2005

Modeling mechanisms

Stuart Glennan
Butler University, sglennan@butler.edu

The original publication appeared at:

Modeling Mechanisms

Stuart Glennan

Department of Philosophy and Religion

Butler University

4600 Sunset Avenue

Indianapolis, IN 46228

Abstract

Philosophers of science increasingly believe that much of science is concerned with understanding the mechanisms responsible for the production of natural phenomena. An adequate understanding of scientific research requires an account of how scientists develop and test models of mechanisms. This paper offers a general account of the nature of mechanical models, discussing the representational relationship that holds between mechanisms and their models as well as the techniques that can be used to test and refine such models. The analysis is supported by study of two competing models of a mechanism of speech perception.

Keywords

The past several years have seen the growth of a mechanisms movement in the philosophy of biology and in the philosophy of science more generally. Mechanist philosophers of biology believe that mechanisms are the key to understanding biological phenomena. Perhaps because of the realist tendencies of the philosophers involved, most of the literature has focused on the properties of mechanisms themselves and has not said much about the relationships between mechanisms and their models or theoretical representations. My goal in this paper is to redress this deficiency and to sketch an account of the relationship between mechanisms and our models of them. I will describe a particular kind of model, called a , and will discuss the relationship between this sort of model and other models that have been discussed in the philosophical literature. I will then offer a detailed case study of two competing models of a phenomenon in the field of speech perception. This case study will illustrate several important issues regarding mechanical models. It will give substance to the claim that models are related to mechanisms via similarity relations; it will demonstrate how it is possible to make inferences about mechanisms when one cannot directly study the properties of the mechanism's parts; and it will illustrate the connection between mechanistic inference and the problem of underdetermination.

I. Mechanisms and Mechanical Models

There is a large literature devoted to the nature of biological models and the strategy of model building in biology (e.g., Levins, 1968; Lewontin, 1974; Wimsatt, 1987), but such discussions generally have proceeded without any detailed analysis of what models are. To the extent that there is an accepted "theory of models," philosophers of biology (e.g., Beatty, 1981; Lloyd, 1994) have adopted the view of models developed by advocates of the semantic view of theories (see Suppe, 1974; 1989)

Most advocates of the semantic view of theories characterize models in terms of state spaces.[1] According to the state space approach, the state of a physical system can be characterized by a set of state variables—variables measuring the values of various physical magnitudes. The set of logically possible states of the system can be identified with the set of all possible combinations of values of each of the variables, and these combinations in turn are treated as vectors in the state space. The dynamical behavior of a modeled system can be characterized in terms of the trajectory of the system through this vector space over time. Physically possible changes in the state of a system may be characterized by laws of succession, while physically possible combinations of values of state variables can be characterized by laws of coexistence (Suppe, 1989).

---

[1] Others (e.g., Suppes, 1960) characterize models as set-theoretic structures, but this approach has not received much discussion among philosophers of biology.

The state space representation is powerful and flexible, and it is hard to imagine anything that scientists might call a model that could not be represented in this way. Indeed, this type of representational format is borrowed from physics, and has been used in biology as well (Lewontin, 1974). But while state space models provide a convenient formalism, this characterization is too abstract to give much insight into the nature of the relationship between a model and the system modeled or into the strategies of model building, testing and revision.

Bill Wimsatt has claimed that "[a]t least in biology, most scientists see their work as explaining types of phenomena by discovering mechanisms . . .' (Wimsatt, 1972, p. 67). If Wimsatt is right about this, it seems plausible that most of the models developed by biologists will be models of mechanisms.[2] Accordingly, we can develop a deeper understanding of the nature of modeling in biology by developing a theory of mechanical models.

A mechanical model is (not surprisingly) a model of a mechanism, so to present an account of mechanical models, a first step is to offer some analysis of the concept of mechanism. In an earlier paper (Glennan, 2002) I have offered the following definition of a mechanism:

(M)   A mechanism for a behavior is a complex system that produces that behavior by the interaction of a number of parts, where the interactions between parts can be characterized by direct, invariant, change-relating generalizations.[3]

While I shall not defend this definition here, a few clarifications are in order. In the first place, mechanisms <u>underlie behaviors</u>. The behavior that the mechanism underlies, or, more simply, the behavior of the mechanism, is what the mechanism does. A heart is a mechanism for pumping blood, a Coke machine is a mechanism for dispensing Cokes in return for coins, and so on. Craver and Darden (2001) emphasize the same point about mechanisms, referring to what the mechanism does as the "phenomenon" that the mechanism produces.

Mechanisms can have behaviors of a variety of forms. Perhaps the most familiar are mechanisms that respond to inputs with outputs (e.g, Coke machines or neurons), but

---

[2] This claim I would argue is true of many domains beside biology. It should be true of any domain where explanations or predictions are sought by examining the organization and interaction of parts of complex systems – including much of physics and chemistry, psychology, and economics.

[3] This definition reflects a slight modification from a definition I proposed earlier (Glennan, 1996). I have replaced an appeal to laws with the more specific notion of a "direct, invariant, change relating generalization. This analysis of mechanism is substantially in agreement with a number of other analyses, including those of Bechtel and Richardson (1993) and Machamer, Darden and Craver (2000).

there are also, for instance, mechanisms that maintain systems in stable states or mechanisms that produce a periodic behavior (e.g., the mechanism that explains Old Faithful's eruptions. Many mechanisms are, like hearts and Coke machines, designed to behave as they do, but one can equally well talk about mechanisms underlying behaviors that, like the eruptions of Old Faithful, are not the product of design.

The second part of the definition that requires explanation is the notion of 'direct, invariant, change-relating generalizations.' These generalizations characterize the causal interactions between parts of mechanisms. They function as laws in Mitchell's (1997) pragmatic sense, though, unlike laws on the positivist conception, they depend in essential ways on particulars, are subject to breakdowns, and are not of unrestricted scope. The term 'change-relating generalization' is borrowed from Woodward (2000), to indicate a generalization that characterizes how a change in the property of one or more parts <u>brings about</u> a change in the property of another part. It is meant to exclude generalizations that hold true in virtue of common causes, global constraints, conservation principles, and the like. If models were represented in terms of laws describing transitions in a state space, these laws would be the laws of succession. The most typical form for this kind of generalization would be a differential equation, though many change relating generalizations might describe discrete changes. The stipulation that the generalizations be <u>direct</u> is meant to rule out generalizations that describe more remote effects. For instance, depressing the button on the Coke machine causes a Coke to be dispensed, but only via the complex interaction of a number of parts. Hence, there is no direct generalization connecting button presses to the dispensing of Cokes.

An important feature of this conception of mechanism is that it is hierarchical (cf. Glennan 1996; Machamer, Darden and Craver, 2000; Craver, 2002). The parts of mechanisms may themselves be mechanisms, and the change-relating generalizations connecting those parts may themselves be mechanical processes. The behavior of the mechanism can often be described by a change-relating generalization. We might describe the behavior of a simple Coke machine in the following change-relating generalization: "Whenever $1 is inserted and the button marked 'Coke' is pressed, a Coke appears in the slot at the bottom of the machine." The fact that the statement is true is explained by facts about how the Coke machine works. Generalizations of this kind I call <u>mechanically explicable</u> (cf. Glennan 1996; 1997). As a less homely example, Mendel's first law, which we might state as "Whenever a parent is heterozygous at a locus, the proportion of gametes produced by the parent carrying each allele will be .5," is also mechanically explicable. It follows from facts about the mechanisms used in organisms to produce gametes.

There is thus a two-way relationship between invariant generalizations and mechanisms. First, reliable behavior of mechanisms depends upon the existence of invariant relations between their parts, and change-relating generalizations characterize

4

these relations. Second, many such generalizations are mechanically explicable, in the sense that they are just generalizations about the behavior of mechanisms. A single generalization can both be explained by a mechanism and characterize the interaction between parts of a larger mechanism. For instance, Mendel's first law is explained by the standard mechanisms of gamete formation, while evenly segregated gametes, produced in accordance with Mendel's first law, play a part in larger mechanisms, for instance the mechanism by which deleterious alleles will be (generally) driven to extinction.

Given this understanding of what a mechanism is, we now can define a mechanical model:

(MM) A mechanical model consists of (i) a description of the mechanism's behavior (the behavioral description); and (ii) a description of the mechanism that accounts for that behavior (the mechanical description).

The two-part characterization of a mechanism, in terms of a behavior and the mechanism that produces it, leads naturally to a two-part characterization of a mechanical model. The behavioral description is a description of the overall behavior of a mechanism. The mechanical description is a description of the mechanism's parts and their functional arrangement. Another way to put it is to say that the behavior description tells one what a mechanism is doing, while the mechanical description tells one how the mechanism does it.[4]

The sense in which the two parts of a mechanical model are descriptions requires clarification. They are not descriptions in the sense of 'descriptive phrases' familiar from Russell's theory of descriptions, nor indeed can they be identified as sets of sentences or as syntactically defined entities of any kind. Rather, they are semantic entities, in the sense that there can exist many syntactically different formulations of the same description. For instance, the behavioral description of an ecological model where a population oscillates around a carrying capacity could be presented either diagrammatically or by means of differential equations while still counting as the same description. For one to have different behavioral descriptions, the behavior described

---

[4] This way of expressing the relationship suggests a close affinity between this distinction and Marr's (1982) distinction between computational and algorithmic theories.

A number of readers have suggested that the behavioral description should not be part of the mechanical model. I have adopted the two-part approach principally to emphasize that the behavioral description is itself a model that is only similar to the modeled behavior in certain degrees and respects. In evaluating a model one must evaluate the adequacy of both the mechanical description and the behavioral description. One might alternately take the view that these descriptions are separate models, with the behavioral description being a "model of data" in Suppes' (1962) sense.

must be different. Thus, if in one model the oscillation of the population dampens while in another it does not, we have different behavioral descriptions, and hence different models.

One way to look at the relationship between the behavioral and mechanical descriptions is as a distinction between a description of the external behavior of a mechanism and a description of its internal workings, but sometimes the spatial terminology must be construed metaphorically. In their discussion of strategies for discerning mechanical structure, Bechtel and Richardson (1993) have usefully distinguished between decomposition and localization. Localization is a spatial notion – where one identifies parts of mechanisms via their locations – while decomposition identifies parts of mechanisms by their functional relations. When functions are spatially localizable, decomposition and localization will yield the same analysis, and as Bechtel and Richardson emphasize, localization can provide a powerful heuristic for the discovery of functional decompositions. Functional localization can fail, however, when the parts responsible for particular functions are distributed in space or are not stable in location over time. It should be clear, given that (M) defines mechanisms in terms of functional relations between parts, that in such cases it is the functional structure revealed by decomposition that is constitutive of mechanism.

The most important difference between mechanical models and state-space models is that mechanical models have two parts. According to the state-space conception, the possible states of the system modeled are represented by the set of vectors in a state space, and a particular model can be understood as a curve through this space. As a very simple example, consider how we might construct a state-space model of an analog watch. The state of the watch at a given time can be characterized by two variables, one representing the position of the hour hand and the other the position of the minute hand. One can then plot the evolution of the state of the watch through time as a periodic curve in this space. From the mechanistic point of view, this description of the watch is only half a model of the watch. It is, in particular, a description of the behavior of the watch. What is lacking is a description of the mechanism that produces this behavior. A mechanical description would characterize how the various parts of the watch (the battery, the quartz crystal, the hands, the internal gears, etc.) cause the hands to move in the way characterized by the behavioral description.

The internal workings of the mechanism can be represented using the state space approach as well. The relevant properties of the interacting parts of the mechanism could be coded as state variables. Laws of succession could be specified that express how changes in values of state variables representing properties of one part bring about changes in the values of state variables representing properties of directly linked parts. In this case, the laws of succession are the change-relating generalizations referred to in (M). What this example shows is that the notion of a mechanical model is more restricted than

that of a state-space model. Whether a state space model is a mechanical model depends upon what state variables are chosen, and whether the laws of succession used to characterize the state changes represent direct causal interactions between the parts of the mechanism.

It is possible to formulate a mechanical model using a state space representation but that not all state space models are mechanical models. The requirements for a model being a description of a mechanism place substantive constraints on the choice of state variables (such as the fact that state variables should refer to properties of parts), parameters, and laws of succession and coexistence. The satisfaction of these additional constraints is what accounts for the explanatory power of mechanical models.   The division between the behavioral description and the mechanical description is analogous to the division between explanandum and explanans. The mechanism characterized by the mechanical description brings about, and hence explains, the behavior characterized by the behavioral description. [5] If the behavioral description is a statement of a law (in the weak sense described above), then that law is mechanically explicable.

Two further points should be made which will have significant implications for the subsequent discussion of testing mechanical models. First, notice that the concept of a mechanism's behavior generally presupposes a concept of normal functioning. When one describes the behavior of a mechanism, one describes how it will behave if it is not broken. For instance, in describing the behavior of the watch in terms of the periodic rotation of hands, one presupposes that the watch's battery has not worn out and many other things of that sort. The "if it not broken" clause is a kind of ceteris paribus clause for behavioral descriptions.   This idea of "normal function" is required even for mechanisms that are not the product of design or selection. If, for instance, one describes the behavior of the El Ninõ mechanism, the description presupposes that the normal mechanism by which El Ninõ produces its effects is not disrupted by exogenous factors (like the earth being hit by a large asteroid).

The second point is that there is a one-many relationship between behavioral and mechanical descriptions. This is because the same behavior can be produced by different mechanisms. A spring wound watch and a Quartz crystal watch will (relative to most descriptions) behave the same way, even though the mechanism that produces the

---

[5] The division is only analogous. I am adopting an ontic conception of explanation in which the locus of explanatory insight is "in the objects."   It is the fact that the mechanism brings about the behavior, not that the mechanical description entails the behavioral description, which yields the explanation (cf. Glennan, 2002). At the same time, we can only achieve understanding of a phenomena with the help of an adequate representation of the phenomena and of the mechanism that brings it about.

behavior will be quite different. We shall take up the question of how to choose between competing models of the same behavior in the last section of this paper.

One of the most controversial questions in the recent literature on modeling concerns the relationship between models and the real systems they model. At least since Levins (1968) it has been widely recognized that the models scientists use are generally false, in the sense that they are based on clearly false idealizing assumptions. My view is that the key to understanding how false models can be genuinely explanatory is to realize that even a false model can provide a partial representation of a modeled system. I follow Ron Giere (1988) in claiming that the representation relation exists in virtue of similarities in various degrees and respects between the model and the modeled system. Giere (1999) has usefully compared models to maps. The most salient feature of this analogy is that maps, like models, represent only certain aspects of the region that is mapped. Hence, there can be different maps of the same terrain, with each map bearing similarity relations to the terrain only in certain degrees and respects.[6]

Let me conclude with a point regarding terminology. The notion of a mechanical model that I have developed is closely related to what Machamer, Darden and Craver (2000) have called a 'mechanism schema.' While I concur with much of what they say regarding mechanism schemata, I prefer the term 'mechanical model' for several reasons. Most importantly, scientists often use the term 'model' to refer to thing entities I call mechanical models. Also, Machamer, Darden and Craver introduce the term 'mechanism schema' in part in contrast to what they call a 'mechanism sketch.' Their idea is that a mechanism sketch is an incomplete schema – one in which the component entities and activities have not been completely identified. Sketches are filled in to provide schemata, and schemata are what are required for mechanistic explanation. The problem with the sketch/schema distinction is that it makes into a two-step process what is in reality a process of continuous model articulation.

II.   Models of Vowel Normalization

In this section I shall apply the analysis of mechanical models to two models from the cognitive psychology literature on speech perception. These models seek to explain the auditory processing mechanisms by which listeners adjust to variations in the acoustical properties of different talkers' vowel sounds. The process of adjusting to this variation is called <u>vowel normalization</u>. Because this case is not well known to philosophers, I will begin with some background on the basics of speech perception. I will then describe the models, demonstrating how they both satisfy the characterization of

---

[6] There is a voluminous literature on idealization in modeling and the nature of representation. See, e.g., Wimsatt (1987), Morgan and Morrison (1999), Bailer-Jones (2003) and Odenbaugh (under review).

a mechanical model given in the previous section. In the concluding section of the paper I will discuss experimental work that has been done to evaluate these models, showing how this case supports the view that models represent mechanisms only in degrees and respects.

While many of the points about representation and testing could have been made using a variety of examples, I have chosen this case for two reasons. First, unlike most paradigm cases in the literature, the parts of these putative mechanisms are, at least at this point, difficult to localize. I want to show how it is possible to construct and test a mechanical model even in such cases. Second, and relatedly, this case illustrates how it is possible to evaluate competing models of the same phenomena even when one can't directly identify and manipulate the parts of the mechanism responsible. I will argue that these methods, which I call methods of indirect testing, provide us with some grip on a methodologically interesting version of the underdetermination problem.

Researchers in speech perception seek to understand the mechanisms by which listeners transform acoustic signals into sequences of phonemes.[7] Acoustically speaking, a speech signal is just a sound wave. One way to characterize a speech signal is with an oscillogram, which is a plot of the wave's amplitude over time. An alternative way to characterize a wave is with a continuous spectogram. In a continuous spectogram, the sound wave is decomposed into frequency components that are analyzed continuously from short snippets of the speech signal. The result is a three-dimensional plot representing intensity (wave amplitude) at each frequency over time. The majority of acoustical information required for vowel phonological recognition is coded by the wave spectrum, and in particular by peaks (local maxima) of intensity in the spectrum called formants.

Normal (i.e., non-whispered) vowels and other voiced speech sounds are produced by passing vibrating air from the vocal cords, which function as a forced harmonic oscillator, through the articulatory tract (chiefly the mouth and nose), which functions as resonator. By altering physiological features of the articulatory tract (the openness of the mouth, the position of the tongue, etc.), a talker can change the resonant frequencies of the tract.

The configuration of the articulatory tract at any given time can be characterized by a frequency response curve. The local maxima in this curve (i.e., the frequencies at which the resonator resonates) are the formants. Physiologically, formant positions are determined chiefly by the shape of the mouth and lips and the position of the tongue. Formants are numbered from lowest to highest (F1, F2, etc.) according to the frequency

---

[7] For further background on the production and perception of speech, the reader may wish to consult (Moore, 2003).

of their peak resonance. The frequency of the formants (especially F1, F2 and F3) turns out to be a major determinant in the perception of different phonemes. Besides formant frequencies, the fundamental frequency of the wave (which is the frequency of oscillation of the vocal cords), called F0, also influences phoneme perception.

The vowel perception models I shall discuss are concerned with explaining the ability of American English speaking listeners to identify English language vowels. Table 1 lists the International Phonetic Alphabet (IPA) symbols for the ten English vowels studied in these models:

IPA Symbol	Sample Word
i	heed
I	hid
E	head
Θ	had
℘	hud
a	hod
□	hawed
Y	hood
u	who'd
ə ⇓	heard

Table 1: IPA Symbols for American English Vowels[8]

The most significant pieces of information used by listeners to recognize vowel sounds are the frequencies of the first and second formants. Most listeners accurately recognize artificial vowels that are constructed by masking all but F1 and F2 cues from naturally produced vowels. Differences in F1 and F2 intensities also correspond fairly closely to different places of articulation of vowels in the articulatory tract. Front vowels (i.e., vowels where the tongue is positioned in the front of the mouth) like /i/ have relatively higher F2 frequencies than back vowels like /u/. Low vowels (i.e., vowels where the tongue is positioned low in the mouth) like /æ/ have relatively higher F1 frequencies than high vowels like /i/. Using F1 and F2 frequencies it is possible to construct for a map of the speaker's vowel space. An F1/F2 map for an average talker is pictured in figure 1:

---

[8] For more on phonetic vowels, including information on the articulatory features that produce them, the reader may wish to consult the website of the International Phonetic Association (http://www2.arts.gla.ac.uk/IPA/ipa.html).
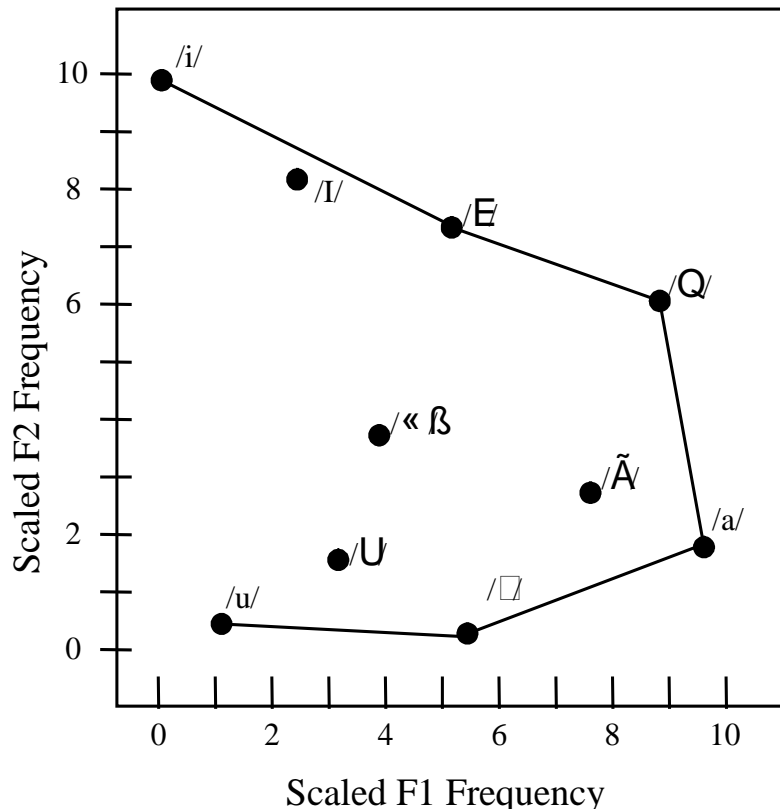
Figure 1: An Average Talker's Vowel Space[9]

The difficulty with this simple characterization of vowel perception is that it does not take into account variations in the vowel spaces across different talkers. Because different talkers' F1 vs. F2 maps are different, signals that are similar with respect to F1 and F2 frequencies can represent different vowels in different talkers. The process by which listeners adjust to different talkers' vowel spaces is called <u>vowel normalization</u>.

In the remainder of this section, I describe two competing models of vowel normalization, one proposed by L. J. Gerstman (1968) and the other proposed by A. K. Syrdal and H. S. Gopal (1986).[10] While my primary concern will be with vowel

---

[9] This graph is produced from average F1 and F2 frequencies collected by Peterson and Barney (1952) and scaled to values between 0 and 10 by Gerstman (1968). This graph is identical to one that appears in Gerstman 1968 except that I have replaced Gerstman's notation for vowels with standard IPA symbols.

[10] Unlike Syrdal and Gopal, Gerstman's primary concern was to develop models for machine recognition of human vowels, but the sort of information appealed to by Gerstman is likely involved in human perception. In this paper we shall, following Nusbaum and Morin (1992), consider Gerstman's approach as a model for human vowel recognition.

normalization itself, it is not possible to examine this component of the auditory system in isolation. To evaluate different models of vowel normalization, one must first integrate them into models of the larger recognition process. Models of normalization can be compared indirectly by comparing models of vowel recognition that incorporate different models of normalization. The two models that I will describe are thus not models of vowel normalization alone, but models of the vowel recognition mechanism that incorporate a normalization component.

These models are competing because they represent alternative hypotheses about what mechanism underlies a particular behavior. The two models have a common behavioral description, but differ in their mechanical description. To describe, and ultimately evaluate, these models, we must first examine this common behavioral description.

In the broadest sense, the behavior of a vowel recognition mechanism can be described by a function from acoustic signals to vowel tokens. The acoustic signals of interest are segments of human-produced speech that contain vowels. The study of human vowel recognition is complicated considerably by two problems. In the first place, acoustic properties of vowels can be altered by the phonetic context (i.e., what consonants surround the vowel). Second, human subjects will use features of the lexical, semantic and pragmatic context to provide additional information to aid in vowel recognition. In order to isolate the acoustic features used in vowel recognition, the authors focused on the analysis of a limited data set produced originally by Peterson and Barney (1952). Peterson and Barney recorded vowels produced by 76 talkers (a mixture of men, women and children). Each talker read from a randomly ordered list containing two repetitions of the ten words listed in table 1. Peterson and Barney then made spectrographic measurements of each vowel, measuring values of F0, F1, F2 and F3. Finally, they tested the intelligibility of these vowels by playing them to twenty-six listeners. 1199 of the 1520 were unanimously recognized.

With respect to this data set, human listeners display good but not complete accuracy in recognizing vowels. This level of performance with respect to the Peterson and Barney data set serves as the (partial) behavioral description of the "average" human vowel normalization mechanism. A model of vowel normalization should identify vowels with approximately the same degree of accuracy as human listeners do. Both the Gerstman and Syrdal/Gopal models satisfy this requirement. The two models therefore have roughly the same behavioral description, but they offer different accounts of the mechanism responsible for this behavior.

A.   The Gerstman Model

The basic idea behind the Gerstman model is that as a talker begins to speak the listener's auditory system acquires information about the F1 vs. F2 position of a small

number of vowels, called point vowels, which are used calibrate (or "tune") the recognition mechanism to that talker's particular vowel space. The model suggested by Gerstman is composed of four parts, which I shall call the formant frequency analyzer, the calibration mechanism, the normalization mechanism and the identification mechanism. The frequency analyzer is the part of the system that analyses a signal into a frequency spectrum and identifies formants. The calibration mechanism analyzes a calibration signal consisting of an initial segment of a particular talker's speech in order to determine the dimensions of that speaker's vowel space. The normalization mechanism transforms absolute formant frequencies F1 and F2 into frequencies F1' and F2' that are normalized to eliminate variations between different speakers' F1 vs. F2 spaces. The identification mechanism uses F1' and F2' to identify the vowel token. The structure of the overall mechanism is indicated in figure 2:
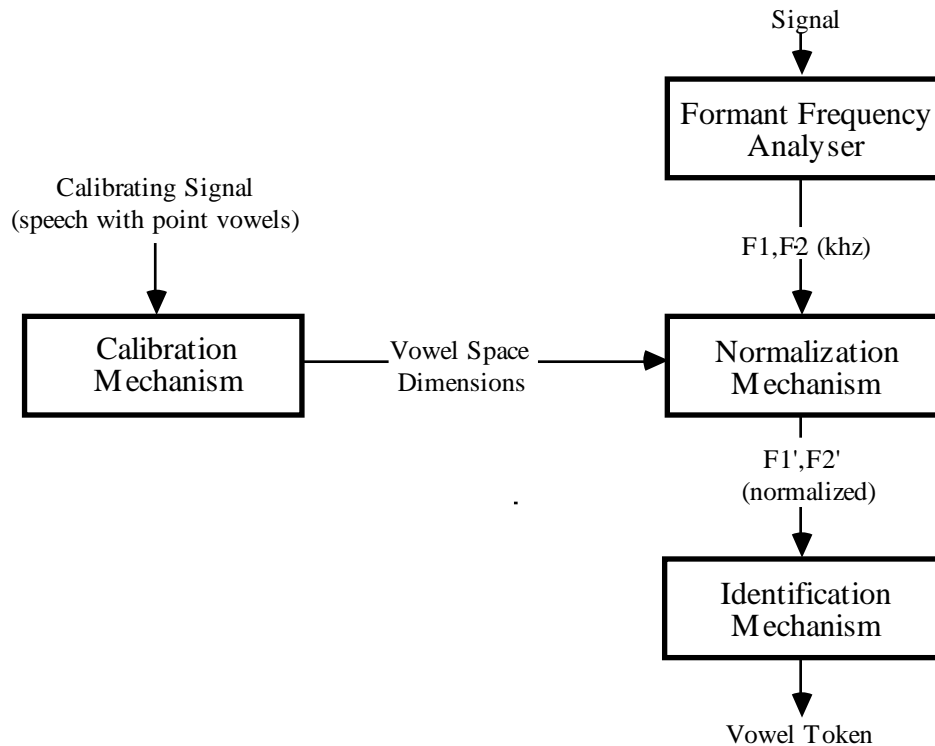

Figure 2: The Gerstman Vowel Normalization Model

The distinguishing feature of the Gerstman model is that the normalization mechanism requires certain parameters that are not determined by the vowel signal being processed. It is what Nusbaum and Morin (1992) call a "contextual tuning" mechanism. What is needed to tune the mechanism are signals that provide the minimum and maximum values of F1 and F2 for a given speaker's vowels. These are given by the point vowels (which are the leftmost, rightmost, topmost and bottommost vowels shown in figure 1). Using the F1 and F2 values for these vowels (the vowel space dimensions), the normalization mechanism can produce a normalized value of F1 and F2 for subsequent

vowel tokens. The identification mechanism works by comparing the position of the normalized vowel (F1', F2') in the scaled vowel space (see figure 1) with the position of average vowels within that space.

A mechanical model is more than just a set of generalizations about the external behavior of a system; a model purports to describe the structure of the mechanism that accounts for the behavior. Because the Gerstman model correctly classifies those vowel tokens which are unanimously categorized by human listeners, it passes the first test of accounting for the behavior of the system, at least insofar as identifying vowels in the Peterson and Barney data set with accuracy comparable to humans counts as predicting the behavior of the actual auditory mechanism. But beyond this, what claims has Gerstman made about the internal structure of the auditory mechanism?

Figure 2 summarizes the claims about the structure of the mechanism. Acceptance of the model involves commitment to the view that there are four distinct components of the auditory mechanism that exhibit the behaviors described in the above paragraphs. The internal structure of these components has been left largely unspecified. For instance, the calibration mechanism must in some way acquire F1 and F2 frequencies for point vowels, but Gerstman has not specified how these vowels are identified. They would have to be identified by a mechanism that either normalizes vowels in a different way or does not normalize at all. There are a number of ways in which the calibration mechanism might work: it might rely on other features of the vowel token besides F1 and F2 to identify unnormalized vowels; alternatively, it might rely on lexical or pragmatic information. One might for instance suppose that a listener might rely on the fact talkers say fairly standard things when they first begin speaking ("Hello!" etc.) and use this information to help disambiguate any vowels that are initially ambiguous. A major consideration in evaluating the Gerstman model is whether or not one can give plausible accounts of how this and the other parts of the mechanism hypothesized by that model might work.

B.   The Syrdal/Gopal Model

Syrdal and Gopal (1986) have developed a model of vowel recognition that uses a different normalization model. Unlike the Gerstman model, the Syrdal/Gopal model suggests that each vowel is self-normalizing. In other words, the model suggests that the F1 and F2 values of each vowel token can be normalized by using other parts of the acoustic signal and do not need tuning parameters extracted from the context of the utterance. Syrdal and Gopal's model is represented schematically in figure 3:
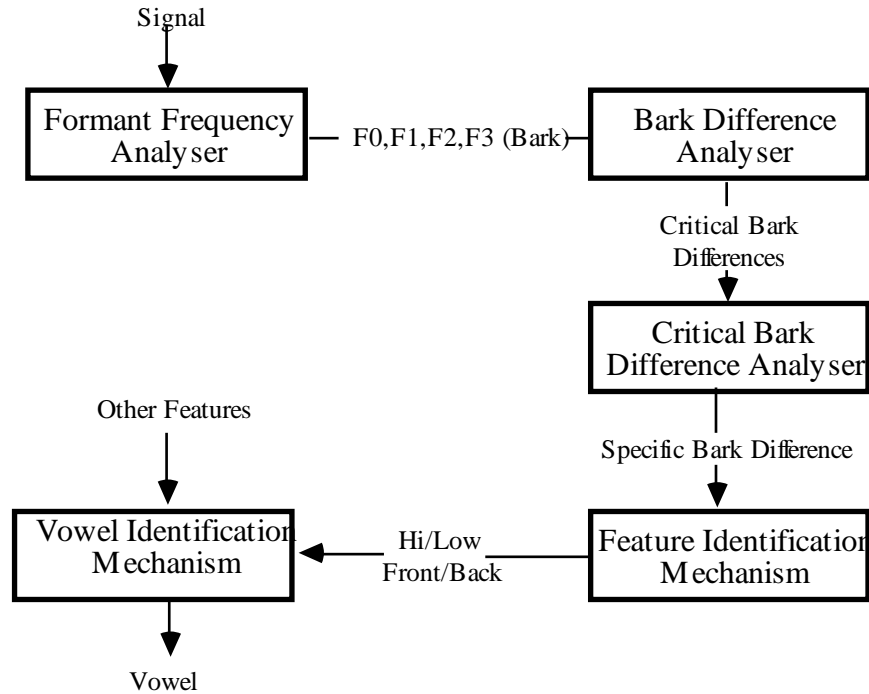
Figure 3: The Syrdal/Gopal Vowel Normalization Model

Processing of each vowel, according to this model, occurs in five stages. First, the acoustic signal is analyzed into formant pitches measured using a scale called the Bark scale.[11] Second, differences between F3 and F2 and between F1 and F0 are calculated. Third, these differences are classified into one of two categories depending upon whether or not their difference exceeds a critical distance of 3 Bark. Syrdal and Gopal show that this twofold binary classification is related to the vowel's phonetic features. The relation is given by the following map:

	< 3 Bark	> 3 Bark
F3-F2	front	back
F1-F0	high	mid, low

Since this classificatory schema only distinguishes four possible categories, it is not by itself sufficient to uniquely identify the ten different American vowels being considered. Additional information is required to identify the vowels. Syrdal and Gopal have analyzed critical bark differences in several other dimensions (F2-F1, F4-F2 and F4-F3) and determined that these measures could be used reliably to further discriminate between all but three pairs of vowels (/E/ and /æ/, /u/ and /Y/, /a/ and /□/). Other kinds of information (like vowel duration) can also be used to discriminate vowels.

---

[11] Perceived pitch does not vary linearly with actual frequency of pure tones. The Bark scale is one of a number of non-linear scales. The significance of the Bark scale is discussed in the final section of this paper.

Although Syrdal and Gopal think there is good evidence that the human auditory system uses binary feature classifications to minimize inter- and intraspeaker variability with respect to the production of vowels, they do not commit themselves to one particular computational model for vowel recognition. Instead they categorize vowels using a number of different features including critical bark differences, absolute bark values and vowel duration. They do not try to argue which of these techniques (if any) the human auditory system uses for feature discrimination. For this reason I have left some of the inputs in the vowel identification mechanism in figure 3 unspecified. These details turn out not to be significant for the kinds of tests described in the next section. What is critical is the general approach to normalization and identification. Syrdal and Gopal believe that these discriminations can be made based upon the structure of the signal being identified, regardless of what particular acoustic features are actually used to make binary discriminations. This is the sense in which the Syrdal/Gopal model is self-normalizing.

The Gerstman and Syrdal/Gopal models illustrate several noteworthy features of mechanisms and mechanical models. The most important characteristic of mechanisms is that they are complex structures consisting of a number of parts. Thus, a mechanical model must specify the hypothesized parts of the mechanism. The parts of the Gerstman and Syrdal/Gopal mechanisms are given respectively in the boxes found in figures 2 and 3. These diagrams do not simply specify a list of parts, but also their functional arrangement. In particular, the arrows represent the causal interactions that occur between the different parts. The flow chart representation I have used to summarize these mechanisms is certainly not the only possible way to present a mechanical model, but it is an especially natural one.

If one compares the diagrams in figures 2 and 3 with visual representations of more paradigmatic mechanisms, like schematics for electronic or mechanical devices, what is striking is that there is no indication of the size, location or arrangement of the parts. Parts are instead specified functionally, that is, in terms of the causal role of the part within the overall mechanism. Note that the parts are themselves complex mechanisms. Their behavior is described but the mechanism underlying the behavior left unspecified. This strategy has an important advantage. By leaving out the details by which these functions are implemented, the model highlights the explanatorily relevant features of the mechanism. The correctness of these explanations will not turn on how it is, for instance, that the formant frequency analyzer works, so long as there exists some mechanism that performs this function.

III.   Evaluating Mechanical Models

The Gerstman and Syrdal/Gopal models represent competing models of the mechanism responsible vowel normalization. It is tempting to ask which (if either) of these models has gotten the mechanism right, but that question is naïve. The reason for this is not just that decisive evidence is hard to come by, but rather that the posited relationship between a model and the mechanism it models is one of similarity rather than isomorphic correspondence. This similarity comes in varying degrees and respects; and while some authors (e.g., Hughes, 1997) argue that the concept of similarity is a vague and misleading, in the context of particular models, similarity claims can be spelled out in ways that make it unproblematic.

What distinguishes mechanical models from models generally is that they must articulate a set of components whose activities and interactions produce the phenomenon in question. For models of this sort, there are a number of questions one can ask about respects of similarity. We may divide these respects into two classes. The first class concerns of the adequacy of the behavioral description, or simply behavioral adequacy:

1. Does the model predict (quantitatively or qualitatively) the overall behavior of the mechanism? Do these predictions hold for all inputs, or only for some ranges?

The second class concerns adequacy of the mechanical description, or mechanical adequacy:

2. Has the model identified all of the components in the mechanism? Have the components been localized?
3. For each component, has the model correctly identified its causally relevant properties – i.e., the properties whose changes figure into interactions with other components?
4. Does the model provide quantitatively accurate descriptions of the interactions and activities of each component?
5. Does the model correctly represent the spatial and temporal organization of the mechanism?
6. If the model includes submodels of the mechanical structure of components, are these submodels good representations of these components?
7. Is the mechanism identified by the model the sole mechanism responsible for the production of the behavior, or are there multiple mechanisms? If there are multiple mechanisms, do they operate concurrently and redundantly, or do different mechanisms operate in different contexts?

The distinction between behavioral and mechanical adequacy has been recognized by other authors. Lloyd makes a similar distinction, where she observes that a model can adequately predict outcomes, but be merely understood as "calculating devices… because their isomorphism with the natural system is so limited" (1994, p. 148). Similarly, as I have argued elsewhere (2002b) much of the debate over the adequacy of genic selection

models and "beanbag genetics" can be seen as a debate about whether behaviorally adequate genic selection models correctly represent causal mechanisms.

Let us now consider the adequacy of the behavioral descriptions given by the Gerstman and Syrdal/Gopal models. These models were constructed after the fact, as sets of algorithms for correctly classifying those vowels in the Peterson and Barney dataset that were unanimously classified by a panel human listeners. As such, both models are "observationally equivalent" to this panel of human listeners, and so it is plausible to say that the models meet the criterion of behavioral adequacy. On the other hand, it is important to remember that this equivalence has only been tested for a very small range of possible inputs. A more adequate model of the mechanism would approximate the behavior of human listeners for vowels spoken by a wider variety of speakers in a wider variety of phonetic, prosodic, syntactic and semantic contexts. The difficulty of such a demand is that to emulate human speakers in such contexts would require the construction of models of large parts of the complex mechanisms of speech perception. Human listeners use higher-level information to help in speech perception tasks (See, e.g., Warren, 1970 on the phoneme restoration effect). Unless these models of vowel recognition are embedded in larger models of speech recognition, it will not be possible to fully judge the behavioral adequacy of these models.

But let us set aside the limitations on the behavioral adequacy of these models, and turn to their mechanical adequacy. To the extent that both of these models "save the phenomena" (i.e., are behaviorally adequate), they illustrate how the relationship between behavioral and mechanical adequacy parallels the well-known problem of underdetermination.

The classic formulation of the underdetermination problem presupposes a distinction between observation and theory statements. The problem then is that any of an infinite number of theories (i.e., sets of theoretical statements together with bridge principles) may entail a given set of observation statements. Critics of the underdetermination thesis (e.g., Laudan, 1990) argue that while there is an "in principle" logical sense in which observation will always underdetermine theory, the practical import of the result is overblown.

While I agree with such assessments of the logical problem, a more interesting version of the problem surfaces in assessing mechanical models. In this case, the problem is that multiple mechanical descriptions (analogous to theories) may entail similar or identical behavioral descriptions (analogous to observations). In the case of the two models discussed here, the mechanical descriptions are not ad hoc constructions, but represent plausible competing hypotheses about the mechanisms responsible for observable behaviors.

In its mechanistic version, the underdetermination problem is not a logical conundrum but a methodological challenge. If two models have equivalent behavioral

descriptions, how does one decide between them? There are two sorts of approaches. The first approach, which I call <u>direct inference</u>, is to try to take the mechanism apart and study the behavior of its parts.

The functional localization strategies explored by Bechtel and Richardson (1993) and the strategies discussed by Craver and Darden (2001) for discovering neurobiological mechanisms are principally strategies of this kind.   Direct inference strategies defeat the underdetermination problem by breaking the boundary between theory and observation. While the internal workings of mechanisms may not be readily apparent, they may be accessible via special observational and experimental methods.

Direct inference strategies are important, and indeed a good deal of work on mechanisms of speech perception relies upon them.   However, there are sometimes circumstances in which it is impractical or impossible to dissect a mechanism. In the case of high level cognitive mechanisms, the parts themselves may be complex and highly distributed and may defy our strategies for localization. Moreover, in the particular case of speech research where we have no model organisms other than human beings, ethical considerations place considerable constraints on the kind of physical dissection of mechanisms that we might otherwise undertake. Thus in cases such as this one must utilize testing techniques that allow one to make inferences about the internal structure of a mechanism simply by examining the mechanism's external behavior. These techniques, which I call <u>indirect inference</u> methods, examine the behavior of the mechanism under non-standard conditions.   They defeat the standard underdetermination problem by expanding the range of behavioral information (i.e., by increasing the range of phenomena the models must "save"). Placing a mechanism in non-standard conditions will will often cause it to behave in unusual ways—that is, it may break the mechanism. Observing when and how a mechanism breaks can tell one a lot about how a mechanism works.   One tests the models by seeing if they break in ways similar to the actual mechanism.

I shall now discuss some of the experimental evidence that can be used to evaluate these models, focusing on three contrasting features of the models and showing how techniques of direct or indirect inference can be used to evaluate them.[12] The first of these features concerns the temporal organization of the mechanism, and in particular the time at which information required to normalize the vowel space is acquired. Recall that the basic purpose of these mechanisms is to normalize the vowel space in order to compensate for variations among talkers. In the initial description of the Gerstman model we observed that the model uses a "contextual tuning" mechanism, whereby the listener acquires information about the highest and lowest F1 and F2 values for a particular

---

[12] I discuss a broader array of features in more technical detail in Glennan 1992.

speaker's vowels, and uses this information to adjust to that speaker. How this calibration mechanism works is left unspecified, but presumably it works by using higher level information (e.g., involving expectations about what people say) to make guesses as to what vowels are being used. Thus, in order for a signal representing a vowel token to be correctly identified, the mechanism must already have been calibrated by an earlier signal. In contrast, the Syrdal and Gopal model supposes that the signal representing the vowel token is self-normalizing, in the sense that adjustments to inter-speaker variations in F1 and F2 can be made using other features of the signal itself.

Nusbaum and Morin (1992) devised an experiment to test whether human vowel recognition depends upon a contextual turning or a non-contextual self-normalizing mechanism. This experiment is an exemplar of an indirect inference technique. Under normal conditions, listeners have access to both the contextual information that would tune a Gerstman-type mechanism and the structural features of the speech signal that would tune a Syrdal/Gopal-type mechanism. Their idea was to construct a non-standard condition in which contextual information was removed, and to see if this would "break" the mechanism. If this occurred, it would be evidence against the contextual tuning model.

To do this Nusbaum and Morin provided listeners with sequences of syllables and asked them to identify which of these syllables contained a target vowel. Subjects were given two kinds of sequences. In the first type of sequence, all syllables came from the same talker, while in the second, syllables came from different talkers. The second condition then eliminated contextual cues that could be used to calibrate the mechanism.

Nusbaum and Morin's experiments had mixed results. On the one hand, they found no significant difference in listener's performance in the two conditions. This fact suggests that listeners do not need the context to tune their vowel normalization mechanism. On the other hand, they found that listeners were considerably faster in making identifications in the same-talker condition. These results lead Nusbaum and Morin to hypothesize that human listeners have redundant normalization mechanisms, with the more computationally efficient contextual tuning mechanism being used unless contextual clues were eliminated.

The second feature with respect to which we can compare these models concerns the features of the speech signal that are used to identify vowels. The Gerstman model uses only the first and second formants, while the Syrdal and Gopal model uses additional information – notably the fundamental frequency (F0) and third formant (F3). Nusbaum and Morin (1992) again devised an experiment that used indirect inference to determine how the models compared to actual auditory mechanisms in this respect. While the standard data set that was used to assess the adequacy of these models' behavioral descriptions contained the full range of acoustical information associated with speech signals, Nusbaum and Morin constructed modified signals in which F0 and F3 were

removed. Listeners were presented with the modified set of stimuli and were again asked to identify which syllables were instances of target vowels. Once again, they were presented with sequences of syllables in either a same-talker or mixed-talker condition. The results of these experiments were consistent with Nusbaum and Morin's hypothesis of redundant mechanisms. In same-talker condition, where contextual cues were available, the absence of F0 and F3 information did not impair speakers' ability to recognize target vowels. In the mixed-talker condition, where both the contextual cues and the F0 and F3 information were absent, there was significant degradation in the ability of listeners to identify target vowels. In this condition, neither of the redundant mechanisms would have the parameters required to operate correctly.

The third and final feature which I shall discuss concerns the hypothesized mechanism of frequency analysis. Both models suppose there is an initial component of the mechanism that analyzes the acoustic signal into pitch and formant frequencies, but they differ in the output of these components. The Gerstman model proposes that these frequencies reflect the actual acoustic properties of the signal while the Syrdal/Gopal model supposes that perceived frequencies are not linear with respect to acoustic frequencies. Accordingly, the two models describe the output of this component using different scales – a kilohertz scale for Gerstman and a perceptual scale called the Bark scale for Syrdal and Gopal.

At issue here is not the accuracy of vowel classification made by the two models, but the accuracy of the description of the behavior of one of the mechanism's parts. This part of the mechanism can be studied more directly than other hypothesized components. First, it is possible to study pitch perception apart from phoneme recognition by performing experiments in which subjects are asked to discriminate pure pitches. Second, there is physiological evidence that the task of frequency analysis is carried out within the ear. Current evidence suggests that (1) all sound intensity within a certain critical band of acoustic frequencies is integrated and perceived as a single pitch; (2) the width of this critical band is not constant over the perceptual range of the auditory mechanism. (Scharf, 1970); and (3) critical band corresponds to a fixed length along the basilar membrane (Syrdal and Gopal, 1986).

There is thus substantial evidence that any vowel normalization mechanism would operate on perceived rather than acoustic frequencies of pitch and formants. While in this respect the Syrdal/Gopal model is superior, it is interesting that the fact that the Gerstman model uses acoustic frequencies does not impair the model's accuracy. Indeed, it is important to recall in this connection that, because his original purpose was to develop a technique for machine recognition, Gerstman did not need to utilize the same frequency analysis mechanism that is used in the human auditory system.

This is only a partial review of the evidence in favor of the two models, but it is enough to draw two conclusions about the testing of models generally. First, it shows that methods of indirect inference make it is possible to study the properties of mechanisms whose parts are, for whatever reason, difficult to examine and manipulate directly. Second, it shows that there is not typically a Baconian crucial experiment to decide between rival models. Models can't be falsified because models are not true or false, but are rather similar to the systems they model in degrees and respects of the sort listed at the beginning of this section. This fact also explains why it is not possible to cleanly separate discovery and testing, (cf. Darden, 1991, Craver and Darden, 2001). Models aren't generally thrown out after failures in crucial experiments, but are rather elaborated to gradually increase their degrees and respects of similarity.

It is difficult to specify a principled way in which the degrees and respects of similarity and dissimilarity should be weighed against each other, and hence it may be difficult to say which of two models is the more similar to the modeled system. But this is hardly a defect in the analysis of models presented here. Indeed, it goes some way towards explaining why scientists may continue to test and articulate a number of different and incompatible models.

It might be argued that my choice of models has lead to an overstatement of my case. The models discussed in this paper are representations of a small part of an extremely complicated speech recognition mechanism, and there are reasons to believe that a fuller understanding of speech recognition will show that the relationship between vowel recognition and other parts of speech processing are more complicated than current models suggest. We have also seen evidence that the mechanism responsible for this aspect of speech recognition appears to be redundant, so that each model tells only part of the story. Moreover, except in the case of the mechanism of frequency analysis, there is very little direct evidence about where or how the hypothesized components operate. Collectively, these facts might suggest that the models in question are unusually partial and sketchy, an investigation of more developed models would not support my conclusions regarding the nature of models as representations.

There will certainly be models in which the range of behavior is more completely specified, the parts and their properties more clearly identified, and the like. In such cases, it may be reasonable to judge a model of a mechanism correct or incorrect. If, for instance, one has a circuit diagram for a radio, it would seem possible to judge it as either correct or incorrect. But even a correct circuit diagram will not be similar to the radio in all respects. It will not, for instance, represent the physical size of components or the physical geometry of their layout on the circuit board. Models, as Giere (1999) tells us, are like maps, and maps always represent only selected features of the terrain. Equally important, my hunch (and it is admittedly only a hunch) is that the sorts of models one constructs to describe complex natural phenomena will more often be of the partial kind I

have described than of the sort represented by a circuit diagram. At any rate, we should recognize that there is no firm line between complete and partial models (or between mechanism schemata and mechanism sketches), but that models undergo continuous processes of articulation and refinement.

Not all models studied by scientists are mechanical models, but many are. While there will always be domain specific strategies of model formulation and testing, recognizing the pervasiveness of mechanisms and their models will increase understanding of similarities in strategies across domains.

Acknowledgments

References

Bailer-Jones, D. (2003). When scientific models represent. International Studies in the Philosophy of Science, 17, 59-74.

Beatty, John. (1981). What's wrong with the received view of biological theory. In P. Asquith and R. Giere (Eds.), PSA 1980: Proceedings of the Philosophy of Science Association, Vol. 2 (pp. 397–426). East Lansing, MI: Philosophy of Science Association.

Bechtel, W. & Richardson, R. (1993). Discovering complexity: cecomposition and localization as strategies in scientific sesearch. Princeton: Princeton University Press.

Craver, C. (2002). Interlevel experiments and multilevel mechanisms in the neuroscience of memory. Philosophy of Science, 69 (Proceedings), S83-97.

Craver, C. & Darden, L. (2001). Discovering mechanisms in neurobiology: the case of spatial memory. In P. Machamer, R. Grush, & P. McLaughlin (Eds.), Theory and method in neuroscience (pp. 112-137). Pittsburgh: University of Pittsburgh Press.

Darden, L. (1991). Theory change in science: strategies from Mendelian genetics. New York: Oxford University Press.

Gerstman, L. J. (1968). Classification of self-normalized vowels. <u>IEEE Transactions on Audio Electroacoustics, AU-16</u>, 78-80.

Giere, R. (1988). <u>Explaining science.</u> Chicago: University of Chicago Press.

Giere, R. (1999). Using models to represent reality. In L. Magnani, N. Nersessian, & P. Thagard, (Eds.), <u>Model-based reasoning in scientific discovery</u> (pp. 41-57). Dordrecht: Kluwer.

Glennan, S. (1996). Mechanisms and the nature of causation. <u>Erkenntnis, 44</u>, 49-71.

Glennan, S. (1997). Capacities, universality and singularity. <u>Philosophy of Science, 64</u>, 605-626.

Glennan, S. (2002). Rethinking mechanistic explanation. <u>Philosophy of Science, 69 (Proceedings),</u> S342-53

Glennan, S. (2002b). Contextual unanimity and the units of selection problem. <u>Philosophy of Science, 69</u>, 118-137.

Hughes, R. I. G. (1997). Models and representation. <u>Philosophy of Science, 64 (Proceedings),</u> S323-S336.

Laudan, L. (1990). Demystifying underdetermination. In C. W. Savage (Ed.), <u>Scientific theories, Minnesota studies in the philosophy of science, Vol. 14</u> (pp. 267-297). Minneapolis: University of Minnesota Press.

Lewontin, R. C. (1974). <u>The genetic basis of evolutionary change</u>. New York: Columbia University Press.

Levins, R. (1968). The strategy of model building in population biology. <u>American Scientist, 54,</u> 421-31.

Lloyd, E. (1994). <u>The structure and confirmation of evolutionary theory</u>. Princeton: Princeton University Press.

Machamer, P., Darden L & Craver, C. (2000). Thinking about mechanisms. <u>Philosophy of Science, 67</u>, 1-25.

Marr, D. (1982). <u>Vision: a computational investigation into the human representation and processing of visual information</u>. San Francisco: W. H. Freedman.

Mitchell, S. (1997). Pragmatic laws. <u>Philosophy of Science, 64 (Proceedings),</u> S468-479.

Moore, B. (2003). <u>An introduction to the psychology of hearing, 5th Edition</u>. New York: Academic Press.

Morgan, M. & Morrison, M. (Eds.) (1999). <u>Models as mediators.</u> Cambridge: Cambridge University Press.

Nusbaum, H. & Morin, T. (1992). Paying attention to differences among talkers. In Y. Tohkura, Y. Sagisaka, & E. Vatikiotis-Bateson (Eds.), <u>Speech perception, production, and linguistic structure</u> (pp. 113-134). Tokyo: OHM Publishing Company.

Odenbaugh, J. (in preparation). True lies: robustness and idealizations in ecological explanations.

Peterson, G. E., & Barney, H. L. (1952). Control Methods used in a study of vowels. Journal of the Acoustical Society of America, 24, 175-84.

Scharf, B. (1970). Critical bands. In J. V. Tobias (Ed.) Foundations of modern auditory theory, Vol. 1. New York: Academic Press.

Suppe, F, (Ed.). (1974). The structure of scientific theories. Chicago: University of Illinois Press.

Suppe, F. (1989). The semantic conception of theories and scientific realism. Chicago: University of Illinois Press.

Suppes, P. (1960). A comparison of the meaning and uses of models in mathematics and the empirical sciences. Synthese, 12, 287-300.

Suppes, P. (1962). Models of data, In E. Nagel, P. Suppes and A. Tarski (Eds.), Logic, methodology and the philosophy of science: proceedings of the 1960 international congress (pp. 252-61). Stanford, CA: Stanford University Press.

Syrdal, A.K., & Gopal, H.S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. Journal of the Acoustical Society of America 79, 1086-1100.

Warren, R. M. (1970). Perceptual restorations of missing speech sounds. Science, 167, 392-393.

Wimsatt, W. (1972). Complexity and organization. In K. F. Schaffner & R. S. Cohen (Eds.), PSA 1972, Proceedings of the Philosophy of Science Association (pp. 67–86). Dordrecht: Reidel.

Wimsatt, W. (1987). False models as a way to truer theories. In M. Nitecki & A. Hoffman (Eds.), Neutral models in biology (pp. 23-55). Oxford: Oxford University Press.

Woodward, J. (2000). Explanation and invariance in the special sciences. British Journal for the Philosophy of Science, 51, 197-254.