collections and University Archives
The Stanford University Libraries
Stanford, CA 94305-6004

Cold Spring Harbor Laboratory Archive
I Bungtown Road
Cold Spring Harbor, NY 11724
Appendix II  Interviews with the Author

Ed Applewhite: Washington DC, 21 January 2000
Sydney Brenner: Telephone Interview, 27 August 2001
Don Caspar: Tallahassee, 4-5 December 1998; Washington DC, April 30, 2002
Carolyn Cohen: Telephone Interview, 12 June 2000; Boston, 26 July 2001
Dick Crane: Telephone Interview, 9 Dec 1998; 13 January 1999
Francis Crick: Telephone Interview, 23 Nov 1998
Jeremy Goldberg: Washington DC, 4 April 2001
Ken Holmes: Pangbourne, England, 3 July 2000
Hugh Huxley: Boston, 26 July 2001
Barbara Low: Telephone Interview, 15 August 2001
Magda Cordell McHale: Telephone Interview, 3 November 1999
Ivan Rayment: Telephone Interview, 12 June 2000
Willy Russell: Telephone Interview, 29 January 1999
James Watson: Telephone Interview, 1 March 1999
Jo Wildy: Hereford, England, 10 June 1999
Bibliography


Caspar, D. L. (1984), “This Week’s Citation Classic: Caspar D L D & Klug A”, *Current Contents* **27**: 15


255


Scholarly Life

I was born on June 3, 1971 and raised on “Two Mile Station,” a mountainous, 30,000-acre high-country sheep, cattle, and deer ranch near St Bathans, Central Otago, New Zealand. Due to the sparsely populated surrounding district, I was sent to Otago Boys' High School, a boarding school in Dunedin, at the age of 12. After high school, I enrolled at University of Otago, gaining a BA in economics, a BSc in biochemistry, and a Dip Grad in medical ethics. While at University of Otago, I was member of Knox College and was elected president of the students' executive. In my senior year, I was selected to be Otago University's first exchange student at University of California at Berkeley. It was in liberating Berkeley, after reading a book by Rudolf Carnap and taking a class with Prof. Lisa Lloyd, that I decided to pursue a career in the philosophy of science. I returned to New Zealand and began to study philosophy under the mentorship of Prof. Paul Griffiths and Prof. Alan Musgrave. In 1994, I began graduate school at the University of Pittsburgh and was awarded an MA in history and philosophy of science and an MS in molecular biology (resulting in two publications: Morgan 1998 and Morgan et al. 2002). The former degree was influenced most by the historian of biology, Prof. Robert Olby and the late philosopher of science, Prof. Wesley Salmon. The latter degree involved a stay in Prof. Roger Hendrix's laboratory, where I studied the genome and evolution of the virus/transposon Bacteriophage Mu. In Pittsburgh I met a medical ethics student who would become my future wife. To continue our respective careers, we decided to move to the Baltimore/Washington DC area. In 1997 I began a PhD in philosophy at JHU. As well as take classes at Johns Hopkins University, I worked with Prof. Ken Schaffner at George Washington University on the role of ontogeny in evolution (Morgan 2001). My dissertation studies began under the direction of Prof. Peter Achinstein and Prof. Karen Neander. When Prof. Neander moved to University of California at Davis, Prof. Michael Williams became my second reader. I chose to look at the Caspar-Klug theory of virus structure, a theory that had not been
previously researched by either historians or philosophers of science. Some of my historical research was published (Morgan 2003; Morgan 2004). I will begin my first full-time academic job as Assistant Professor of Philosophy at Spring Hill College, Mobile AL in the fall of 2004.