
Manipulating Underdetermination in Scientific Controversy: The Case of the Molecular Clock

Michael R. Dietrich
Dartmouth College

Robert A. Skipper, Jr.
University of Cincinnati

Where there are cases of underdetermination in scientific controversies, such as the case of the molecular clock, scientists may direct the course and terms of dispute by playing off the multidimensional framework of theory evaluation. This is because assessment strategies themselves are underdetermined. Within the framework of assessment, there are a variety of trade-offs between different strategies as well as shifting emphases as specific strategies are given more or less weight in assessment situations. When a strategy is underdetermined, scientists can change the dynamics of a controversy by making assessments using different combinations of evaluation strategies and/or weighting whatever strategies are in play in different ways. Following an underdetermination strategy does not end or resolve a scientific dispute. Consequently, manipulating underdetermination is a feature of controversy dynamics and not controversy closure.

1. Introduction

Underdetermination is common in scientific practice. It is widely recognized by scientists even if they do not discuss it in terms frequently used by philosophers. We argue that in the course of scientific controversies underdetermination can be and often is manipulated by scientists to direct the terms and course of dispute. During a dispute, scientists use a wide array of different strategies in different combinations in their comparative

Thanks to Dick Burian, Roberta Millstein, and two anonymous reviewers for *Perspectives on Science* for comments on earlier drafts of this paper. RASjr also acknowledges the support of the Charles P. Taft Research Center for their support in the form of a Fellowship during the time this paper was written.

Perspectives on Science 2007, vol. 15, no. 3
©2007 by The Massachusetts Institute of Technology

assessment of theories. This assessment process involves trade-offs between different strategies as well as shifting emphases as different strategies are given more or less weight in an assessment situation. The multidimensional framework of theory assessment provides the resources for the manipulation of underdetermination. Manipulating underdetermination is not a feature of controversy closure in that it does not end or resolve dispute. Rather, it is a feature of a continuing controversy and so of controversy dynamics.

We elaborate and illustrate our view with a case study from molecular evolution, i.e., the controversy over the molecular clock. As molecular biology emerged as a discipline in the 1960s, some biologists began to consider how molecules themselves had evolved. Two startling and controversial concepts emerged from this early work on molecular evolution. In the early 1960s, Emile Zuckerkandl, Linus Pauling, and others recognized that they could use similarities and differences between molecules to infer their degree of evolutionary relatedness. When Zuckerkandl and Pauling compared protein sequences, however, they discovered that molecules like hemoglobin seemed to be changing at a constant rate. They called this approximate rate constancy the molecular clock. For evolutionary biologists used to thinking in terms of natural selection, a constant rate of evolution was extremely unlikely since selection depended on interactions with the environment which was understood to be continuously changing. The rate of evolution under selection should also change or fluctuate. The molecular clock was made even more controversial by its association with a new theory of molecular evolution. In 1968 Motoo Kimura argued that most detected changes in molecules were not due to the influence of natural selection. Instead, they were governed by genetic drift and were therefore considered neutral with respect to selection. The neutral theory was perceived as a direct threat to selectionism then dominating organismic evolution (Dietrich 1994). Since the early 1970s the molecular clock has been a very important part of the neutral theory of molecular evolution. Indeed, an important part of the appeal of the neutral theory has been that it provides a very elegant explanation for constant rates of substitution. We consider efforts to argue for a selectionist alternative for the molecular clock and so to undermine the clock's support for the neutralist position. While the neutralist and selectionist positions are often represented as polar opposites, the efforts to propose a selectionist clock do not proceed by simply asserting the correctness of the selectionist clock and the failure of the neutralist clock. Instead selectionists have proposed a new set of arguments through the manipulation of underdetermination. These arguments attempt to put the neutralist and selectionist interpretations of the clock on equal evidentiary footing, but do not resolve the controversy.

We begin in section (2) with a discussion of theory assessment in biological science that emphasizes the numerous standards, categories, and dimensions of evaluation. In section (3), we explore the connection between theory evaluation and underdetermination. With these apparatuses in place, in section (4) we show, by way of an analysis of the controversy over the molecular clock, precisely what we think manipulation of underdetermination consists in and how we think it works. Section (5) is reserved for discussion and concluding remarks.

2. Theory Evaluation.

In the assessment of scientific theories, a number of methodological principles (epistemic, pragmatic, and social) may be at work. For instance, Thomas Kuhn offered a list of five criteria for good theories, viz., accuracy, consistency, broad scope, simplicity, and fruitfulness (Kuhn 1977, 322), and W. H. Newton-Smith has offered a list of eight “good-making features of theories,” viz., observational nesting, fertility, track record, inter-theory support, smoothness, internal consistency, compatibility with well-grounded metaphysical beliefs, and simplicity (Newton-Smith 1981, 226–230). There are other such lists. Moreover, these lists could be expanded to include pragmatic criteria, such as cost and tractability, and social criteria, such as opportunism in context. Some of these criteria, such as simplicity, are considered epistemic by some philosophers (Laudan 1981, 196) and social by some sociologists (Bloor 1981, 201). Indeed, the debate between philosophers and sociologists is often construed in terms of the epistemic and the social, but it need not be. Bruno Latour, for instance, advocates a move beyond the social and epistemic, beyond relativism and realism (Latour 1992; cf. Longino 1990; Solomon 2001).

Biologists have used an identifiable constellation of standards in their comparative evaluation of most theories, and philosophers of biology have done considerable work to understand the ways in which the process works. Three main approaches may be identified: confirmation, epistemology of experiment, and strategies for generating scientific change. Each of these approaches has been championed in the philosophy of biology: confirmation by Elisabeth Lloyd (1987; 1988), epistemology of experiment by David Rudge (1998, 2001), and assessment during cycles of scientific change by Lindley Darden (1991). Elsewhere, Skipper (2000) has argued that there is no good reason to think that these three approaches to theory assessment are mutually exclusive. Indeed, based on a critical historical analysis of a persistent controversy in evolutionary genetics, Skipper developed a comprehensive framework of theory evaluation highlighting the ways in which the three approaches complement each other. We think this framework holds considerable promise for under-

standing theory assessment in the biological sciences. In particular, we think it is a useful framework for understanding the kinds of standards biologists appeal to in making theory choices in underdetermination situations. Herein, we review this multidimensional framework for theory assessment, applying it and elaborating on how it can be used to make theory choices in situations of underdetermination in subsequent sections of the essay. The framework is summarized in Figure 1 below; discussion of the approaches in which they are embedded follows.¹

Many in the philosophy of science are convinced that the core of theory assessment is the examination of the very specific relationship between hypotheses and the data adduced for them, i.e., the confirmation relation. This view may well be right. However, let us be clear that confirmation is not all there is to theory assessment. Nevertheless, let us start there. In philosophy of biology, Lisa Lloyd (1987, 1988) has developed a well-known, qualitative account of confirmation of evolutionary and ecological models drawn from a series of case studies in population genetics and ecology.

Lloyd has argued that confirmation of evolutionary and ecological models can be practically specified by three criteria, viz., fit between theoretical model and data, independent support for aspects of a particular theoretical model, and variety of evidence (Figure 1: B1–B3). Variety of evidence (Figure 1: B3) is at the core of Lloyd's account. Variety of evidence as a strategy itself is at bottom iterations and/or combinations of her other two strategies, i.e., fit between model and data and independent support for aspects of a model (Figure 1: B2–B3). Lloyd's intuition behind variety of evidence is that numerous and varied demonstrations of, e.g., fit, can increase the strength of the hypothesis that a theoretical model corresponds to the natural system modeled. A similar claim can be made for independent support. The more times, the more different ways and types of ways, and the more assumptions are tested independently, the greater the empirical support for a hypothesis made about a model that appeals to that assumption. Variety of types of support is a mixture of variety of instances of fit and variety of independent support. So, a hypothesis made about a theoretical model that has a variety of instances of fit along with a variety of instances of independent support for its assumptions is in a different situation regarding its confirmation than a theory with only one instance of fit.

1. The sheer number of strategies, 30 if we are counting, makes it impractical to explicate each one in the present paper. Consequently, only a few will be discussed. See Darden 1991, Lloyd 1987, 1988, Rudge 1998, and Skipper 2000, 2002, 2004a, 2004b for a thorough explanation and justification of each.

A. FORMAL ELEMENTS	B. RELATIONSHIP BETWEEN THEORY & EVIDENCE	C. RELATIONSHIP BETWEEN EXPERIMENT & EVIDENCE	D. RELATIONSHIP TO OTHER THEORIES	E. FUTURE PROSPECTS
A1. Internal consistency and non-tautologousness	B1. Fit between model & data	C1. Experimental checks & calibration	D1. Relations with other accepted theories	E1. Extendability
A2. Systematicity vs. modularity	B2. Independent support for aspects of a model	C2. Reproduction of known artifacts		
A3. Clarity	B3. Variety of evidence	C3. Experimenter intervention	D2. Relations to rivals	E2. Fruitfulness
A4. Experimental testability	B4. Explanatory adequacy	C4. Independent confirmation of results by different experiments		
A5. Quantitative predictability	B5. Predictive adequacy	C5. Elimination of error & alternatives		
A6. Simplicity	B6. Scope & generality	C6. Using results themselves to argue for their validity		
A7. Dimensional expandability	B7. Lack of ad hocness	C7. Use of independently corroborated theory to support results		
A8. Dynamical behavior		C8. Using an apparatus based on well-corroborated theory		
		C9. Using statistical arguments		
		C10. Replication		
		C11. use of controls		

Figure 1. Categories and strategies for theory assessment. There are five main categories listed across the top row, viz., (A)-(E). Strategies from the approaches discussed in the text are represented. We do not take the list to be exhaustive.

Lloyd sees herself as describing at least part of the set of techniques evolutionary biologists use to relate theoretical models to natural systems. More specifically, she sees her three confirmation criteria as concretely articulating techniques that are part and parcel of an experimental model (following Suppes 1962). The experimental strategies Rudge (1998, 2001) delineates are also part and parcel of experimental models in much of biology. Rudge (1998, 2001) considers the evaluation of experimental results made via observational and experimental procedures in the context of detecting natural selection and includes such strategies as calibration (of apparatuses), artifact reproduction, experimental interventions, and so on (see Figure 1: C1–C11 for all of Rudge’s strategies). Rudge’s account is based directly on Allan Franklin’s (1986, 1990) epistemology of experiment.

Via historical case studies from high energy physics, Franklin introduced a set of experimental strategies he packaged as an epistemology of experiment. Franklin’s broad aim was to demonstrate that claims based on experimental results are epistemically warranted, *pace* the then key social constructivists of science, such as Harry Collins (1985). Franklin offered

nine experimental strategies in his earlier work. To demonstrate the reasonableness of the strategies, Franklin justifies each using Bayesian epistemology. Franklin (2002) has continued and expanded his work on epistemology of experiment. Rudge (1998) takes Franklin's epistemology and critically examines it against a handful of detailed historical case studies, most notably H. B. D. Kettlewell's studies of industrial melanism in the peppered moth, *Biston betularia*. Rudge found that Franklin's experimental strategies were used in the Kettlewell work and, further, that additional Franklin-like strategies could be drawn from that work, such as the use of experimental controls (Figure 1: C11). Further, like Franklin, Rudge "epistemologizes" his experimental strategies using the Bayesian epistemology.²

Rudge's experimental strategies are meant as standards for evaluating experimental or observational data, whereas Lloyd's confirmation criteria are meant as standards for assessing the relationship between that data and specific theoretical hypotheses. Indeed, looking at the relationship between the practical procedures used to produce data and the formal tools for evaluating the relationship between hypotheses and data is looking at the opposite sides of the same coin. Take, for instance, variety of evidence or, more specifically, variety of instances of fit between model and data. Replication of experimental results is one way of concretely specifying that variety of evidence. And in so doing, we can assess the techniques of replication used as a way of qualitatively evaluating the variety of instances of fit between model and data. And in fact this kind of assessment happens often in comparative assessment situations (e.g., Skipper 2002). Other examples of the ways in which experimental strategies may elaborate confirmation criteria are available. Recently, for instance, Skipper (2004b) showed how calibrating laboratory populations using data from natural populations can be used to assess the relative strength of claims that there is independent support for aspects of a model. The important point here is that experimental evaluation strategies are available for critically examining the strength of empirical claims made about models. Moreover, it should be clear that different combinations of experimental and confirmational criteria may be used.

Now consider Darden's account of strategies for theory assessment during cycles of scientific change. Darden draws many of her strategies out of

2. Recently, Rudge 2001 has compared his Bayesian analysis of the Kettlewell work with an analysis of that work using Deborah Mayo's 1996 error statistical epistemology of experiment. Rudge now thinks that Mayo's epistemology of experiment better captures the epistemology of the Kettlewell experimental work. But the issue exercising Rudge here concerns Bayesian vs. error statistical approaches to understanding the growth of knowledge, not whether the experimental strategies identified in Rudge 1998 apply.

a substantial case-study in the history of Mendelian genetics (Darden 1991). Other of her strategies come from a survey of the philosophical literature. For Darden, scientific change proceeds through interacting cycles of discovery, evaluation, and revision, and her task is to delineate the constellation of reasoning strategies scientists use (or might use) to catalyze change during these cycles. Darden approaches theory evaluation by considering how theories are evaluated on four of the five categories of theory evaluation in Figure 1 (columns A1–A6, B4–B8, D1–D2, E1–E2; see Skipper 2000, 2004a on A7–A8). Indeed, the categories of theory evaluation we delineate in Figure 1 are based on Darden's organization of her own evaluation strategies (Darden 1991, 258).

In terms of categories of theory evaluation, we have so far explored the related categories of the relationship between hypotheses (or models or theories) to data and the relationship between data and the experimental techniques used to generate them (columns B and C in Figure 1). Among the strategies Darden delineates, several are meant as standards for assessing the relationship between theory and evidence. Indeed, we think that such strategies Darden discusses, including explanatory adequacy, predictive adequacy, and scope/generalizability can be particularized by Lloyd's confirmational criteria. Consider Darden's strategy of assessing explanatory adequacy (Figure 1: B4). How might a scientist determine whether some explanation is adequate? Darden does not give many details. We think that Lloyd's confirmational criteria (as well as Rudge's experimental strategies) can play this role in a straightforward way. Moreover, we think that information gleaned from appealing to criteria of confirmation and more specific assessment of experimental results may be used by scientists to judge claims about predictive adequacy, scope/generalizability, and lack of ad hocness (Figure 1: B5–B7).

But Darden's strategies for theory evaluation go beyond just the relationship between theories and evidence to include the assessment of the formal elements of theories (Figure 1, column A), the assessment of theories in relation to other theories (column D), and the future prospects of theories (column E). That is, in addition to empirical strategies for theory assessment, Darden includes non- or extra-empirical ones. Now, some philosophers argue that non- or extra-empirical assessment standards are not epistemic but are instead pragmatic (e.g., Sober 1994). We do not wish to enter into this debate here and, anyway, have already committed ourselves to the view that theory evaluation standards cut across epistemic, pragmatic, and social dimensions. In that case, it is fair to say that Darden's framework for theory assessment is broader than either Lloyd's or Rudge's (which is not really an account of theory assessment, but can be understood to contribute to Lloyd's). And, importantly, it seems clear that

Darden's framework exemplifies the multidimensionality of theory evaluation.

The central point of the examination of theory evaluation in this section is this: Biologists, and surely scientists more generally, appeal to a broad array of theory assessment strategies that cut across epistemic, pragmatic, and social dimensions. Moreover, these strategies may be used in various combinations. In addition, we will see that assessment strategies are assigned different weights in different combinations during comparative evaluation of theories in situations of underdetermination. And in that context, it will become clear that scientists are able to shift the grounds and terms of dispute by deploying the multidimensional framework of theory assessment which we call "manipulating underdetermination" in different ways. Let us now turn to the relationship between comparative theory assessment, or theory choice, and underdetermination.

3. Underdetermination and Theory Choice.

In its strongest form, the Duhem–Quine thesis claims that a single theoretical hypothesis cannot be conclusively falsified, so "any statement can be held to be true come what may if we make drastic enough adjustments elsewhere in the system" (Quine 1953, 43).³ In effect, this form of the Duhem–Quine thesis claims that a very strong type of underdetermination follows from scientific holism: when faced with contravening evidence, it is always possible in principle to save some favored hypothesis by making enough revisions in other parts of the system associated with it. In other words, any number of theories can be generated which agree with the evidence presented. Philosophers and sociologists often take Quine at his word when he claims that "any statement can be held to be true come what may if we make drastic enough adjustments elsewhere in the system" (Quine 1953, 43).⁴ Interpreting Quine's famous claim, however, is anything but straightforward (Laudan 1990).

When philosophers and sociologists talk about underdetermination, the usual case concerns the underdetermination of theory by evidence. From this pattern of usage, it would seem then that, in general, underdetermination is a two part relation: $\{X_1, X_2, \dots X_n\}$ is underdetermined by $\{Y_1, Y_2, \dots Y_m\}$, where the members of the X-set are usually theoretical systems, theoretical models, or theories and members of the Y-set are usually statements of experimental results or models of data. In fact, however,

3. Quine is referring to a system of beliefs composed of an array of more or less theoretical and observational propositions.

4. On the limits of interpretive flexibility and the debate over underdetermination between philosophers and sociologists, see Dietrich 1993.

underdetermination is not so much about theories, as it is about the grounds for judging between theories. Put another way, theories are not underdetermined; choices between theories are underdetermined.

Given this distinction, underdetermination can be defined as follows: A choice between members of the set $\{X_1, X_2, \dots, X_n\}$ based on the set $\{Y_1, Y_2, \dots, Y_m\}$ is underdetermined if and only if: (1) there is a set $\{X_1, X_2, \dots, X_n\}$, (2) each X in $\{X_1, X_2, \dots, X_n\}$ is pair-wise contrary with every other member of $\{X_1, X_2, \dots, X_n\}$, (3) there is a set $\{Y_1, Y_2, \dots, Y_m\}$, (4) there is a relation C such that for each member X of $\{X_1, X_2, \dots, X_n\}$, X stands in the relation C to $\{Y_1, Y_2, \dots, Y_m\}$, where C (the underdetermination relation) is some relation concerning compatibility which should be specifically defined in each case. Underdetermination, then, is always relative to some basis for choice used by scientists in their deliberations over alternative theories. We will articulate this basis for choice in terms of evaluation standards. Thus, underdetermination is relative to the grounds for claims made by way of the assessment standards in Figure 1. However, the basis for choice need not be articulated in terms of evaluation standards.

Different versions of underdetermination will be a result of differences in the X -set and Y -set and/or in the definition of the underdetermination relation. The range of the Y -set is always specified and relative to a particular judgment at a particular time. The range of the X -set varies, but it will always include at least two members. Using this general schema, one can clarify exactly what is meant by various claims of underdetermination. In contrast to general and vague claims about underdetermination, the schema advocated here provides the basis for what we will call specified underdetermination.

Consider for instance the general and familiar claim that a theory is underdetermined by evidence. In order to clarify this claim the nature of the relation between the theories and evidence must be articulated; in terms of the scheme presented here, the underdetermination relation must be specified. If we define C , the underdetermination relation, as logical entailment, then we have what Larry Laudan calls deductive underdetermination. In his words, deductive underdetermination can be formulated as a claim similar to the following one: "For any finite body of evidence, there are indefinitely many mutually contrary theories, each of which logically entails the evidence" (Laudan 1990, 269).

Notice that defining C in terms of logical entailment restricts the ways the theories (the X -set) and the evidence (the Y -set) can be formulated—they must be formulated such that they can stand in a relation of logical entailment to each other, e.g., as theoretical and observational statements. Deductive underdetermination produces the familiar result that deductive logic alone is not capable of determining whether or not an empirical the-

ory is true with certainty, regardless of the evidence. However, deductive underdetermination alone does not mean that theory choice is completely underdetermined, but only that deductive logic is not going to be the way the choice is made. Deductive underdetermination does not rule out the possibility that other grounds for theory choice could be fully determinate (Laudan 1990, 270).

A more typical formulation of the underdetermination relation found in claims about the underdetermination of theory by evidence involves some form of epistemic relation. We will call an underdetermination relation epistemic if it is chiefly concerned with some type of empirical support or the relation of the theories to the world (Laudan 1990, 271, McMullin 1982, 18, van Fraassen 1980, 88).⁵ For instance, Laudan distinguishes three possible epistemic relations: (1) any theory in the set of rival theories can be made *logically compatible* with the evidence; (2) any theory in the set of rival theories can *explain* the evidence; and (3) any theory in the set of rival theories can be *empirically supported* by the evidence (Laudan 1990, 275). Note that implicit in these underdetermination relations are theory evaluation standards. So, for instance, the standard “prefer theories with greater explanatory adequacy” (Figure 1: B4) is implicit in Laudan’s second alternative listed above, while the standard “prefer theories which are empirically supported by all the presented evidence” (Figure 1: B1) is implicit in the first alternative. It should be evident that numerous ways of specifying epistemic underdetermination relations (C) are possible.

Underdetermination need not be limited to grounds for theory assessment that appeal to empirical support; it can be extended to cover almost any assessment strategy used, e.g., pragmatic or sociological. Consider the sociological standard “prefer theories which when implemented in laboratory practice will have the lowest reasonable cost.” If in some specific case the Y-set is just the lowest reasonable cost in monetary terms, the X-set is a set of rival implementation plans for related theories, and C is defined as “is the cost of,” then we can make the underdetermination claim that any implementation plan in the set of rival implementation plans will cost the same—they all fit equally well with the lowest reasonable cost. For any strategy of assessment in the multidimensional framework of Figure 1, there is plausibly some underdetermination relation that can apply to it provided the Y-set includes the appropriate entities.

As we have seen, on our view any given theory assessment involves a constellation of evaluation strategies. Because in any given situation a number of different standards may be at work, the degree of underdeterm-

5. Laudan refers to these epistemic relations as ampliative relations. We prefer the more general term “epistemic.”

ination must be specified. The degree of underdetermination refers to the number of standards that are shown to be underdetermined relative to some particular theory. The degree of underdetermination can be weak when only one or two of the standards are underdetermined or it can be very strong when many of the strategies are shown to be underdetermined. We might say, for instance, that some X-set is underdetermined with respect to explanatory adequacy (Figure 1: B4); this is a low degree of underdetermination. Or, we might say that some X-set is underdetermined with respect to several forms of variety of evidence (Figure 1: B3); this is a higher degree of underdetermination. Specifying the degree of underdetermination requires that the strategies which could possibly be underdetermined in the case at hand be articulated. For the moment we are not assuming that any kind of strategy is more important than any other in theory choice or controversy resolution. But, as we will see in our discussion of the controversy over the molecular clock in the next section, differential weighting of standards could affect the significance attached to different underdetermination claims.

Another important way in which underdetermination claims ought to be specified involves the range of underdetermination, where that refers to the number of members of the X-set. For instance, Laudan's formulation of deductive underdetermination, above, allowed the X-set to range over indefinitely many theories, but it could, and often does, have a much smaller range, e.g., two genuine rivals. Differences in the range of the X-set provide further grounds for distinguishing between different types of underdetermination. Differences in the range of the X-set, for instance, lie behind Laudan's distinction between what he calls the nonuniqueness thesis and the egalitarian thesis. Put in terms of theoretical systems, the nonuniqueness thesis is the thesis that:

For any theory, T, embedded in a system, S, and any body of evidence, e, there will be at least one other system, S' (containing a rival to T), such that S' is as well supported by e as S is (Laudan 1990, 292).

Similarly stated, the egalitarian thesis is the thesis that:

For any theory, T, embedded in a system, S, and any body of evidence, e, there will be systems, S₁, S₂, . . . , S_n, each containing a different rival to T, such that each is as well supported as T (Laudan 1990, 292).

The basic difference here is that the egalitarian thesis indicates that there are a number of equally supported rivals and the nonuniqueness thesis claims that there is at least one equally supported rival. The egalitarian

thesis is, thus, stronger than the nonuniqueness thesis. Since the difference between nonunique forms of underdetermination and egalitarian versions of underdetermination is a matter of the number of theories involved, we will refer to the range of underdetermination rather than to nonunique or egalitarian underdetermination. The question of what the range of underdetermination is in a specific case is empirical. If one wishes to claim that an indefinite number of theories are compatible with the evidence, then one must provide some reason for advocating such an extended range.

But suppose that there is a strong degree of underdetermination of an indefinite range of rival theories by some set of data. It would follow then that in a Quinean fashion one could conclude that any theory may be held regardless of the evidence it faces. As Laudan notes, this claim can be interpreted descriptively or normatively (Laudan 1990, 272). If, for instance, Quine is just claiming that it is possible to retain beliefs despite contravening evidence, then he is making a purely descriptive claim about underdetermination. If he is claiming it is rational to hold on to any statement regardless of the evidence facing it, then he is making a normative claim about underdetermination. Laudan argues that if underdetermination is to have any implications for normative epistemology, then it must be construed normatively or in terms of rationality (Laudan 1990, 272).⁶ Differently put, descriptive underdetermination refers to all the possible courses of action available, while normative underdetermination refers to only the rational courses of action. In most cases, the set of rational courses of action will be a subset of the set of possible courses of action. So, in effect, considerations of rationality or reasonableness constrain the courses of action that are available. For example, Laudan considers construing underdetermination in terms of theories being equally compatible with the evidence. Logical compatibility can be achieved quite easily by simply eliminating without replacement those parts of the theory that make it incompatible with the data. This kind of elimination, however, comes with a price; eliminating large chunks of a theory may drastically and negatively affect that theory's explanatory adequacy (Figure 1: B4). If explanatory adequacy is valued in theory assessment, then it may not be reasonable to sacrifice explanatory adequacy for compatibility. So, in the case of compatibility, the question is: Is it rational to preserve a theory by making it compatible with the evidence if this action is going to harm the theory's explanatory adequacy? Similar tradeoffs are not hard to imagine. The bot-

6. Rationality is understood by Laudan to be instrumental rationality. The sense of normativity advocated here need not be limited to instrumental rationality.

tom line here is that if underdetermination claims are to have any epistemic bearing, their advocates must attend to the question of the rationality of a course of action, not just its possibility (Laudan 1990, 276).

We have argued for a schema that suggests that underdetermination must be specified so that an underdetermination relation, i.e., epistemic, pragmatic, sociological, must be articulated. In addition, the view requires specification on a case-by-case basis the relevant strategies for evaluating alternative theories, the degree of underdetermination, the range of underdetermination, and the normative status of the underdetermination claim (Dietrich 1993). The controversy over the molecular clock in molecular evolution, discussed in the next section, illustrates these facets of underdetermination. As we will see, the case establishes a claim of epistemic underdetermination, making use of some of the strategies from the framework of theory evaluation delineated in Figure 1. Most importantly, the following discussion of the molecular clock controversy illustrates this paper's main, and novel, claim: During the course of scientific controversies underdetermination of the strategies for theory assessment is manipulated by the scientists to direct the course and terms of dispute. Indeed, the case presented below makes manifest one example of the way in which this manipulation, or "underdetermination strategy," has been used.

4. The Case of the Molecular Clock: Illustrating the Underdetermination Strategy.

Although they had articulated the idea earlier, the molecular clock was christened in 1965 by Emile Zuckerkandl and Linus Pauling (Zuckerkandl and Pauling 1965, Morgan 1998). The idea behind the molecular clock was that the observed changes in the amino acid sequences of a protein from different species should be "approximately proportional in number to evolutionary time" (Zuckerkandl and Pauling 1965, 148). The differences observed across species could then serve as an evolutionary timescale. As a timekeeping device, the molecular clock was inherently stochastic. The "ticks" of the clock were not uniform, but were best understood as a regular statistical process producing a distribution of rates of substitution for the species and molecule under consideration. The molecular clock immediately attracted interest and controversy (Dietrich 1998, 2006).

4.1. The Neutralist Approach.

In their initial attempts to explain the mechanism for the clock, Zuckerkandl and Pauling invoked both natural selection and random drift to explain the clock's apparent constancy (Morgan 1998). After 1968,

Motoo Kimura, Allan Wilson and others used the neutral theory of molecular evolution to explain the mechanism of the clock. Indeed, Motoo Kimura's original arguments in 1968 for the significance of neutral mutations and random drift were based in part on the hemoglobin data presented by Zuckerkandl and Pauling in 1965. By 1971, Tomoko Ohta and Motoo Kimura asserted that the "remarkable constancy of the rate of amino acid substitutions in each protein over a vast period of geologic time constitutes so far the strongest evidence for the theory (Kimura 1968, King and Jukes 1969) that the major cause of molecular evolution is random fixation of selectively neutral or nearly neutral mutations" (Ohta and Kimura 1971, 18). The reason the constancy provided such strong support is that the neutral theory provided a mechanism for that constancy—the prediction of rate constancy followed easily from basic theoretical commitments of Kimura and the neutralists. According to the neutralists, the rate per generation of mutant substitutions in a population (k) is equal to the mutation rate per gamete (v):

$$k = v. \quad [1]$$

The result in [1] is derived as follows. In a population of actual size N , there are $2Nv$ new mutations produced in the entire population per generation. Only a certain fraction of these new mutations will become established in the population; that is, only a certain fraction will reach fixation. Let u represent the probability that a new mutation will reach fixation. Then, in Kimura's (1979, p. 108) words, "in a steady state in which the process of substitution goes on for a very long time, the rate k of mutant substitution per unit time is given by the equation

$$k = 2Nvu." \quad [2]$$

For selectively neutral mutations, however, the probability of fixation (u) is equal to $1/(2N)$, because "any one of the $2N$ genes in the population is as likely as any other to be fixed, and so the probability that the new mutant will be the lucky one is $1/(2N)$ " (Kimura 1979, p. 108). Substituting $1/(2N)$ for u in equation [2] yields equation [1]. It is important to note here that the rate of mutant substitution is independent of population size.

The rate of evolution for selectively advantageous mutants, in contrast to neutral mutants, is dependent on both population size and selection pressure. Kimura (1979, p. 110) uses a lengthy derivation to show that for selectively advantageous mutants the equation for the rate of evolution is

$$k = 4Nsv. \quad [3]$$

Thus, in order for a selectionist model to account for the constancy of the rates of evolution it must show how constancy is possible when the rates are strongly dependent on the environment, as represented by the selection coefficient s and the measure of population size N , which can be quite variable. Under the selectionist model, the rates of molecular evolution should show nearly the same variability as the rates of phenotypic evolution. Unfortunately for the neutralists, the rates of molecular evolution were found to be far from uniform.

The rate of amino acid substitution was known from the beginning (1965) to vary among different proteins. The neutralists explained this difference in terms of different proteins having different fractions of neutral mutants; the number of neutral mutants depends on the functional constraints for each protein. So, for instance, fibrinopeptide A has a much higher rate of substitution than histone IV, which is highly constrained (King and Jukes 1969, p. 792). But even within protein families variation was observed. So, for instance, insulins in the line leading to guinea pigs seem to have evolved faster than insulins in other lines (King and Jukes 1969; Ohta and Kimura 1971, 19). The neutralists needed a way to explain these deviations from the intrinsic rate of molecular evolution.

In 1971, Tomoko Ohta and Motoo Kimura analyzed these variations in proteins statistically. Claims about rates of evolution are based on the number of amino acid substitutions that actually occurred during the course of evolution and an accurate estimate of the time of divergence for the proteins being compared. The fraction of amino acid differences between two protein sequences (p_d) is equal to the number of differences in the two sequences (d_{aa}) divided by the total number of amino acid sites in the sequences (n_{aa}); so, says Kimura (1969, p. 1182),

$$p_d = d_{aa}/n_{aa} \quad \{4\}$$

In the course of evolution, however, a given amino acid site may have changed more than once; it may have had multiple hits. To correct for multiple hits, Zuckerkandl and Pauling and then Kimura assumed that the process of amino acid substitution could be modeled by a simple Poisson process.⁷ With this assumption, the probability of no substitution occurring at a given site is $e^{-K_{aa}}$, where K_{aa} is the average number of amino

7. A Poisson process is a purely random stationary process. In a purely random process, the behavior of the process in the past has no influence on future behavior (Lindley 1965, 67). A process is stationary if the probability of the incidents in the process are not influenced by temporal order (in other words, if it is invariant with respect to time) (Lindley 1965, 68). The classic example of a Poisson process is radioactive decay.

acid substitutions per site for the protein sequences being compared. It follows then that

$$1 - p_d = e^{-K_{aa}} \quad [5]$$

and that

$$K_{aa} = -\ln(1 - p_d). \quad [6]$$

The rate of amino acid substitution per site per year is

$$k_{aa} = K_{aa}/(2T), \quad [7]$$

where T is the number of years since divergence. From here the expected variance of k_{aa} can be computed and compared to its observed variance. When Ohta and Kimura did this for different alpha and beta hemoglobins, and for cytochrome c, they found that observed variance in the beta hemoglobin and the cytochrome c were significantly larger than expected. From this they concluded that “the variations in evolutionary rates among highly evolved animals are larger than expected from chance” (Ohta and Kimura 1971, 21). Ohta and Kimura did not take this as a reason to give up the neutral theory. The increased variance in substitution rates was chalked up to a small fraction of advantageous mutations that affect the molecule’s function but do not interfere with the constancy of the overall rate of substitution (Ohta and Kimura 1971, 23).

After Ohta and Kimura’s paper in 1971, a tremendous amount of empirical research was done on the molecular clock (Wilson et al. 1977). Charles Langley and Walter Fitch, for instance, used a procedure based on minimum phyletic distances to test whether or not the process of nucleotide substitution was a constant Poisson process. They concluded that “it is clear that the *total* rate of substitution (as observed through the minimum phyletic distance procedure) varies markedly in geological time and among divergent lines of descent” (Langley and Fitch 1974, 174). Similar conclusions stressing the non-uniformity of rates in hemoglobin and a slow down of rates in primate lineages were offered by Goodman, Moore, and Matsuda (1975). Even Kimura himself admitted that the rate of molecular evolution is not perfectly uniform (Kimura 1983, 79), but in his opinion, “emphasizing local fluctuations as evidence against the neutral theory, while neglecting to inquire why the overall rate is intrinsically so regular or constant is picayunish. It is a classic case of ‘not seeing the forest for the trees’” (Kimura 1983, 85).

Kimura backs up his account of rate constancy in his 1983 book, *The Neutral Theory of Molecular Evolution*, by providing a more thorough statistical treatment of the variations in the rate of molecular evolution. Kimura's analysis uses what has since been called a star phylogeny (Gillespie 1984a). The lineages in a star phylogeny are taken to have all diverged from a common ancestor in a relatively short period of time. Kimura considers the case of six mammals; humans, mice, rabbits, dogs, horses, and cattle, which diverged from each other about 80 million years ago (Kimura 1983, 76). What Kimura wants to know is "whether the intrinsic rates of amino acid substitutions among the six lineages are equal and whether variation of the observed numbers of substitutions as shown in Table 1 below lie within the limits of normal statistical fluctuations" (Kimura 1983, 76–77). Kimura's method is as follows for the case of alpha hemoglobin (see Table 1): The number of amino acids (n_{aa}) for alpha hemoglobin is 141. Using the information in Table 1 and equation [6], the number of substitutions corrected for multiple hits can be calculated and then used to calculate the mean and variance of the number of amino acid substitutions. In order to see if the variation in amino acid substitutions among lineages is larger than expected, the value of R is then determined, where R is the ratio of the observed variance to the expected variance. But, for Poisson processes, the mean is equivalent to the expected variance, so R can be readily computed as the ratio of the observed variance to the mean. For the case of alpha hemoglobin, Kimura calculated that $R = 1.26$. In order to find out if R 's deviation from the expected value of 1 is significant, Kimura argued that "the statistic $(n - 1)R$ should follow the X^2 (chi-square) distribution with $n - 1$ degrees of freedom" (Kimura 1983, 77). For the alpha hemoglobin case, the table of X^2 values shows that the deviation is not larger than that expected by chance, so Kimura infers that the hypothesis, i.e., that the intrinsic rate of evolution in these six lineages is the same, stands. Kimura then extended his analysis to include four other proteins, beta hemoglobin, myoglobin, ribonuclease, and cytochrome c. Of these additional proteins, myoglobin and ribonuclease showed no significant variation, but beta hemoglobin and cytochrome c showed significantly higher variation than expected. In Kimura's words, "these results suggest that although the strict constancy may not hold, yet a rough constancy of the evolutionary rate for each molecule among various lineages is a rule rather than an exception" (Kimura 1983, 79). Moreover, Kimura notes that the average value of R for the five molecules used is 2.6, which is consistent with earlier results showing that observed variances up to 2.5 times larger than expected are allowable if the variation is a result of chance alone (Kimura 1983, p. 79; Ohta and Kimura 1971). So, in the

Table 1. Observed numbers of amino acid differences between six mammals when their hemoglobin alpha-chains are compared. The ratio of the observed to the expected variances turns out to be as follows: $R = 1.26$. (Adapted from Table 4.3 in Kimura 1983.)

Human	Mouse	Rabbit	Dog	Horse	Bovine
Human	18	25	23	18	17
Mouse		27	25	24	19
Rabbit			28	25	25
Dog				27	28
Horse					18
Bovine					

end, Kimura admits that an approximate rate constancy holds as a rule, but he also admits that there may be deviations from the rule; hence, his admonition about not seeing the forest for the trees, as mentioned above.

4.2. The Selectionist Alternative and Critique of the Neutralist Approach.

Soon after Kimura published his account of variations in the rate of evolution, John Gillespie used the same data that Kimura had used to propose a rival interpretation that called for an episodic molecular clock (Gillespie 1984a). Where Kimura emphasized the underlying constant rate, Gillespie argued that the significance of the two proteins with larger than expected variances was “by no means obvious” (Gillespie 1984a, 8010).

Gillespie and Kimura were not new opponents in 1984; Gillespie had proposed a natural selection model for molecular evolution in 1978 (Gillespie 1978) and had questioned rate constancy as early as 1979 (Gillespie and Langley 1979). In 1984, Gillespie also published a negative review of *The Neutral Theory of Molecular Evolution* that acknowledged the achievements of the neutral theory but questioned Kimura’s style of advocacy (Gillespie 1984, 733). In effect, Kimura and Gillespie represent opposite poles in the molecular evolution community. Kimura was the senior advocate of the neutral theory. As head of the department of population genetics at Japan’s National Institute of Genetics, he had gathered around him a team of scientists, including, for instance, Tomoko Ohta and Naoyuki Takahata, who developed the neutral theory and explored its implications. Gillespie was a senior professor at the University of California at Davis. He made his career devising and promoting sophisticated selectionist approaches to molecular evolution.

The debate over rates of molecular evolution, as Gillespie saw it, was concerned with estimates of “the ratio of the variance to the mean of the number of substitutions per lineage,” what Gillespie calls “ $R(t)$ ” (Gillespie 1986a, 140). According to Gillespie, the neutralists need to explain why the value of $R(t)$ is larger than unity in many cases, and the selectionist needs to explain why the value of $R(t)$ is not estimated as being any larger than 3.4. Gillespie, of course, takes up the challenge for the selectionist position.

Gillespie’s main argument for an episodic molecular clock is that “the variation in the numbers of substitutions per lineage must ultimately be attributable to variation in the rates of substitutions” (1986b). Gillespie cashes out his position by proposing a model of the molecular clock based on a doubly stochastic Poisson process. This doubly stochastic Poisson process is “a Poisson process for which the rate of the process itself is a stationary stochastic process” (Gillespie 1984a, 8011). An important feature of this kind of model is the relation between the clocks in different lineages. Gillespie assumes that the clocks have “equal tick rates at the time of radiation” and that “the correlation in tick rates between lineages drops off at the same rate as the correlation within a lineage” (1984a, 8011). So, the rates can change randomly within lineages. Gillespie then assumes that the clock changes rapidly, or, as he says, “the rate of change of the rate of molecular evolution is assumed to occur on a time scale that is much shorter than the lengths of the lineages under study” (1986a, 141). Gillespie then says, “this assumption is motivated by the fact that major environmental changes, such as the recent ice ages, occur on a time scale of thousands to tens of thousands of years while the time between substitutions is typically on the order of millions of years” (1986a, 146). With this assumption Gillespie inferred that in order to fit the parameters of his statistical model the clock should be episodic, i.e., it should reflect a substitution process where there are long periods with no substitutions broken up with occasional short bursts of substitutions (Gillespie 1986a, 141).

There are two basic versions of the episodic clock: one is the two-state clock and the other is the gamma clock. The two-state clock is based on a two-state Markov process. According to Gillespie, “the process remains in state zero for an exponentially distributed time with mean $1/u_0$ and then jumps to the value $1/(u_0t)$, where it remains for an exponentially distributed time with mean t before returning to zero, and so forth” (Gillespie 1986a, 147–8). The gamma clock jumps between a gamma-distributed height and zero and has a substantially different mathematical basis from the two-state clock. Despite these differences, the two clocks produce similar estimates of both the mean number of substitutions per episode and the mean number of episodes. According to Gillespie, “given the very dif-

ferent characters of the two clocks, this suggests that the inferred episodic structure is robust to the assumptions made about the clock, at least for the restricted range of values of $R(t)$ for the currently available data” (1986a, 149). This demonstration of robustness puts the episodic clock on firmer footing by making it more resilient to attacks on its assumptions. In this way it also becomes harder to dismiss as non-genuine.

The episodic clock has a number of advantages from Gillespie’s selectionist perspective. First, the episodic nature of the clock agrees well with the selectionist model of molecular evolution developed by Gillespie (1984b). Gillespie’s (1986a, 141) model has three components:

- 1) a changing environment,
- 2) an epistatic scheme in which each environmental change presents a challenge to the species that may be met by substitutions at one of several nearly equivalent loci, and
- 3) a mutational landscape that makes it unlikely that any particular locus will experience the substitution of an allele that is two mutational steps away from the allele that is currently fixed in the population.

This model assumes that both the environmental changes and the number of loci available to respond are large. Second, the episodic clock fits well with the data. Gillespie’s model of the episodic clock predicts that the values of $R(t)$ should be in the range of $1.0 < R(t) < 3.5$, if “the number of loci that can respond to a particular environmental change is large” (Gillespie 1986a, 152–3). This was exactly the range of values observed at the time. Thus, in Gillespie’s words, “from the perspective of purely statistical considerations, our model of evolution by natural selection actually seems to fit the data better than does the neutral model” (Gillespie 1986a, 153). Given this agreement with observed values of $R(t)$ and the agreement between the episodic dynamics of the model and the episodic structure of the data, Gillespie concludes that he has shown that “a very plausible model of molecular evolution by natural selection fits the sequence of data just as well as does the neutral allele model” (Gillespie 1986a, 141).

Notice that what Gillespie has done is establish a case of epistemic underdetermination that renders the ability to generate predictions that agree with the observed values of $R(t)$ insufficient grounds for choice between existing neutralist and selectionist models of the rate of molecular evolution. Both models fit the data and are predictively accurate (Figure 1: B1, B5), so that the degree of underdetermination is two, even though Gillespie’s may be a slightly better fit. But that is not enough for Gillespie, who shifts the grounds and terms of assessment off the underdetermined assessment standards of fit and predictive accuracy and emphasizes

the independent support of the underlying selectionist assumptions of his model (Figure 1: B2), i.e., by producing what he says is “a very plausible model of molecular evolution” (Gillespie 1986a, p. 141). He still wants to charge that the predictions of the neutral model do not fit well with the large values of $R(t)$. And the neutralists want to explain these large values away as exceptions. The type of underdetermination argued for by Gillespie has the effect, therefore, of partially shifting the grounds of the controversy away from emphasizing fit between model and data and the ability to predict certain values of $R(t)$ (Figure 1: B1, B5), given certain theoretical commitments, to emphasizing other features of each of the models, particularly independent support for underlying assumptions (Figure 1: B2), as well as, simplicity and lack of ad hoc theoretical assumptions (Figure 1: A6, B7) as we shall see below. Differently put, Gillespie manipulates the underdetermination situation, changing and re-weighting the combination of epistemic standards for evaluation. This underdetermination strategy has the effect of shifting both the grounds and the course of comparative assessment.

Gillespie fleshes out his attack by claiming that the neutral theory’s model of the molecular clock, as found in Kimura’s 1983 presentation, is at a disadvantage because of the artificiality of its assumptions. In Kimura’s model the rates of evolution are fixed at the origin of each lineage and each lineage’s rate is independent of the others. But, “why,” Gillespie asks, “should the rates of mutation be altered only at the time of origin of the orders of mammals and then remain unaltered for the next 60 Myr even though other lineages are branching off to form families, genera, and species?” (Gillespie 1986a, 153). This artificiality leads Gillespie to assert that

[o]ne conclusion is clear: the neutral theory cannot, at present, account for the statistical patterns in the sequence data without the use of a very artificial model of rate variation. While the theory could be modified to be in better agreement with the data, such modifications will make the theory less compelling as a uniquely parsimonious explanation for molecular evolution (Gillespie 1986a, 153).

The formal or mathematical simplicity (Figure 1: A6) of the neutral theory is taken by Gillespie and others to be a major point in its favor, although not enough of an advantage as it stands to withstand his underdetermination strategy. Indeed Gillespie believes that the good standing of the neutral theory is partly a result of the difficulties inherent in producing a relatively simple selectionist alternative and is partly a result of the possibility of doing “a little rearranging” and having the theory

“emerge as strong as ever” (Gillespie 1987, 33–4). In the end, if the neutral theory is to emerge from his challenge, Gillespie predicts that “salvation will most likely come with a redefinition of the neutral model as exemplified in the recent work of Takahata (1987)” (Gillespie 1987, 33).

4.3. The Neutralist Response.

The controversy over the molecular clock became fully articulated when Gillespie introduced an alternative account of the molecular clock that challenged the mathematics, assumptions, and mechanisms of the neutralist’s interpretation of the clock. In the face of Gillespie’s challenge that the molecular clock is episodic, Naoyuki Takahata proposed a number of possibilities to try to diffuse the negative implications that an episodic clock has for the neutral theory. Takahata’s account is particularly revealing because he offers three ways to bring the observation of large values for $R(t)$ in line with the neutral theory, but rejects two of them as implausible. So, Takahata’s strategy is to produce a neutralist model that can account for the entire range of observed values of $R(t)$ from small to large.

Gillespie’s claims, according to Takahata, are that rate constancy is an artifact of a slow rate of amino acid substitution, that “molecular evolution may be episodic, with bursts of rapid evolution separated by periods of very slow evolution,” and that directional selection may explain the episodic nature of the clock (Takahata 1987, 170). Given these claims, Takahata admits that “it seems certain that some, if not all, genes evolve with sloppier clocks than those of radioactive materials” (Takahata 1987, 170). In other words, the rate of molecular evolution shows significant departures from the expectations of a simple Poisson process model, which can usually be exemplified by the process of radioactive decay. Takahata’s goal is to propose possible alternatives that are compatible with the neutral theory.

In order to account for values of $R(t)$ greater than unity, Takahata argues that we need to account for values of the dispersion index, $I(t)$, which are greater than unity. $I(t)$ is easier to manipulate than $R(t)$, but is related to $R(t)$.⁸ Takahata proposes three ways of generating large values of $I(t)$:

- 1) *Multiple substitutions*. The idea behind this option is that multiple substitutions at one time would produce a larger than expected variance. Takahata lists a number of authors who recognize that “some, if not many, substitutions in a gene might not have occurred singly

8. The difference between $R(t)$ and $I(t)$ is that $R(t)$ is an actual estimate of the ratio of the sample variance to the sample mean for the number of substitutions among lineages, while $I(t)$ is a ratio of the expectation of the sample variance to the expectation of the sample mean.

because of highly specialized intramolecular interactions” (Takahata 1987, p. 171). According to this model, in order for the index of dispersion, $I(t)$, to be much larger than unity, substitutions would have to have both a high multiplicity and rate of occurrence.

- 2) *Fluctuating substitution rate*. This option employs a doubly stochastic Poisson process, as Gillespie did, but tries to reach a different conclusion. In order to explain the bursts of changes and long periods of stasis that follow from this model, Takahata appeals to higher rates of occurrence and accumulation of deleterious mutations coupled with a reduction of population size (Takahata 1987, 173). With the help of a computer simulation, Takahata concludes that,

[i]f there are very severe bottlenecks whose mode of occurrences differs from lineage to lineage and deleterious mutation rates are much higher than neutral mutation rates, then there would be bursts of substitutions in each lineage and at the same time it can be expected that $I(t) > 1$ (Takahata 1987, p. 174).

- 3) *Fluctuating neutral space*. This option is based on a time-dependent renewal process that assumes that “the substitution rate fluctuates through changes of selective constraints as new substitutions take place one after another” (Takahata 1987, 174). The degree of selective constraint is what Takahata calls the neutral space. If changes only occur at the time of substitution, then this renewal model predicts that if the neutral space fluctuates to a large extent, $I(t)$ can be fairly large (Takahata 1987, 175). Fluctuations in the neutral space are supported by research on opossum alpha hemoglobin and guinea pig insulin. Both of these molecules have shown large changes in selective constraints as a result of intragenomic causes, such as the loss of a previously invariant amino acid (Takahata 1987, 176).

Takahata’s alternatives are meant to demonstrate that the neutral theory is compatible with large values of $R(t)$ which are held as evidence against it by Gillespie. So, Takahata accepts the reliability of the experimental result; namely, that the values of $R(t)$ are significantly larger than the expected value of 1. Takahata’s commitment to the neutral theory thus demands that he find components of the old theoretical system proposed by Kimura that can be revised or replaced such that the neutral theory will then be compatible with the experimental results. In his response to Gillespie, Takahata is following his own strategy of underdetermination; he is striving to formulate models that will generate high values of $I(t)$, and hence $R(t)$. In this way he can bring the neutral theory better in line

with the data. That is, he wants to refine the fit and predictive accuracy of specific assumptions of the neutral theory (Figure 1: B1, B5) to gain improved overall fit with the data (Figure 1: B1). But notice that observed values of $R(t)$ greater than unity will not decide between the rival models proposed because of Gillespie's underdetermination strategy; instead, being able to generate an expected value of $R(t) > 1$ is a minimal requirement for the adequacy of a proposed model. Takahata places greater emphasis on the predictive accuracy of specific assumptions that achieve greater fit than on overall fit. Being able to generate a high value of $R(t)$, however, is not sufficient for a model to be considered choice-worthy, as can be seen by Takahata's own evaluation of the models he proposes.

Takahata's first alternative pushed for a different assumption about how substitutions occurred. If multiple substitutions occurred, then Takahata could retain a Poisson process model, although it would be compound, not simple. The rates of substitution and the multiplicity of substitutions observed, however, were not as high as required for this model to produce significantly large values of $I(t)$. This option was thus rendered "unlikely" as a result of a poor empirical support for the specific assumption about multiple substitutions (Figure 1: B2). This explains why the multiple substitution model was not pursued, but not why the revision was made in the first place. An explanation and justification for this is suggested by Takahata's citing of six papers ranging from 1968 to 1985 that substitutions may be influenced by intramolecular interactions (Takahata 1987, 171). Multiple substitutions are a critical part of Kimura's original model as evidenced by the concern for estimating multiple substitutions. If several authorities suggest that perhaps our ideas about substitutions need revision, then the importance of this element warrants its attempted revision.

Takahata's second alternative follows Gillespie's lead in using a doubly Poisson process model, but adds to it a neutralist twist borrowing from work by Tomoko Ohta and Motoo Kimura on slightly deleterious mutations and on the effects of population bottlenecks (Ohta 1973, 1977; Kimura 1979). The emphasis this model places on deleterious mutations that arise during bottlenecks in the population makes the model dependent on the periods that each lineage spends in a bottleneck phase. For bottleneck events to cause significant variation in the rate among lineages, they would have to "vary considerably from lineage to lineage" (Takahata 1987, 176). However, Takahata, like Kimura, has assumed that all the lineages under study follow common rules, so mutation rates themselves are not allowed to differ among lineages. Takahata's emphasis on bottlenecks in this model also has the result that it can account for the large values of $R(t)$ only by creating another problem of explaining why bottlenecks af-

fect only some protein lineages when they all experience the bottleneck. These constraints and problems raised by the neutralist interpretation of the doubly stochastic Poisson model argue against its plausibility, according to Takahata.

Takahata's third alternative, the fluctuating neutral space model, was developed using a previously unused type of mathematical model (a time dependent renewal process), but it was able to account for the troublesome evidence using intragenomic causes sympathetic to the neutralist view of molecular evolution—changes in selective or functional constraints after substitutions. The idea behind this revision, however, seems to be closely tied to Gillespie's argument that it is changes in the action of natural selection that accounts for the higher rates.

Takahata clearly favors the fluctuating neutral space model, but admits that a satisfactory description of the mechanisms underlying the statistical models is still lacking. The heart of the difference between Takahata and Gillespie is in their basic assumptions about molecular evolution; the fluctuating neutral space model "emphasizes intragenic or intragenomic causes" while the episodic model "stresses external or environmental causes." Mathematically, the two models show different evolutionary patterns. In Takahata's words,

... in the fluctuating neutral space model, some lineages fortuitously undergo more substitutions than do others and it is this variation that inflates $R(t)$. This pattern of evolution may be exemplified by insulin genes of guinea pig and coypu (Kimura 1987). In the episodic clock model, on the other hand, the rate drastically changes many times along each lineage and it is this change that increases $R(t)$. Direct evidence for this pattern currently appears to be nil (Takahata 1988, 388).

So, according to Takahata, Gillespie has no empirical support for his assumption that there are rate changes in each lineage. Takahata's attack on Gillespie's assumptions is particularly important here, because the strength or plausibility of underlying assumptions is taken by both Takahata and Gillespie to be an important basis for choice among rival models. Remember that Takahata ruled out two possible neutralist models in part on the basis of the quality of their assumptions. Further, recall that Gillespie has shifted the grounds of assessment away from fit and toward independent support. Of course, Takahata used his own underdetermination strategy, but not by ignoring Gillespie's.

From the standpoint of manipulating underdetermination, Takahata's paper is particularly illuminating because he actually goes through a list of rivals to Gillespie's model. Each of Takahata's models can generate pre-

dictions of high values of $I(t)$ and so put the neutralist position on equal footing with regard to fit between model and data for values of $I(t)$. For each model, however, Takahata examines the price to be exacted, and only in the case of the fluctuating neutral space model decides that the price is not too high. The grounds for this decision, improved fit and independent support for assumptions, are thus revealed as open routes for choice between his and Gillespie's models. Takahata's underdetermination strategy shifts the grounds of assessment away from emphasizing overall fit with and predictive accuracy of values of $R(t)$ (Figure 1: B1, B5), the grounds on which Gillespie launched his attack, to the specific predictions of $I(t)$ generated by the kinds of assumptions and mechanisms embodied in each of the rival models resulting in a differently weighted combination of fit and independent support for assumptions (Figure 1: B1, B2, B5).

The outcome of the debate between Takahata and Gillespie was not to settle any part of the larger neutralist selectionist controversy. If anything, the manipulation of underdetermination in this debate strategically leveled the playing field by equating neutralist and selectionist models in terms of their ability to explain high values of $I(t)$. Given that neutralist models were judged to have an advantage in terms of the explaining rate constancy in a very parsimonious way, Gillespie's use of underdetermination represents a modest success for the selectionist perspective.⁹ Gillespie's manipulation of underdetermination to shift the grounds of the controversy did not resolve the dispute; rather, it used new phenomena (overdispersion) to drive a search for new models, but most importantly it shifted the standards by which those models would be evaluated.

5. Discussion and Conclusion.

The controversy between Gillespie and Takahata over the molecular clock reveals that scientists can manipulate underdetermination, or deploy an underdetermination strategy. By drawing on the multidimensional framework of theory assessment, scientists can re-combine and re-weight assessment strategies in response to underdetermination so that the grounds of assessment, terms, and course of dispute change. Gillespie began by claiming that both Kimura's neutral theory and his own selectionist model fit the relevant molecular data (about) equally well and are both predictively accurate (Figure 1: B1, B5). He has identified a case of epistemic underdetermination in which two assessment strategies are underdetermined for two models. The consequence of Gillespie's underdeter-

9. Throughout the 1990s neutralists, such as Tomoko Ohta, developed more sophisticated models of nearly neutral or slightly deleterious mutations to try to capture the overdispersion in the molecular clock. See Cutler 2000 for a review.

ination claim is that any choice between the two models must be made on other grounds. Gillespie begins the assessment process by giving Kimura's neutral theory an advantage; not only does it fit the data, it is also parsimonious (Figure 1: A6). And he admits that it is difficult to generate a parsimonious alternative of his selectionist model. But the neutral theory's advantage does not last long. Gillespie argues for the superiority of his selectionist model essentially by arguing that the theoretical assumptions of the neutral theory are flawed and that the selectionist assumptions of his own model have independent support (Figure 1: B2). For Gillespie, the independent support for underlying assumptions of his selectionist model trumps the neutral theory's parsimony; independent support for aspects of a model (Figure 1: B2) is given more weight than simplicity (Figure 1: A6).

When Takahata responds to Gillespie, he agrees that fit between model and data is important and also, given epistemic underdetermination, that the grounds of assessment must change. There is no question for Takahata that he will stay true to at least the basic framework of the neutral theory. So he needs a way of countering the independent support for the underlying selectionist assumptions claimed as an advantage for Gillespie's selection model. In doing so, he shifts the grounds of assessment off "independent support vs. parsimony." When Takahata revises the neutral theory he shifts the grounds for assessment to fit between its assumptions and values of $I(t)$, and better overall fit with values of $R(t)$. Takahata generates three options in revising the neutral theory's assumptions and chooses the third, i.e., the fluctuating neutral space option. Indeed, Takahata argues that Gillespie lacks evidence for his claim that rates fluctuate in each lineage and pushes the point that there is independent support for the revised assumptions of the neutral theory (Table 1: B2). According to Takahata, the support for the revised assumptions is stronger than the support for Gillespie's selectionist assumptions. Now what carries the most weight is the ability for the models under scrutiny to fit and accurately predict values of $I(t)$ in a way that generates a better overall fit of the model with the data for $R(t)$. Both the (revised) neutral theory and Gillespie's selectionist model are in line with the data. And Gillespie's selectionist assumptions have independent support. But the neutral theory has the advantage that its assumptions fit the data for values of $I(t)$ and that fit improves overall fit of the theory with the data for values of $R(t)$. For Takahata, that combination trumps Gillespie's independent support for the underlying assumptions of his selectionist model.

Gillespie and Takahata manipulate underdetermination by arguing that different combinations of differently weighted assessment strategies both create and break the "tie" the two models have regarding overall fit

with model and data in different ways. Gillespie and Takahata were able to construct (apparently quite good) arguments in favor of their respective models of the clock by manipulating the combination of assessment strategies and the weights of the specific strategies in those combinations. They adopted specific underdetermination strategies. Indeed, this point is precisely what we argued for at the outset of the paper. Where there are cases of underdetermination, such as the case of the molecular clock, scientists may direct the course and terms of dispute by playing off the multi-dimensional framework of theory evaluation in the ways we have described. This is because assessment strategies themselves are underdetermined. Within the framework of assessment, there are a variety of trade-offs between different strategies as well as shifting emphases on those strategies as different ones are given more or less weight in assessment situations. Our analysis focused on a small set of assessment strategies. Numerous other strategies delineated in Figure 1 may also be involved in the manipulation of underdetermination. If we are right that the underdetermination relation can be specified for epistemic, pragmatic, or sociological grounds for choice, then the sheer number of assessment strategies biologists have recourse to, the permutations of trade-offs, combinations, and weightings within combinations is enormous, and that makes for a broad range of underdetermination strategies. The case discussed here presents a very specific controversy illustrating a very specific case of scientists following underdetermination strategies. While we are confident that the history and sociology of science affords ample data for the thesis of this paper, it is a massive empirical enterprise to demonstrate the multitude of ways in which underdetermination may be manipulated.

We claim that when a strategy is underdetermined, scientists can change the dynamics of a controversy by making assessments using different combinations of evaluation strategies and/or weighting whatever strategies are in play in different ways. Following an underdetermination strategy does not end or resolve a scientific dispute. Consequently, manipulating underdetermination is a feature of controversy dynamics and not controversy closure.

References

- Bloor, David. 1981. "The Strengths of the Strong Programme." *Philosophy of the Social Sciences* 11: 199–213.
- Collins, Harry. 1985. *Changing Order*. London: Sage Publications.
- Cutler, David J. 2000. "Understanding the Overdispersed Molecular Clock." *Genetics* 154: 1403–1417.
- Darden, Lindley. 1991. *Theory Change in Science: Strategies from Mendelian Genetics*. New York: Oxford University Press.

- Dietrich, Michael. 1993. "Underdetermination and the Limits of Interpretive Flexibility." *Perspectives on Science* 1: 109–126.
- . 1994. "The Origins of the Neutral Theory of Molecular Evolution." *Journal of the History of Biology*, 27: 21–59.
- . 1998. "Paradox and Persuasion: Negotiating the Place of Molecular Evolution Within Evolutionary Biology." *Journal of the History of Biology* 31: 85–111.
- . 2006. "From Mendel to Molecules: A Brief History of Evolutionary Genetics." Pp. 3–13 in *Evolutionary Genetics: Concepts and Case Studies*, eds. Charles W. Fox and Jason B. Wolf. New York, NY: Oxford University Press.
- Franklin, Allan. 1986. *The Neglect of Experiment*. Cambridge: Cambridge University Press.
- . 1990. *Experiment: Right or Wrong*. Cambridge: Cambridge University Press.
- . 2002. *Selectivity and Discord: Two Problems of Experiment*. Pittsburgh, PA: University of Pittsburgh Press.
- Gillespie, John. 1978. "A General Model to Account for Enzyme Variation in Natural Populations. V. The SAS-CFF Model." *Theoretical Population Biology* 14: 1–45.
- . 1984a. "The Molecular Clock May Be an Episodic Clock." *Proceedings of the National Academy of Sciences* 81: 8009–8013.
- . 1984b. "Molecular Evolution Over the Mutational Landscape." *Evolution* 38: 1116–1129.
- . 1984c. "The Status of the Neutral Theory." *Science* 224: 732–733.
- . 1986a. "Natural Selection and the Molecular Clock." *Molecular Biology and Evolution* 3: 136–155.
- . 1986b. "Variability of Evolutionary Rates of DNA." *Genetics* 113: 1077–1091.
- . 1987. "Molecular Evolution and the Neutral Allele Theory." *Oxford Surveys of Evolutionary Biology* 4:10–37.
- Gillespie, John and Langley, Charles. 1979. "Are Evolutionary Rates Really Variable?" *Journal of Molecular Evolution* 13: 27–34.
- Goodman, M., Moore, G., and Matsuda, G. 1975. "Darwinian Evolution in the Genealogy of Hemoglobin." *Nature* 253: 603–608.
- Kimura, Motoo. 1968. "Evolutionary Rate at the Molecular Level." *Nature* 217: 624–626.
- . 1969. "The Rate of Molecular Evolution Considered from the Standpoint of Population Genetics." *Proceedings of the National Academy of Sciences* 63: 1181–1188.

- . 1979. "The Neutral Theory of Molecular Evolution." *Scientific American* 241: 98–127.
- . 1983. *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press.
- . 1987. "Molecular Evolutionary Clock and the Neutral Theory." *Journal of Molecular Evolution* 26: 24–33.
- King, Jack and Jukes, Thomas. 1969. "Non-Darwinian Evolution." *Science* 164: 788–798.
- Kuhn, Thomas. 1977. *The Essential Tension*. Chicago: University of Chicago Press.
- Langley, Charles and Fitch, Walter. 1974. "An Examination of the Constancy of the Rate of Molecular Evolution." *Journal of Molecular Evolution* 3: 161–177.
- Latour, Bruno. 1992. "One More Turn After the Social Turn." Pp. 272–294 in E. McMullin ed., *The Social Dimensions of Science*. Notre Dame: University of Notre Dame Press.
- Laudan, Laurence 1990. "Demystifying Underdetermination." W. Savage ed., *Scientific Theories. Minnesota Studies in the Philosophy of Science* 10. Minneapolis: University of Minnesota Press.
- Lindley, D. 1965. *Introduction to Probability and Statistics From a Bayesian Viewpoint. Part 1: Probability*. Cambridge: Cambridge University Press.
- Lipton, Peter 1991, *Inference to the Best Explanation*. London: Routledge.
- Lloyd, Elisabeth A. 1987. "Confirmation of Evolutionary and Ecological Models." *Biology and Philosophy* 2: 277–293.
- . 1988. *The Structure and Confirmation of Evolutionary Theory*. Princeton: Princeton University Press.
- Longino, Helen. 1990. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton: Princeton University Press.
- Mayo, Deborah. 1996. *Error and the Growth of Experimental Knowledge*. Chicago: University of Chicago Press.
- McMullin, E. 1982. "Values in Science." Pp. 3–28 in *PSA 1982*, vol. 2. P. Asquith and T. Nickles ed. East Lansing, MI: Philosophy of Science Association.
- Morgan, Gregory. 1998. "Emile Zuckerkandl, Linus Pauling, and the Molecular Evolutionary Clock, 1959–1965." *Journal of the History of Biology* 31: 155–178.
- Newton-Smith, William. H. 1981. *The Rationality of Science*. Oxford: Oxford University Press.
- Ohta, T. 1977. "Extension to the Neutral Mutation Random Drift Hypothesis", *Molecular Evolution and Polymorphism*. Motoo Kimura, ed. Mishima, Japan: National Institute of Genetics Publications.

- Ohta, T. and Kimura, M. 1971. "On the Constancy of the Evolutionary Rate of Cistrons." *Journal of Molecular Evolution* 1: 18–25.
- Pickering, Andrew. 1984. *Constructing Quarks*. Edinburgh: Edinburgh University Press.
- Quine, W. V. O. 1953. *From a Logical Point of View*. 2nd edition. New York: Harper and Row.
- Sarich, V. and Wilson, A. 1967. "Immunological Time Scale for Hominid Evolution." *Science* 158: 1200–1202.
- Rudge, David W. 1998. "A Bayesian Analysis of Strategies in Evolutionary Biology." *Perspectives on Science* 6: 341–360.
- . 2001. "Kettlewell from an Error Statistician's Point of View." *Perspectives on Science* 9: 59–77.
- Schaffner, Kenneth F. 1993. *Discovery and Explanation in Biology and Medicine*. Chicago: University of Chicago Press.
- Skipper, Jr., Robert A. 2000. *The R. A. Fisher-Sewall Wright Controversy in Philosophical Focus: Theory Evaluation in Population Genetics*. Dissertation: University of Maryland, College Park.
- . 2002. "The Persistence of the R. A. Fisher-Sewall Wright Controversy." *Biology and Philosophy* 17: 341–367.
- . 2004a. "The Heuristic Role of Sewall Wright's 1932 Adaptive Landscape Diagram." *Philosophy of Science* 71: 1176–1188.
- . 2004b. "Calibration of Laboratory Models in Population Genetics." *Perspectives on Science* 12: 369–393.
- Sober, E. 1994. "Let's Razor Ockham's Razor." Pp. 136–157 in *From a Biological Point of View*. Edited by Elliott Sober. Cambridge, UK: Cambridge University Press.
- Solomon, M. 2001. *Social Empiricism*. Cambridge, MA: MIT Press.
- Suppes, Patrick. 1962. "Models of Data." Pp. 252–261 in *Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress*. Edited by E. Nagel, P. Suppes, and A. Tarski. Palo Alto, CA: Stanford University Press.
- Takahata, Naoyuki. 1987. "On the Overdispersed Molecular Clock." *Genetics* 116: 169–179.
- . 1988. "More on the Episodic Clock." *Genetics* 118: 387–388.
- van Fraassen, Bas C. 1980. *The Scientific Image*. New York: Oxford University Press.
- Wilson, A., Carlson, S. and White, T. 1977. "Biochemical Evolution." *Annual Review of Biochemistry* 46: 573–639.
- Wimsatt, William C. 1980. "Reductionistic Research Heuristics and Their Biases in the Units of Selection Controversy." Pp. 213–259 in T. Nickles ed., *Scientific Discovery*, Vol. 2: Case Studies. Dordrecht, The Netherlands: Reidel.

- Zuckermandl, E. and Pauling, L. 1962. "Molecular Disease, Evolution, and Genetic Heterogeneity." *Horizons in Biochemistry*. Kasha, M. and B. Pullman, eds. New York, NY: Academic Press.
- . 1965. "Evolutionary Divergence and Convergence in Proteins." *Evolving Genes and Proteins*. V. Bryson and H. Vogel, eds. New York, NY: Academic Press.