From Wimsatt, W. C. (in press) *Re-engineering philosophy for limited beings: Piecewise approximations to reality*. Cambridge, MA: Harvard University Press. Originally published in M. Nitecki and A. Hoffman, eds., *Neutral Models in Biology*. London : Oxford University Press, pp. 23-55

Chapter 6
False Models as Means to Truer Theories

Introduction

Many philosophers of science today argue that scientific realism is false. They often mean different things by this claim, but most would agree in arguing against something like the view that scientific theories give, or at least aim to give, approximate, or are approaching asymptotically to giving, a true description of the world. All theories, even the best, make idealizations or other false assumptions that fail as correct descriptions of the world. The opponents of scientific realism argue that the success or failure of these theories must therefore be independent of or at least not solely a product of how well they describe the world. If theories have this problematic status, models must be even worse, for models are usually assumed to be mere heuristic tools to be used in making predictions or as an aid in the search for explanations, and which only occasionally are promoted to the status of theories when they are found not to be as false as we had assumed.

While this rough caricature of philosophical opinion may have some truth behind it, these or similar views have lead most writers to ignore the role that false models can have in improving our descriptions and explanations of the world. (Nancy Cartwright's (1983) excellent studies of the use and functions of models, though aiming at other conclusions, are a striking exception here). While normally treated as a handicap, the falsity of scientific models is in fact often essential to this role. I will not discuss the larger issue of scientific realism here: the way in which most philosophers have formulated (or mis-formulated) that problem renders it largely irrelevant to the concerns of most scientists who would call themselves realists. (Philosophers attack a realism that is “global” and metaphysical. Most scientists use and would defend a more modest (or “local”) realism, and would do so on heuristic rather than on metaphysical grounds.)

By local realism, I mean something like the following: on certain grounds (usually, for example that the existence of an entity or property is known, derivable or detectable through a variety of independent means--see Wimsatt, 1980), scientists would argue that an entity or property is real, and they cannot imagine plausible or possible theoretical changes which could undercut this conclusion. Furthermore, they might argue that their experimental and problem-solving approaches require them to presuppose the existence of that entity, property or phenomenon--a heuristic argument. I suspect that many philosophical opponents of scientific realism could accept this kind of local and heuristic realism. I think that it is the kind of realism most worth defending, though I think that it may also give handles for defending a more ambitious kind of scientific realism. In any case, there is much more of use to be found in the topic of false models, to which I will now turn.

Neutral models will be discussed mostly indirectly in this paper. After an attempt at characterizing what a neutral model is, I will show that this idea relates more generally to the use of false models, and then discuss a variety of ways in which false models are used to get to what we, at least provisionally, regard as the truth. Many or most of the papers in this volume turn out to use their models in one or more of these ways. Thus, the focus of my paper may be conceived as generalizing beyond the case of neutral models to consider the variety of uses of models, in spite of, or even because of the fact that they are assumed to be false, to get at the truth or our best current approximations to it.

Even the best models have "biases":

The term "neutral model" is a misnomer if it is taken to suggest that a model is free of biases such as might be induced by acceptance of a given hypothesis (such as that the patterns to be found among organisms are products of selection). Any model must make some assumptions and simplifications, many of which are problematic, so the best working hypothesis would be that there are no bias free models in science.

This observation has a parallel in the question, "What variables must be controlled for in an experimental design?" There are no general specifications for what variables should be controlled, since (1) what variables which should be controlled or factored out through appropriate attempts to isolate the system, (2) what variables should be measured, (3) what errors are acceptable, and (4) how the experiment should be designed are all functions of the purpose of the experiment. Similarly, (1) what models are acceptable, (2) what data are relevant to them, and (3) what counts as a "sufficiently close fit" between model and data is a function of the purposes for which the models and data are employed. (As one referee for this paper pointed out, many people who would applaud the use of appropriate controls and isolations in an experimental design inconsistently turn on mathematical or causal models of the system and criticize them for doing the same thing! This activity is desirable and necessary in either case.)

Any model implicitly or explicitly makes simplifications, ignores variables, and simplifies or ignores interactions among the variables in the models and among possibly relevant variables not included in the model. These omitted and simplified variables and interactions are sources of bias in cases where they are important. Sometimes certain kinds of variables are systematically ignored. Thus, in reductionist modeling, where one seeks to understand the behavior of a system in terms of the interactions of its parts, a variety of model-building strategies and heuristics lead us to ignore features of the environment of the system being studied (Wimsatt, 1980b, 231-235). These environmental variables may be left out of the model completely, or if included, treated as constant in space or in time, or treated as varying in some particularly simple way, such as in a linear or random fashion. In testing the models, with the focus on the interrelations among internal factors, environmental variables may be simply ignored, (Wimsatt, 1980b) or treated in some aggregate way to simplify their analysis (Taylor, 1985, Wimsatt, 1985a).

A model may be bias free in one case (where its variables and parameters refer, at least approximately correctly to causal factors in nature, and where its "accessory conditions" (Taylor, 1985) are satisfied) but biased for other cases it is applied to because it worked well in the first case (Taylor, 1985, discusses examples of this for ecological models in greater detail). Even where it is recognized that a model must be changed, there may be biases (1) in how or where it is changed, or (2) in the criteria accepted for recognizing that it must be changed.

An example of the first type is found in reductionist modeling: whereas both the description and interaction of parts internal to and of variables external to the system are usually oversimplified, there will often be a bias towards increasing the internal realism of the model in cases where the failure of fit of the model with the data is due to unrealistic assumptions about the environment (Wimsatt, 1980b).

A potential example of the second kind of case is provided by Williams' (1966, p. 17) "principle of parsimony," in which he recommends "recognizing adaptation at no higher level than is required" (presumably by the data). If this recommendation is taken as an invitation to find *some* set of parameter values for the simple model for which it fits the

data, then one may be engaged in a "curve-fitting" exercise which may hide the need for a higher level or more complex model, unless one takes the pains to determine whether the parameter values for which fit is achieved are actually found in nature. This is seldom done, at least in the kind of optimization modeling frequently found in evolutionary biology.

The concept of a "neutral model":

In evolutionary biology and ecology, a "neutral model" usually means a model without selection. Thus, Raup (this volume—Nitecki and Hoffman 1987) in his work with Gould, Schopf and Simberloff considers "random phylogenies"--phylogenetic descent trees in which originations and extinctions are determined by random variables. These artificial phylogenies in many respects resemble those found in nature. They thus reasoned that the similarities between the artificial and natural phylogenies were not products of selection processes operating at that level. Their model did not, as they pointed out, rule out selectionist explanations for speciations and extinctions at a lower (e.g., intra- or inter-populational) level. Similarly, the work of Kimura, Crow (this volume), and others on "neutral mutation" theories modeled and evaluated patterns of molecular variability and change on the assumption that selection forces on these variants could be ignored. Similarities between their predicted patterns and what was found in nature led to various versions of the hypothesis that the evolution of various systems or of various kinds of traits was driven not by selection, but by various forms of "genetic drift." Finally, Kauffman's work on "generic constraints" in development (Kauffman, 1985, and in this volume) identified features which, because they were near universal properties of his randomly constructed model genetic control networks, are taken to provide "baselines" for the properties of systems on which selection acts. He argues that these "generic properties" will usually survive *in spite of* selection rather than because of it. Here Kauffman is using a comparison of models with and without selection to argue that selection is not important.

One must not assume that if the data fit the "neutral" model then the excluded variables are unimportant. The researches of Raup, Crow, and Kauffman each suggest that selection may be unimportant in some way in producing the phenomena being modeled, but they do not rule it out entirely. Thus, Raup's "random phylogenies" do not exclude selection as a causal agent in producing individual extinctions (e.g., through overspecialization to a temporally unstable niche) or speciations (e.g., through directional selection in different directions producing isolating mechanisms between two different geographically isolated subpopulations of the same species), because his model simply does not address those questions.

Similarly, the work of Crow and Kimura could be consistent with the (very plausible) hypothesis that the very neutrality of most mutations is a product of selection. This could occur through selection for developmental canalization of macroscopic morphological traits, selection for redundancy of design through duplicated DNA, codon synonymy, or redundancy at the level of parallel synthetic paths in metabolic pathways, or at other higher levels. It could also represent selection for an architecture for the program of gene expression to preserve relatively high heritability of fitness or other phenotypic traits in the face of random sexual recombination, and uses any of the preceding mechanisms or others yet to be discovered to accomplish this.

Finally, the ubiquity of Kauffman's "generic properties" among genetic control networks of a certain type (which he takes as a constraint on his simulations) does not rule out the possibility that that type may itself be a result of prior selection processes. Thus, Kauffman's genetic control networks have an average of two inputs and outputs per control element. But he chose this value of these parameters in part because his own prior work (Kauffman, 1969) showed that such networks had a shorter mean cycle time than networks containing more or fewer inputs and outputs, a feature that he argues is advantageous.

I have elaborated elsewhere a model of developmental processes (Wimsatt, 1986), which presents an alternative hypothesis for the explanation of developmental constraints that seems likely in many such cases, and in even more cases where the feature in question is taxonomically very widely distributed, but not absolutely universal. Such broad or "generic" universality of some traits could be produced by strong stabilizing selection due to the dependence of a wide variety of other phenotypic features on them. This view is not new, but to my knowledge, no one has attempted to model its consequences, with the exception of a model proposed by Arthur (1982, 1984) to explain features of macroevolution. (Either Arthur's model or mine gives a plausible explanation, for example, for the near-universality of the genetic code. See, e.g., Kauffman, 1985.)

Jeffrey Schank and I have recently attempted to test my model by doing simulations on networks like those of Kauffman, (1985), where connections do not contribute equally to fitness (as in Kauffman's selection simulations) but the contribution to fitness of a connection is a function of the number of nodes that are affected by that connection. (We have tried 3 different fitness functions as measures of the topological properties of the connection and of the number of nodes that are "downstream" of it, with similar results in all cases.) The results are quite striking. Our networks, like Kauffman's, show decay in a number of "good" connections under mutation and selection. (This effect, which increases in strength with the size of the network, is what leads him to doubt that selection can maintain a network structure of any substantial size against mutational degradation.) But if the number of nodes accessible to a given gene through its various connections is taken as a measure of "generative entrenchment" (Wimsatt 1985), then our results show that the genes (or connections of genes) which have been lost in this mutational decay are those with low degrees of generative entrenchment, and that virtually all of the genes (and connections) which are significantly generatively entrenched are still there after 1000 to 5000 generations!

Thus, in effect, Kauffman's models confirm my theory, and the two theories need to be regarded as complementary rather than contradictory. We are doing a number of simulations to investigate the robustness of this phenomenon for different mutation rates, population sizes, numbers of generations, genome sizes, and connection densities. In the not too far distant future, it should be possible to do it for more realistic biochemical networks. We expect soon to write up some of these results for publication.

The models of Raup, Crow, and Kauffman use and rule out selection factors in different circumstances and in different ways, but they have two things in common. In each, the "neutral" model either does not include selection operating on the postulated variants, or (in Kauffman's case) supports arguments that selection is not efficacious under the circumstances considered. And in each, the "neutral" model is treated as specifying a "baseline" pattern with which natural phenomena and data are to be compared, in order to determine whether selection is required (if the natural phenomena do not fit the "neutral" pattern) or not (if they do fit.) These "neutral models" are treated as "null hypotheses" (a

term frequently found in the literature on ecological models, and of course in statistics--see Stigler's contribution to this volume) which are to be rejected only if the natural phenomena deviate sufficiently from those predicted by the model.

These cases suggest that we characterize a *"neutral model" as a "baseline" model that makes assumptions, which are often assumed to be false for the explicit purpose of evaluating the efficacy of variables that are not included in the model.* This leaves out selection, and thus perhaps falls short of a more complete characterization of "neutral models" in evolutionary biology, but this more general characterization makes more explicit the connection with hypothesis testing in statistics and allows us to bring in features of the use of models from other areas of biology to focus on the heuristic advantages of their use. The characterization of "neutral models" in this way leads naturally to the more general question of *when, how, and under what conditions models which are known or believed to be false can be used to get new or better information about the processes being modeled.* If "neutral" models are useful in biology, this has less to do with their "neutrality" than with more general features of the use of models in science.

The fit of data with a "neutral" model or "null hypothesis" usually establishes that omitted variables do not act in a way specific to the models under comparison, not that they do not act at all. This is consonant with my earlier claim that the adequacy of models is highly context-dependent, and that their adequacy for some purposes does not guarantee their adequacy in general. However, the use of these models as "templates" can focus attention specifically on where the models deviate from reality, leading to estimations of the magnitudes of variables left out, or to the hypothesis of more detailed mechanisms of how and under what conditions these variables act and are important. This is a pattern of inference that is both common and important, and deserves closer scrutiny. The variety of ways in which this is done will be the primary focus of the remainder of the paper.

How models can misrepresent:

The simple observation that most models are oversimplified, approximate, incomplete, and in other ways false gives little reason for using them. Their widespread use suggests that there must be other reasons. It is not enough to say (e.g., Simon, 1981) that we cannot deal with the complexities of the real world, so simple models are all that we can work with, for unless they could help us do something in the task of investigating natural phenomena, there would be no reason for choosing model building over astrology or mystic revelation as a source of knowledge of the natural world.

Nor does the instrumentalist suggestion that we use them because they are effective tools rather than realistic descriptions of nature give us much help, for it presupposes what we want to understand—namely, *how* false models can be effective tools in making predictions and generating explanations. I want to suggest various ways in which false models can (1) lead to the detection and estimation of other relevant variables, (2) help to answer questions about more realistic models, (3) lead us to consider other models as ways of asking new questions about the models we already have, and (4) (in evolutionary or other historical contexts) determine the efficacy of forces which may not be present in the system under investigation but which may have had a role in producing the form that it has.

Before discussing ways in which false models can help us to find better ones, it is useful to have a classification of the ways in which a model can be false. By a model in this context, I mean one of two alternative things: (1) a mathematical model--an equation

or set of equations together with the interpretations necessary to apply them in a given context, or (2) a causal model or proposed mechanism through which a phenomenon or set of phenomena is to be explained. Often one will have both, but the following comments apply roughly equally to either.

The following are ways in which a model can be false. They are ordered roughly in terms of increasing seriousness (except for 6 and 7.):

(1) A model may be of only very *local applicability*. This is a way of being false only if it is more broadly applied.
(2) A model may be an *idealization* whose conditions of applicability are never found in nature, (e.g., point masses, the uses of continuous variables for population sizes, etc.), but which has a range of cases to which it may be more or less accurately applied as an approximation.
(3) A model may be *incomplete*--leaving out 1 or more causally relevant variables. (Here it is assumed that the included variables are causally relevant, and are so in at least roughly the manner described.)
(4) The incompleteness of the model may lead to a *misdescription of the interactions* of the variables which are included, producing apparent interactions where there are none ("spurious" correlations), or apparent independence where there are interactions--as in the spurious "context independence" produced by biases in reductionist research strategies. Taylor, (1985) analyzes the first kind of case for mathematical models in ecology, but most of his conclusions are generalizable to other contexts. (In these cases, it is assumed that the variables identified in the models are at least approximately correctly described.)
(5) A model may give a *totally wrong-headed* picture of nature. Not only are the interactions wrong, but also a significant number of the entities and/or their properties do not exist.
(6) A closely related case is that in which a model is purely "*phenomenological*." That is, it is derived solely to give descriptions and/or predictions of phenomena without making any claims as to whether the variables in the model exist. Examples of this include: the virial equation of state (a Taylor series expansion of the ideal gas law in terms of T or V.); the automata theory (Turing machines) as a description of neural processing; and linear models as curve fitting predictors for extrapolating trends.
(7) A model may simply *fail to describe or predict the data* correctly. This involves just the basic recognition that it is false, and is consistent with any of the preceding states of affairs. But sometimes this may be all that is known.

Most of the cases to be discussed in this paper represent errors of types 2 thru 5, and the productive uses of false models would seem to be limited to cases of types 1 thru 4 and 6. It would seem that the only context in which case 5 could be useful is where case 6 also applies, and often models that are regarded as seriously incorrect are kept as heuristic curve fitting devices. There is no hard and fast distinction between phenomenological and non-phenomenological models, and the distinction between the two often appears to depend on context (see Cartwright, 1983). Thus, the Haldane mapping function can be derived rigorously from first principles, but as such it makes unrealistic assumptions about the mechanisms of recombination. It is sometimes treated as an idealization (case 2) and sometimes as a phenomenological predictive equation (case 6). Finally, a model may make false or unrealistic assumptions about lower level mechanisms, but still produce

good results in predicting phenomena at the upper level. In this case, the upper level relationships may either be robust (see Levins, 1966, Wimsatt, 1980a, chapter 4) or may be quite fragile results which are secured through obvious or unintended "curve fitting" as with many optimization models in evolutionary biology and economics. (The former is a productive use of falsehood--item 10 below--and the latter is an often-unproductive instance of case 6 above.)

Twelve Things to do with False Models:

It may seem paradoxical to claim that the falseness of a model may be essential to its role in producing better models. Isn't it always better to have a true model than a false one? Naturally it is, but this is never a choice that we are given, and it is a choice that only philosophers could delight in imagining. Will any false model provide a road to the truth? Here the answer is just as obviously an emphatic "no!" Some models are so wrong, or their flaws so difficult to analyze that we are better off looking elsewhere. Cases 5 and 7 above represent models with little useful purchase, and "curve fitting" phenomenological models (case 6) would be relatively rarely useful for the kinds of error correcting activity I propose. The most productive kinds of falsity for a model are cases 2 or 3 above, though cases of types 1 and 4 should sometimes produce useful insights. *The primary virtue a model must have if we are to learn from its failures is that it, and the experimental and heuristic tools we have available for analyzing it, are structured in such a way that we can localize its errors and attribute them to some parts, aspects, assumptions, or subcomponents of the model.* If we can do this, then "piecemeal engineering" can improve the model by modifying its offending parts.

There is a mythology among philosophers of science (the so-called "Quine-Duhem thesis") that this cannot be done, that a theory or model meets its experimental tests wholesale, and must be taken or rejected as a whole. Not only science, but also technology and evolution would be impossible if this were true in this and in logically similar cases. That this thesis is false is demonstrated daily by scientists in their labs and studies, who modify experimental designs, models, and theories piecemeal, by electrical engineers, who localize faults in integrated circuits, and by auto mechanics and pathologists, who diagnose what is wrong in specific parts of our artifacts and our natural machines, and correct them. Glymour (1980) and Wimsatt (chapter 4) give general analyses of how this is done, and descriptions of the revised view of our scientific methodology that results. The case for evolutionary processes is exactly analogous. Lewontin (1978) argues that without the "quasi-independence" of traits (which allows us to select for a given trait without changing a large number of other traits simultaneously), any mutation would be a (not-so-hopeful) monster, and evolution as we know it--a process of small piecemeal modifications--would be impossible. (Implications of this insight are elaborated in Wimsatt 1981). In all of these cases then, "piecemeal engineering" is both possible and necessary.

The following is a list of functions served by false models in the search for better ones (parenthetic numbers refer to discussions of appropriate examples in the numbered sections. Multiple references to the same section indicate multiple examples):

(1) An oversimplified model may act as a starting point in a series of models of increasing complexity and realism.

(2) A known incorrect but otherwise suggestive model may undercut the too ready acceptance of a preferred hypothesis by suggesting new alternative lines for the explanation of the phenomena.
(3) An incorrect model may suggest new predictive tests or new refinements of an established model, or highlight specific features of it as particularly important.
(4) An incomplete model may be used as a template, which captures larger or otherwise more obvious effects that can then be "factored out" to detect phenomena that would otherwise be masked or be too small to be seen.
(5) A model that is incomplete may be used as a template for estimating the magnitude of parameters that are not included in the model.
(6) An oversimplified model may provide a simpler arena for answering questions about properties of more complex models, which also appear in this simpler case, and answers derived here can sometimes be extended to cover the more complex models.
(7) An incorrect simpler model can be used as a reference standard to evaluate causal claims about the effects of variables left out of it but included in more complete models, or in different competing models to determine how these models fare if these variables are left out.
(8) Two false models may be used to define the extremes of a continuum of cases in which the real case is presumed to lie, but for which the more realistic intermediate models are too complex to analyze or the information available is too incomplete to guide their construction or to determine a choice between them. In defining these extremes, the "limiting" models specify a property of which the real case is supposed to have an intermediate value.
(9) A false model may suggest the form of a phenomenological relationship between the variables (a specific mathematical functional relationship that gives a "best fit" to the data, but is not derived from an underlying mechanical model). This "phenomenological law" gives a way of describing the data, and (through interpolation or extrapolation) making new predictions, but also, because its form is conditioned by an underlying model, may suggest a related mechanical model capable of explaining it.
(10) A family of models of the same phenomenon, each of which makes various false assumptions, has several distinctive uses: (a) One may look for results which are true in all of the models, and therefore presumably independent of different specific assumptions which vary across models. These invariant results (Levins' (1966) "robust theorems") are thus more likely trustworthy or "true". (b) One may similarly determine assumptions that are irrelevant to a given conclusion. (c) Where a result is true in some models and false in others, one may determine which assumptions or conditions a given result depends upon. (See Levins 1966, 1968, and Wimsatt 1980a and chapter 4 for more detailed discussion).
(11) A model that is incorrect by being incomplete may serve as a limiting case to test the adequacy of new more complex models. (If the model is correct under special conditions, even if these are seldom or never found in nature, it may nonetheless be an adequacy condition or desideratum of newer models that they reduce to it when appropriate limits are taken.)
(12) Where optimization or adaptive design arguments are involved, an evaluation of systems or behaviors which are not found in nature, but which are conceivable alternatives to existing systems can provide explanations for the features of those systems that are found.

The core of the paper, which considers the development of the "chromosomal mechanics" of the Morgan school, is intended to illustrate the first point. Points 2 and 3 are not discussed here, but all of the remaining ones are. In the next sections, I will illustrate these functions by reanalyzing a debate between members of the Morgan school and W. E. Castle over the linearity of the arrangement of the genes in or on the chromosomes in the period 1919-1920.

Background of the debate over linkage mapping in genetics:

In three short years between 1910 when Morgan (1910b) isolated and first crossed his "white eye" mutant *Drosophila*, and 1913 when Sturtevant published his paper mapping relative locations of mutants on the x-chromosome using the frequencies of recombination between them, the Morgan school laid out the major mechanisms of the "linear linkage" model of the location of the genes on the chromosomes. During the next decade Morgan and his colleagues elaborated and defended this model. It won many adherents, and by the early 1920's had become the dominant view, despite several remaining unresolved questions.

This model explained many things. Primary among these was the "linkage" between traits, some of which showed a non-random tendency to be inherited together. Two factors or traits linked in this way (as first discovered by Bateson, Saunders, and Punnett in 1906) violated classical Mendelian models of inheritance. They were not always inherited together, as would be expected for pleiotropic factors that produce multiple effects, such as Mendel's flower and seed coat color. Nor did they assort independently, as Mendel's 7 factors seemed to do. Rather, they showed a characteristic frequency of association that fell in between these extremes. This and similar cases of "partial linkage" were first presented as falsifying instances for the Mendelian theory or model of inheritance. *Here the classical Mendelian model provides a pattern or template of expectations against which the phenomenon of "linkage" acquires significance, as in items 4 or 5 of the above list. Since cases of "partial linkage" represent intermediates between the two classical Mendelian types of total linkage and independent assortment, their classification in this way sets up the problem in a manner that also represents an instance of item 8*. When this phenomenon is first noticed in 1906 there are no theories or models to explain it, though the Boveri-Sutton hypothesis provides a good starting point, one which Morgan and Sturtevant later developed.

With 2 alternative alleles at each of 2 loci, denoted by (**A, a**) for the first pair and (**B, b**) for the second, and starting with genotypes **AABB** and **aabb** as parents, the cases of total linkage or pleiotropy, independent assortment, and partial linkage would show the following proportions among gametes going to make up F2 offspring:

(Parents: **AABB, aabb**:)

Type of Linkage:	Gametic Types:			
	AB	**Ab**	**aB**	**ab**
Total:	50%	0%	0%	50%
Independent:	25%	25%	25%	25%
Partial:	50-**r**%	**r**%	**r**%	50-**r**%

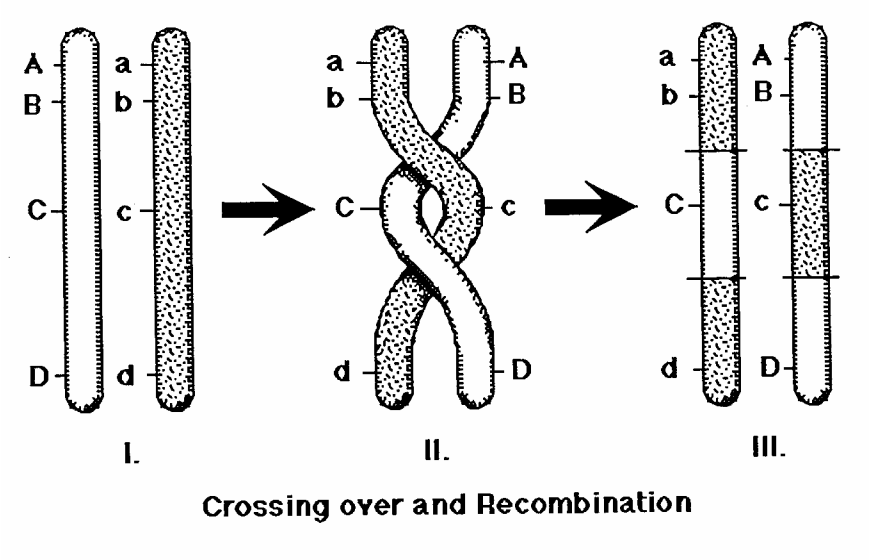
The proportion **r** (which is bounded between 0 and 25%) was found to be (1) constant for any given pair of mutations or factors, and (2) different for different pairs of factors. Most importantly, (3) **r** was independent of the starting combinations of genes in the parents. (Properties (1) and (3) were given by Haldane, 1919, as a *definition* of linkage.) Thus, from (3), if we start with the parental types **AAbb** and **aaBB** instead of **AABB** and **aabb**, the proportions of gametic types in the F2 generation are the exact complement of the original pattern:

(Parents: **AAbb, aaBB**):

Gametes:	**AB**	**Ab**	**aB**	**ab**
	r%	50-**r**%	50-**r**%	**r**%

Let us denote by **R** the proportion of gametic types not found in the parents, e.g., **Ab** and **aB** for parents **AABB** and **aabb**. (Obviously, **R** = 2**r**.) Something is causing reassortment of factors in **R**% of the cases to produce new gametic combinations. Furthermore, this proportion is not a function of what genes are found together in the parents, since the same proportion of **AB** and **ab** gametes are found if we start with parental types **AAbb** and **aaBB**. The hypothesis of the Morgan school (Morgan, 1911) was that the homologous chromosomes were winding around one another and then separating, exchanging corresponding segments.

In figure 6.1, factor **C** is separated from the others (**A, B,** and **D**) through such an intertwining and separation. Morgan suggested that factors that were further apart would be separated more frequently through such a separation. This explains the different values of **r** or **R** for different pairs of factors. Factors with a constant linear location along the chromosome would always separate with the same characteristic frequency. This explains the constant proportion of new types produced by both (**AABB, aabb**) and (**AAbb, aaBB**) parental combinations. Finally, if the factors (and their alleles) kept the same relativ



Crossing over and Recombination

e locations along the chromosome through such interchanges, **r** or **R** should be constant for a given pair of factors, and independent of what other specific alleles are found at other locations along the chromosome.

By 1913, six mutations had been localized to the x-chromosome, and Sturtevant noted that the recombination frequencies between pairs of factors were approximately additive. Thus, of the frequencies of recombinant types between factors, **R(AB), R(BC),** and **R(AC)**, one of them was usually equal to or slightly less than the sum of the other two. This suggested to him that he could use the basic intuitions of Morgan's 1911 paper to construct a map of their relative locations on the x-chromosome from the recombination frequencies between all pairs of the mutations, according to successive applications of the following basic scheme. If three factors, **A, B,** and **C** are arranged in a line, then one of them is between the other two. Which one is in the middle can be determined by seeing which of the following two conditions holds (approximately):

Equation:	Arrangement:
	A B C
(1a) R(AB) + R(BC) = R(AC)	•------•---------•
	A C B
(1b) R(AB) - R(BC) = R(AC)	•------•---------•

With this scheme of mapping and Morgan's proposed mechanism for the reassortment of different factors on different chromosomes, Sturtevant could not only give the relative locations of genes on the chromosomes, but also could explain a new phenomenon which became obvious once the factors were ordered in this way. While one or the other of the above two equations is always met approximately, sometimes the largest recombination frequency is slightly less than the sum of the other two. These deviations from strict additivity in the observed recombination frequencies between different factors are systematic. They occur only for larger recombination frequencies, and increase in magnitude with the size of the recombination frequencies. *This is an instance of the fourth function of false models given above.* It involves first of all recognition of a causal factor (relative location in the chromosome map) that causes behavior like that described in the above equations. Secondly, it involves recognition of deviations from this behavior that, thirdly, are attributed to the action of a causal factor not taken into account in the simple model.

Suppose that crossovers occur at random with equal probability per unit distance along the length of the chromosome map. Then for relatively short distances, in which we should expect only one crossover event, as between **A/a** and **B/b** or **C/c** in figure 6.1, the proportion of observed recombination events should increase linearly with the map distance. But if one is looking at a pair of factors that are sufficiently far apart that two or more crossovers have occurred between them, then the observed recombination frequency should be an *underestimate* of the map distance between the two factors. This is because a

double crossover between two factors (as with **A/a** and **D/d**) brings the factors **A** and **D** back to the same side so that they are included in the same chromosome. At this early stage in the development of their model, they had misidentified the stage in meiosis when crossing over occurs, but the basic argument remains the same today. Under these circumstances, if one is not tracking any factors between the two factors in question, an even number of crossovers will be scored as no crossovers, since the factors will come out on the same side, and an odd number of crossovers will be scored as one crossover, since the factors will come out on opposite sides. Thus an increasing proportion of recombination events will be missed as the distance between the two observed factors increases.

With this bias, factors that are further separated should show increasing deviations between observed recombination frequency and actual map distance. In 1919, J. B. S. Haldane gave a mathematical analysis of the problem and derived from first principles what has come to be known as the "Haldane mapping function," according to which the relation between observed recombination frequency (**R**) and map distance (**D**) is given by the equation:

(2) $$\mathbf{R = .5(1-e^{-2D})}$$

Haldane characterizes the mechanism underlying this equation as assuming an "infinitely flexible chromosome"(1919, p.299). He also describes the case in which multiple crossing over does not occur (producing an exact proportionality between recombination frequency and map distance) as assuming a "rigid" chromosome (p.300.) He also suggests an equation for behavior between these two extremes as representing the case of a stiff or partially rigid chromosome. This kind of mechanical argument underlying the derivation of a mathematical model is extremely useful, because it suggests sources of incorrectness in the idealizing assumptions of the models, and serves to point to specific factors that should be relaxed or modified in producing better models. *As such it exemplifies item 8 from the list of functions of false models.* Haldane does not have a preferred choice for how to model the behavior of a "partially rigid" chromosome, so the best that he can do is to present two largely unmotivated models for the intermediate case, one of which is an analytically tractable but rather *ad hoc* model and the second of which is explicitly just a "curve fitting" attempt for use in prediction.

The Haldane mapping function was either unknown to or at least was never used by the members of the Morgan school during the period of the debate with Castle over the linear arrangement of the genome in 1919 and 1920 (but see Wimsatt 1992)! This was possibly because their empirically determined "coincidence curves" for the various chromosomes, which gave frequency of double crossovers as a function of map distance, indicated the interaction of more complex and variegated causal factors determining multiple crossing over (see Muller, 1916a-1916d).

Whatever the explanation for their failure to use Haldane's model, they clearly had a sound qualitative understanding of the predicted behavior of their proposed mechanism from the first. Such an understanding is displayed already in Sturtevant's 1913 paper, where he pointed out that with the proposed mechanism, there should be a very small frequency of double crossovers even between factors that are quite close together. These could be detected by tracking a factor that was between the two factors in question (as **C** is between **A** and **D** in figure 6.1.) But he detected no close double crossovers, and the

frequency of crossovers between relatively distant factors was substantially less than expected if one crossover had already occurred in that interval. (The relatively distant factors were 34 map units apart, where one map unit is that distance for which 1% recombination is expected.) He therefore hypothesized an "interference effect" acting over a distance around a crossover within which a second crossover was prevented or rendered exceedingly unlikely. *Here is yet another application of the "template matching" function of models of item 4, where the deviation from the expected performance of the template is used to postulate another causal factor whose action causes the deviation.*

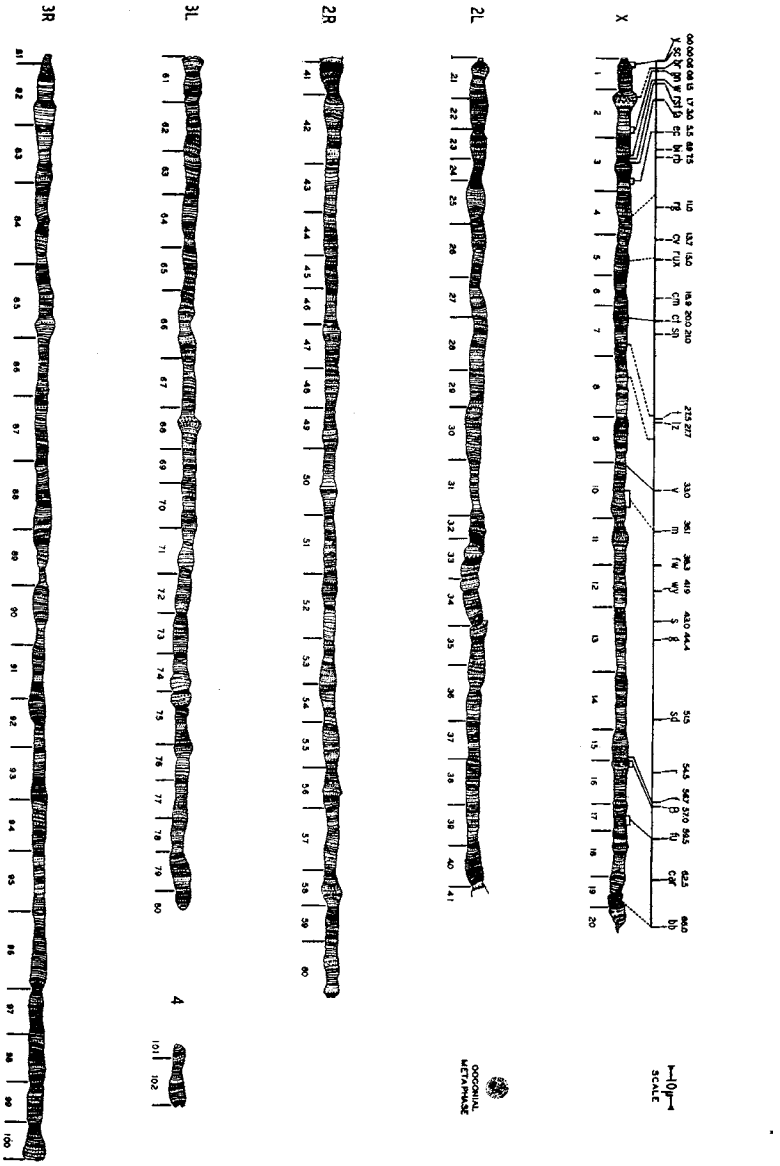
This additional hypothesis of an "interference effect" was required for the model of the Morgan school to account for the exact linearity found in map relations for close map distances (See figure 6.4). This was the subject of extensive investigations by Muller in 1915 and 1916, who considered the mechanisms of crossing-over, proposed a number of alternative mechanical hypotheses to explain this effect, and sought to gather data to choose among them (Muller 1916a-1916d). The mechanical model for recombination was a fruitful source of hypotheses on the cause of this phenomenon. For example, if we assume that chromosomes have a certain rigidity and torsional strength, it is clear that they can be wound up only so tightly (like a double-stranded rubber band) before they break. Thus chromosome rigidity and torsional strength determine a maximum tightness of the loops before breakage would occur. This in turn determines a minimum length of chromosome between crossover events--an "interference distance" (this is just a simple elaboration of the argument of Haldane 1919).

The occurrence of interference also had a beneficial secondary effect, since it meant that they could use recombination frequencies between closely linked factors as a direct measure of map distance, rather than correcting them for unobserved multiple crossovers. (If recombination events occurred at random along the chromosome map, and without interference, so that they were statistically independent, then the multiplicative law for the co-occurrence of multiple events would apply and for a distance in which there was a probability **p** of one crossover, there would be a probability $\mathbf{p^2}$ of 2 crossovers, $\mathbf{p^3}$ of 3 crossovers, $\mathbf{p^n}$ of n crossovers, and so on.)

To appreciate the character of the debate with Castle, it is important to realize the theoretical character of the parameter, "map distance." One could not infer from a chromosome map exactly where the factors were on the chromosome. As Sturtevant noted in 1913, the chromosome might have differential strengths and probabilities of breakage along its length, leading in general to a non-linear relation between map-distance and distance along the chromosome. Nor was it possible to determine which end of the map corresponded to which end of the chromosome without the production of aberrant chromosomes having visibly added or subtracted segments at an end, something that was not done until the 1930's.

In the case of *Drosophila*, the debates continued on a theoretical plane until the (re) discovery of Painter in 1934 of the giant salivary gland chromosomes, which had visible banding patterns whose presence or absence and arrangement could be visibly determined. This allowed the location of genes at specific banding patterns, and the ready detection of inversions, reduplications, translocations, and deletions. Some of these were hypothesized earlier solely on the basis of changed linkage relations--a remarkable (and laborious) triumph which makes the earlier work on chromosome mapping one of the most elegant examples of the interaction of theoretical and experimental work in the history of science. The subsequent localization of genes to bands by studying genetic deletions also permitted

confirmation of Sturtevant's conjecture that the mapping from genetic map to chromosome, while order-preserving, was non-linear: the frequency of crossing over was not constant from one unit of physical chromosome length to the next (figure 6.2).



Castle's attack on the "linear linkage" model:

Not everyone was enamored of the mechanical models of the Morgan school. A number of geneticists (including Bateson, Punnett, Goldschmidt, and Castle) attacked it on a variety of grounds in between 1913 and 1920, and Goldschmidt was still an outspoken critic in 1940 (See Carlson's (1967) excellent history). At the time they had what seemed like good reasons for these attacks, reasons that led them (and Morgan too--as late as 1909 and 1910; according to Allen 1979) also to attack the Boveri-Sutton hypothesis of 1903

that the genes were located on the chromosomes. The Boveri-Sutton hypothesis was a direct ancestor of the linear linkage models of the Morgan school.

These skeptics of the Boveri-Sutton hypothesis and the later models of the Morgan school were bothered that these models had nothing to say about gene action, but explained only correlations in the inheritance of traits. But the tradition from Aristotle down through the beginning of the 20th century was that the hereditary elements, whatever they were and however they worked, determined both the transmission of traits and their development. It seemed unreasonable that any theory dealing with the first of these could be correct: they sought a single theory that would explain both (see Morgan 1909, Allen 1979, and Carlson 1967). The Morgan school chose to defer consideration of this problem. Given the many successes of their research program, transmission genetics went on apace without having much if anything to say about development until the rise of molecular biology and the development of models of gene action like the operon model of the early 1960's.

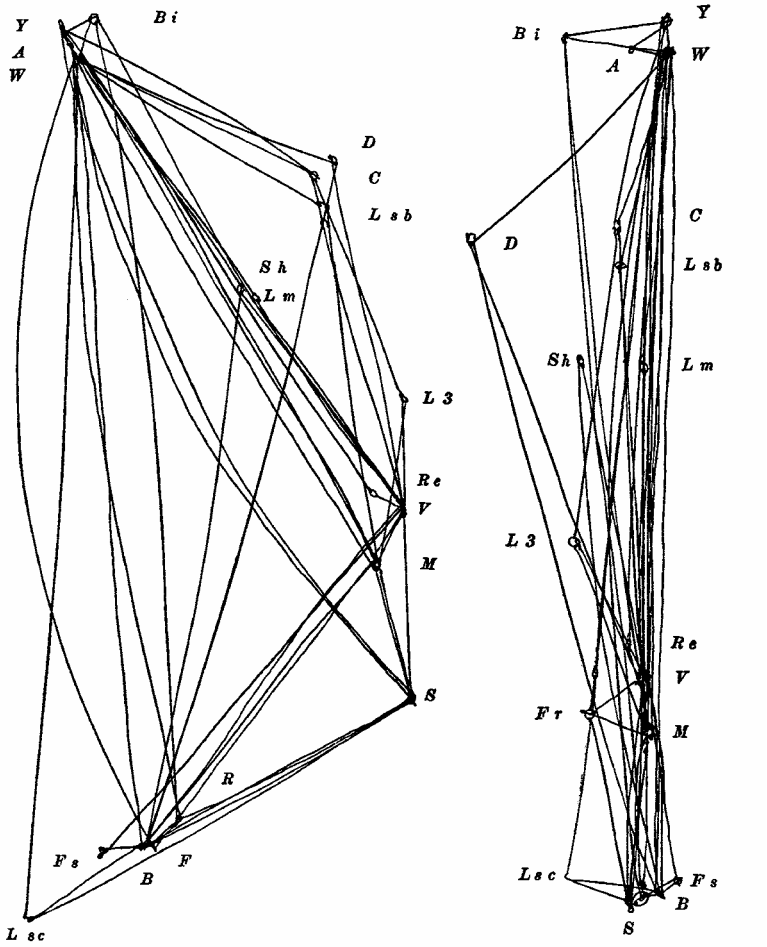
Two things characterized the theoretical attempts of these opponents: (1) all of their models attempted at least to make room for gene action (usually in a variety of different ways), and (2) they all seemed to retain a healthy skepticism, not only about the Morgan model, but about any conclusions which appeared to follow too closely from it. Virtually everyone accepted the data that it explained--that seemed too hard to object to, but this was combined with a healthy skepticism about their mechanical models. A frequent line taken by geneticists that outlived widespread opposition to the model of the Morgan school (found as late as Stadler 1954) was to be "operationalists" or "instrumentalists" about the gene. Given that they could trust the data from breeding experiments, but not the models produced to explain the data, it seemed more reasonable to avoid commitments to the theoretical models and to accept the genes only as "operationally defined" entities which produced the breeding results, in all cases staying as close to the data as possible.

Castle's operationalist attack on the model of the Morgan school is particularly interesting because the subsequent debate provides a clear example of the superiority of a mechanistic or realist research program over an operationalist or instrumentalist one. Some of the reasons why this is so will be apparent in the following discussion, but many of the most convincing cases and arguments would take us too far afield, and will have to be left for another occasion.

In 1919, Castle published his first attack on the "linear linkage" model. He found the linear model and the associated theoretical concept of "map distance" to be too complicated and too far from the data. He was suspicious anyway of the linear model, and complained that "it is doubtful . . . whether an elaborate organic molecule ever has a simple string-like form" (1919a, p. 26). He was bothered even more by the fact that map distances for 3 of the 4 Drosophila chromosomes exceeded 50 units, which he saw as inconsistent with the fact that observed recombination frequencies never exceeded 50%. (This apparent conflict arose from his "operationalist" assumption that map distance should be made proportional to recombination frequency and represents a simple misunderstanding on his part of what the Morgan school was claiming--see below.) To construct the map from recombination frequencies one had to invoke the possibility of multiple crossing over and posit "interference effects," and Castle also regarded both of these assumptions as dubious.

Given his "operationalist" preferences, Castle suggested that the simplest hypothesis would be to assume that the distance between the factors (whatever it signified) was a linear function of recombination frequency. He proceeded to construct a mechanical

model of the arrangement of the genes in the chromosome by cutting wire links with lengths proportional to the recombination frequencies and connecting them, producing the phenomenological model of the x-chromosome of *Drosophila* diagrammed in figure 6.3.



(A phenomenological model (Cartwright, 1983) is one which describes the data without explaining it.) He thereby claimed to have produced a model that fit the data and did so without making the further hypotheses of double crossing over and interference effects. In spite of his "operationalist" stance, he went on to suggest and use two alternative (and mutually inconsistent) mechanistic models to interpret the significance of his construction and to attack the linear linkage model. (Like most other operationalist opponents of the Morgan school, Castle seemed to be a frustrated mechanist!) The force of the counterattacks of the Morgan school was to show that (1) the data Castle used did not support his model, (2) even if he used the data acceptably, which he did not, and (3) that his model would fare even worse if more data were included. Furthermore, they showed (4) that neither of the two alternative mechanisms he proposed to account for his phenomenological model would have the desired effects.

Castle's critique and a later one drew multiple responses, including those by Sturtevant et al. (1919), and by Muller (1920). Muller's response covered virtually all of the ground of the earlier authors and went much further. His critique is elegant, covering not only the selection of the data and the reasons why it had to be treated differently for

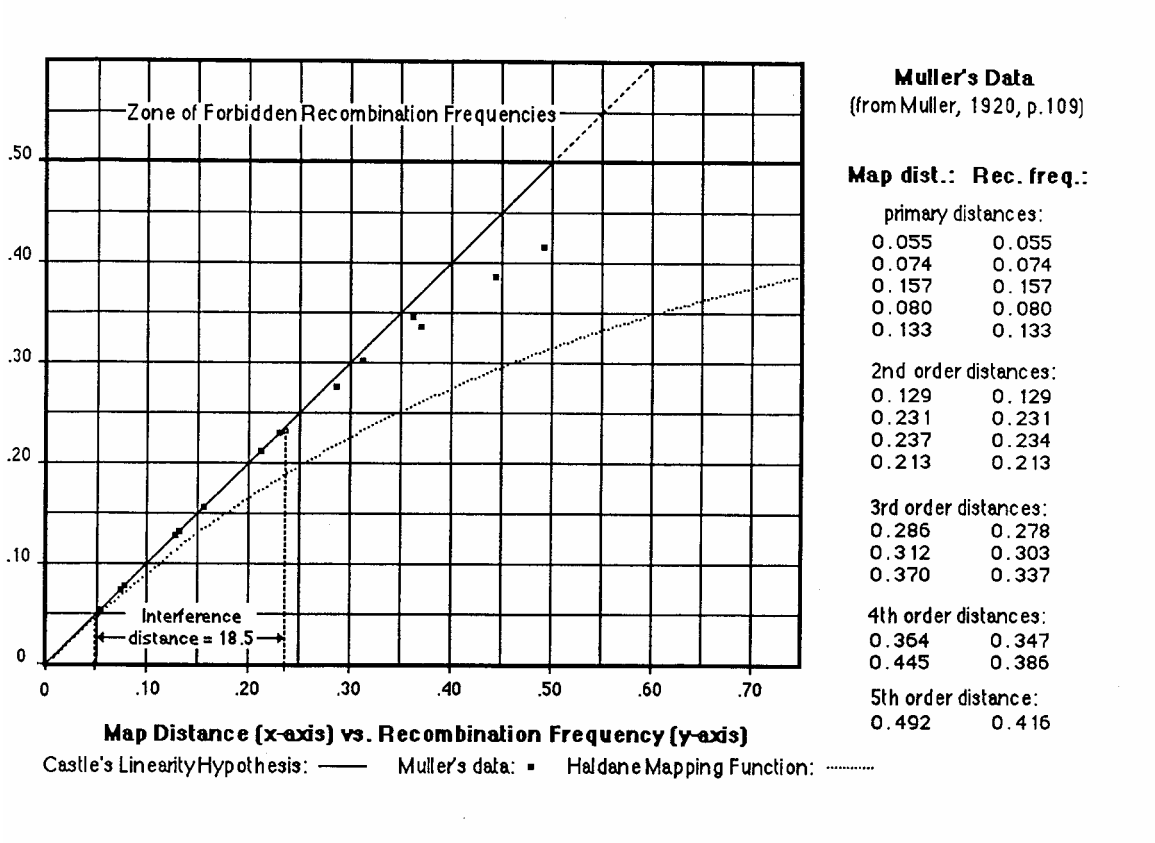
map construction and for testing the hypothesis of linear arrangement, but also an explanation of why Castle's model worked as well as it did (in spite of its incorrectness), predictions that it could be expected to break down and where, a defense of the reasonableness (and inevitability) of interference and multiple crossing over, and a complex set of arguments that Castle's claims were not only inconsistent with the data, but were internally inconsistent as well. His critique included a number of additional arguments not discussed here that further strengthen the case for the chromosomal mechanics of the Morgan school.

Muller's data and the Haldane mapping function:

In his 1920 paper, Muller uses data derived from his own earlier experiments (Muller 1916a-1916d) to argue for the inadequacy of Castle's model. (Wimsatt 1992 discusses this work extensively). These data are of the observed frequencies of recombination between each possible pair of six of the factors on the x-chromosome. In the model of the Morgan school, the recombination frequencies between nearest neighbor factors are assumed to be the map distances. Deviations from additivity for frequencies of recombination between more distant factors (see discussion of Sturtevant above) are seen as products of multiple crossovers between the observed pair of factors. The greatest distance between nearest neighbors among the factors Muller chose is 15.7 map units, and the lowest distance for which non-additivities are observed is 23.7 map units. Thus the assumption that nearest neighbor distances are real map distances, and not underestimates, seems justified. Another problem emphasized by both Castle and Muller is that the rarity of recombination events between very closely linked factors could make very small distance estimates unreliable. Muller obviated this objection by picking 6 factors (out of 12 for which he had data) which were neither too closely nor too distantly spaced along the chromosome map.)

In figure 6.4 I have graphed the relation between recombination frequency and map distance to be expected in 2 cases: (1) the linear relation between the two supposed by Castle's model and (2) the curvilinear relationship expected for the Haldane mapping function (or HMF). With the HMF, the relationship for small distances is approximately linear, but deviates from linearity with increasing map distance and asymptotes for very long distances at 50%. This suggests that factors very far apart on a long chromosome are essentially uncorrelated in their inheritance. (This is a new prediction of the linkage analysis, since it suggests that two factors can be uncorrelated in their inheritance *either* by being on different chromosomes (as before) *or* by being sufficiently far apart on a chromosome with a long map.)

Both of these curves are to be compared with Muller's data, which suggest a relationship which remains essentially linear for much larger map distances than with the HMF (as a product of interference effects), but which also appears to asymptote at a recombination frequency of 50% for very great map distances as with the Haldane mapping function. This actually cannot be seen in Muller's data, [and omits significant complexities, Wimsatt 1992] since the most distant factors he lists do not have very large separations. It can be seen for the data in Haldane's graph (1919, p. 309), which interestingly also shows much greater scatter in the points--largely a product of the fact that Haldane took data from different experiments. (The preceding omits significant complexities discussed in my 1992.) As Muller (1920) argued against Castle, this practice



may show confounding effects due to changing linkage relationships (produced by different translocations, inversions, or deletions in the stocks of flies used in the different experiments), different temperatures, and other factors affecting recombination frequency.

Notice several things here:

(1) The actual data Muller gives behaves qualitatively like the HMF, and might well be mistaken for it if the HMF were not graphed with it for comparison, pointing to the systematic discrepancies. The HMF is being used here as a template, deviations from which point to the importance of interference, which is a causal factor affecting the data but not taken account of in the mechanisms supposed in deriving the HMF. It is being knowingly used as a model that is false because it leaves out a known causal factor, which produces systematic and increasing deviations from the predictions of the HMF with increasing map distance. *This is an instance of the fourth function of false models*. It was done over and over again by members of the "fly group" (Wimsatt 1992.)

(2) If the data curve did not look qualitatively like that of the HMF, there would be less temptation to use it as a template and more temptation to say that it fundamentally misrepresented the mechanisms involved, rather than being false by being merely incomplete. (In that case the partitioning of the effects of causal factors along the lines suggested by the model could not be trusted.) Given their qualitative similarity (in particular, the linearity of **R** and **D** for small distances and the asymptotic approach of **R** to 50% as **D** gets very large) this temptation is strong however. The tendency to treat the HMF as a "baseline" model is further increased by the fact that Haldane presents two other models, one for "rigid" chromosomes (in which **R** is supposed to be linear with **D**, at least up to 50%), and one for "semi-rigid" chromosomes in which interference has an effect intermediate between the two extremes. Muller's data also

falls in between these two extremes (though not exactly on the curve of Haldane's intermediate model) and is thus plausibly treated as due to the joint operation of factors found in the HMF model together with the operation of some kind of interference mechanism (see function 8 of false models in the above list).

(3) The discrepancies between the HMF and the data can actually be used to get a rough estimate of the interference effect. Thus, if one observes where Muller's data begins to deviate noticeably from the line **R = D** (at 23.4) and estimates where the HMF begins to deviate comparably (at about 5), one can get a rough estimate of interference distance as the difference of these two numbers--about 18 or 19 map units. (I subtract 5 from 23.4 is to correct for the fact that double crossovers at close separations may be present in too small a frequency to be detected, but one can use the "measurability lag" for the HMF to estimate that for the real data. *This is another use of a model as template to calibrate a correction factor*, an additional use not on my list since it does not involve the use of false suppositions of a model.) This estimate is a brute empirical estimate from the behavior of the data, not one from any assumptions about the mechanism or mode of action of interference, but illustrates how *a model can be used to estimate a parameter that is not included in it, item 5 from the list of functions of false models.*

Interestingly, the mechanisms of interference have proven since to be quite refractory to analysis in terms of general equations which both apply to a variety of cases and which are derived from underlying mechanical models--probably because the mechanisms are both mathematically complex to describe and vary substantially from case to case. (Thus, the location of the two factors relative to the centromere is a crucial determinant of their recombination behavior, a factor not considered in any of these models.) As a result, many more recent treatments, such as those of Kosambi (1944), Bailey (1961), and Felsenstein (1979) have constructed phenomenological mathematical models with no mechanical underpinning, but which have other advantages. These include (1) providing a relatively good fit with known data and (2) a schema for prediction, (3) generating important prior models (such as Haldane's) as special cases, (4) having nice formal properties, and (5) producing nice operational measures which can be applied to more complex breeding experiments. *The third is an instance of the eleventh function of false models.* Felsenstein (1979) explicitly mentions the last 4 of these. With possibly unintended irony, he summarizes the advantages of his phenomenological model over other more realistic approaches:

> There are a number of papers in the genetic literature in which specific mapping functions are derived which are predicted by particular models of the recombination process . . . While these models have the advantage of precision, they run the risk of being made irrelevant by advances in our understanding of the recombination process. In this respect the very lack of precision of the present phenomenological approach makes it practically invulnerable to disproof. (1979, p. 774)

Felsenstein's more extended defense both illustrates part of the ninth function of model building (predictive adequacy), but also claims that the lack of underlying mechanical detail may make it more robust in the face of new theory (see also Cartwright, 1983). If his model is more immune to falsification, this "advantage" is purchased at the cost of a descent into the possible abuses of "curve fitting," which Felsenstein warns us

against by explicitly noting the "phenomenological" character of his model. As I will show however, being a phenomenological model is no necessary guarantee against falsification.

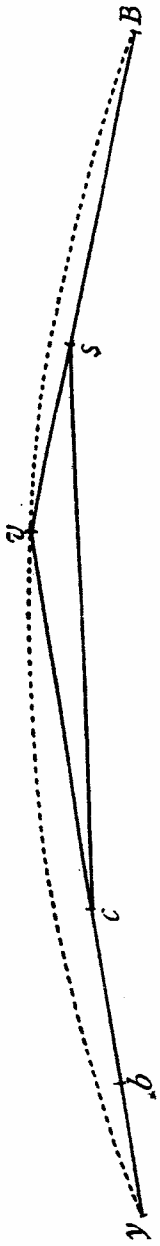
Muller's "2-dimensional" arguments against Castle:

Castle's model generates a 3-dimensional figure because the recombination frequencies between more distant factors are less than the sum of recombination frequencies between nearer factors, a feature earlier referred to as "non-additivity." This produces triangular figures for each triple of factors which, when connected together, produce complex polyhedral structures in 3 dimensions, as in figure 6.3. If we look at three factors at a time, **A, B,** and **C,** and arrange them so that they are at distances proportional to their pairwise recombination frequencies, the fact that **R(AC) < R(AB) + R(BC)** implies that **A, B,** and **C** will be at the vertices of a triangle. (This relationship is the "triangle inequality" of plane geometry.) As the deviation gets smaller, the extreme angles in the triangle get smaller, producing for **R(AB) + R(BC) = R(AC)** the "degenerate triangle" of a straight line. Only if **R(AB) + R(BC) < R(AC)** would it be impossible to construct a triangular figure with straight edges.

If we consider another factor, **D**, such that **R(BC) + R(CD) > R(BD)**, we get another triangle, which must be placed on top of the first, since they share side **R(BC)**. The addition of this triangle to the first generates a new distance, **R(AD)**, which provides a prediction of the recombination frequency between **A** and **D**, which can be compared with the data. In this way, (with successive applications of this construction) Muller constructed a map of the 6 factors, according to Castle's principles. Surprisingly, his map is 2-dimensional, rather than 3-dimensional as Castle required! Why was it legitimate to use a figure of lower dimensionality, which would allow the use of an additional degree of freedom for fitting the data (Figure 6.5)?

(1) Muller argues (1920, p. 113) that his (2-dimensional) figure is the only possible one using the "strongest and second strongest" linkages—that is, using the **R**'s for nearest neighbor and next nearest neighbor factors. This is true for a reason that he does not explicitly mention: because he has used relatively tight linkages (sufficiently close to prevent double crossovers in all cases but one), the "triangle inequality" holds for only one triple of factors--for all others there is an equality. In this special case, we get a "degenerate" figure that is only 2-dimensional, because there is a "bend" at one factor and all of the other relations are linear. Two non-collinear straight lines connected at one point determine a plane. The addition of any more data that had non-additivities would have required (at least) 3 dimensions.

(2) The argument that Muller emphasizes most strongly is that the distance between the most distant factors (with 4 factors in between) on the model constructed according to Castle's principles is too great--it is 49.3 as opposed to his observed recombination frequency of 41.6. (The predicted distances are also too great for the factors that are 3 or



Direct representation of the strongest and second strongest linkages in Table II. (*y bi, bi cl, cl v, v s, s B*, and *y cl, bi v, cb s, v B*, are each represented by a line of length proportionate to the respective frequency of separation.) The dotted curve shows the "average angular deviation" of the line of factors, according to this system.

2 factors apart, as he notes.) He claims that if the figure is bent so as to make the longest distance correct, (his figure 4, p. 113) then the shorter distances become too short, and the model still fails to fit the data. Thus, as he argues earlier, "the data . . . could not be represented either in a three-dimensional or in any other geometrical figure" (p.112).

(3) We can go further than Muller does explicitly, as might be suggested by Muller's references in 2 places to higher-dimensional figures. Suppose that there were non-linearities or bends at two places, but that they were sufficiently small that the predicted recombination frequency between the two most distant factors was still too large. Then

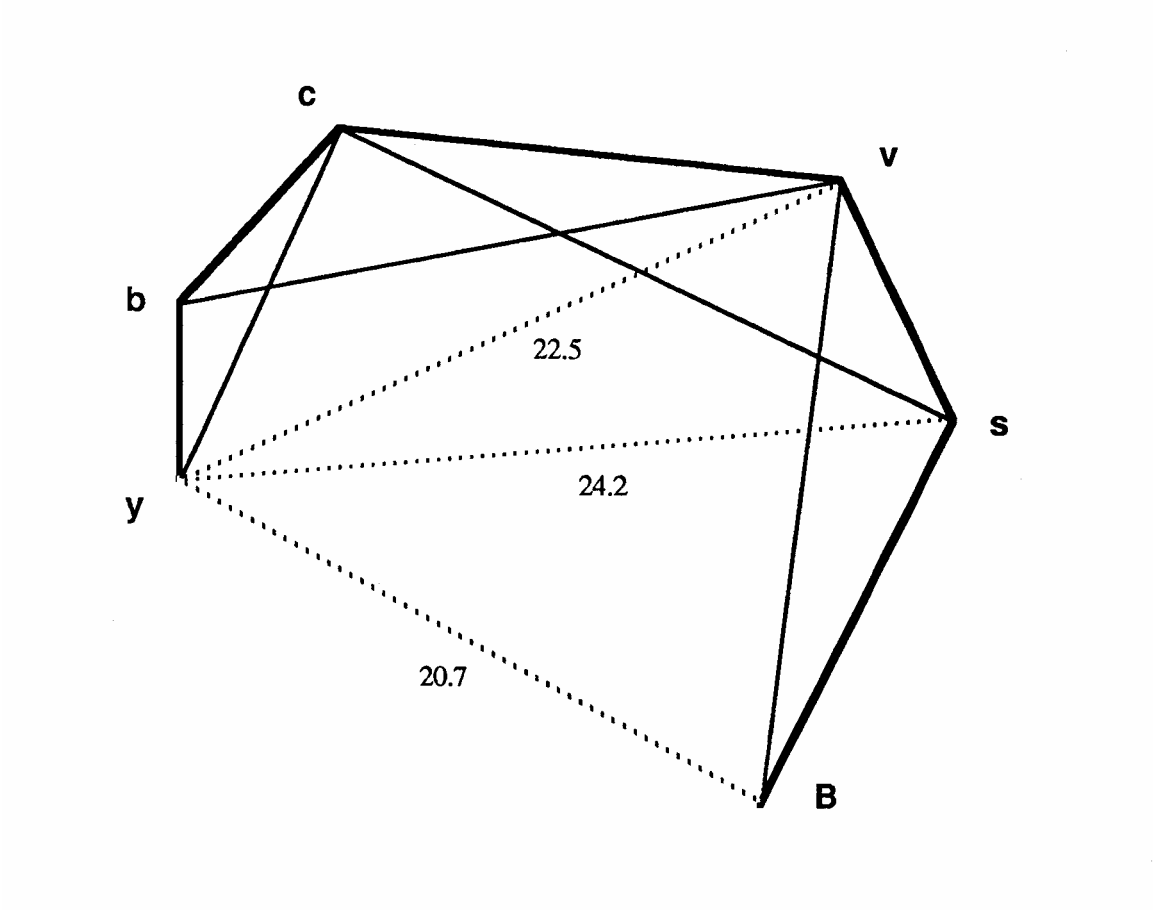
going to three dimensions (i.e., by rotating a part of the figure out of the plane) would not help Castle, since this could serve only to *increase* the distance between the factors, and thus the error in his predictions, if they had already been geometrically arranged in 2-dimensions so as to minimize the distance and predictive errors. If the predictions had been systematically too small, then going to a three-dimensional figure would help, but if they were already too large, as they were, then nothing would work. Even if the greatest distance were too small, one would still have to fit the shorter distances and, if they are supposed to represent physical distances, one is limited to a maximum of 3 dimensions. In any case, this procedure would obviously be nothing more than a "curve (or polyhedron!) fitting" technique, as Muller's arguments suggest.

Here we see arguments in which simpler models (a 2-dimensional model with a reduced data set) are used to draw conclusions about more complex ones (3-dimensional models with larger data sets). In particular, the argument is that given the data in question, no geometrical figure (in any number of dimensions) could consistently represent the data as a set of distances without bending wires or other such "cheating." (Note that Castle's model (in figure 6.3) does contain several bent wires, which represent incorrect predictions, an inconsistency noted by Muller.) *This is an example of the sixth function of false models, in which a simpler false model is used to argue for the inadequacy of a family of more complex related models.*

Multiply-Counterfactual uses of False Models:

We can take one more step, in this case beyond the grounds of the actual dispute, to illustrate a different kind of use of false models to counter a claim of Castle that he can do without the hypothesis of interference distance. We saw that Muller argued that Castle's triangular models produced predictions for recombination frequencies that were too great at larger distances. I will now show that if we transform Muller's data (counterfactually) to produce data that would be expected in the absence of interference, the fit of the data with Castle's model is far worse even than before, with much larger errors in the opposite direction. Thus Castle cannot complain about the Morgan school's use of interference effects, since this additional argument shows that he needs it even more than they do.

I pointed out earlier that interference was very useful to the Morgan school since it allowed them to assume that recombination frequencies between factors which were fairly close together on the chromosome map were true measures of their relative map distances and not biased underestimates as would be the case if multiple crossovers between the factors were possible. Suppose that we drop this assumption, and assume that close distances also are subject to the systematic bias produced by multiple crossovers. Then the "true" map distances (in a world with "infinitely flexible" chromosomes, and thus no interference) could be calculated from the Haldane mapping function, and these new transformed distances could be used to construct a chromosome map according to Castle's model. Such a map is constructed in figure 6.6 using Muller's data transformed according to the HMF.



In this 2-dimensional figure, there is a bend at each factor because the triangle inequality is satisfied for each triple of nearest neighbor factors because of the nonlinearities introduced by the (supposed) multiple crossovers. These non-linearities make the map bend so strongly that the predicted recombination frequencies for more distant factors actually *decrease* as the map curves around on itself (compare distances **ys** and **yB**). This data transformation demonstrates that without interference, Castle's model would have been even more strongly inconsistent with the data. Three dimensions would have helped Castle in this case, but unless the irregular helix which would be produced for a long chromosome by bending in a third dimension was very elongated, one would have had a distance function which oscillated for increasingly distant factors, one again inconsistent with data available at the time. Furthermore, a choice of other intermediate factors would have given different relative locations for the two factors at the ends.) Castle could, of course, have bent successive distances in alternate directions or according to some other still more irregular scheme, but would have had to do so in a way that was consistent with all of the other shorter distances, which are not used in this demonstration. At best, the construction would have been very *ad hoc*, and the addition of any new factors would have produced new problems.

In this argument, the use of false models is very circuitous indeed. True data is transformed according to a false model (the Haldane mapping function) to produce false data, which is then plugged into a false (2-dimensional) but simpler version of a (false) 3-dimensional model to argue that the worsened fit of that model with the data undercuts Castle's attack on the Morgan school's use of interference by showing that he needs it even

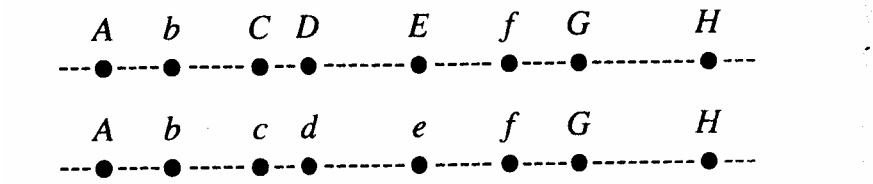
more than they do. (This is a kind of *ad hominem* argument, but one that is entirely legitimate, because it uses false premises that Castle wishes to accept. It is a kind of *reductio ad absurdum* argument, since it involves accepting a premise that they wish to reject to undercut an argument against a premise that they wish to accept.) *This example illustrates the seventh listed function of false models.* It is contrived, since they did not actually make this argument, though it is certainly one that they could have made with the data and models at hand, and it would not have been far from the spirit of either Haldane's analysis or that of Muller to do so. This case is closer to the spirit of many cases of "neutral" model building, where constructed ("random") data sets are used to test hypotheses about selection.

False models can provide new predictive tests highlighting features of a preferred model:

In 1917, Goldschmidt (better known for his later suggestion that macro-evolution took place through the incorporation of macro-mutations or "hopeful monsters") published a paper criticizing the "linear linkage" model. He proposed an alternative explanation for the factor exchanges in crossing over in which factors were tied to specific locations on the chromosome through biochemical forces. Corresponding allelic factors left the chromosome to accomplish their activities in the cell, an attempt to accommodate gene action. They then returned to their places for the chromosomes to undergo any mitotic or meiotic activities. Each factor had a specific "force" which acted with its binding site, and different alleles had somewhat different forces. The differences between forces for alleles at the same locus were much less than the differences between any one of them and the forces of factors at other loci. When they later reassembled on the chromosome, the similarity of allelic forces with their respective binding sites resulted in occasional interchanges between them. The differences between non-allelic factors were assumed to be too great for non-allelic interchanges. The greater the differences between allelic factors, the stronger would be the (apparent) "linkage" of those factors with the others on the same chromosome, since they would have less chance of going to the binding site on the homologous chromosome.

Sturtevant (1917) argued that the pattern of multiple crossovers ruled out Goldschmidt's proposed explanation. On Goldschmidt's model, factor exchanges between alleles at one locus should be uncorrelated with factor exchanges between alleles at other loci, since their occurrence in each case was simply a product of the similarity of alleles at the same locus, and should be independent of events at any other loci. On the model of the Morgan school, the intertwining and linear linkage that produced factor exchange suggested that if a factor crossed over then other factors near it should also cross over. Thus, if we start with homologous chromosomes:

the predicted outcomes of multiple crossovers (looking only at the first chromosome) would be qualitatively as follows:



The data available to them clearly supported the "linear linkage" model.

Bridges (1917) made another crucial criticism, that on Goldschmidt's model, factors with a frequency of crossing over of 1% in one generation (and whose configuration of forces were stable) should have a 99% frequency of "crossing back" to their original location in the next generation. This was at odds with the stability of crossover frequencies for a given pair of factors, and in particular with the independence of these frequencies from the parental types used in the original cross (See the third defining property of linkage of Haldane 1919).

Of these two predictions (both discussed in Carlson, 1967), the first appears not to have been made before the appearance of Goldschmidt's model. (Muller discusses it in 1916, with a reference to what could be this model, attributed to "some geneticists." The Morgan school likely had knowledge of the Goldschmidt model prior to publication, since *Genetics* was a "house" journal for them at that time.) The second made use of an already known property of linkage phenomena, but gave this property a greater importance. *These thus illustrate the third function of the use of false models--to generate new predictive tests of or to give new significance to features of an alternative preferred model.*

Note that Goldschmidt's model required a large array of distinguishable forces which all met the condition that allelic forces were more alike than non-allelic ones. This requirement raised potential problems for mutations that sometimes produced substantial phenotypic effects without changing their linkage relations. Muller used this fact later to argue that the mechanism of gene transmission must be independent of the mechanism of gene expression. Goldschmidt's model would also undercut the advantages for the arrangement of factors in chromosomes first hypothesized by Roux in 1883 and widely accepted since then. In Roux's account, chromosomes were seen basically as devices to increase the reliability of transmission of factors to daughter cells in mitotic division by putting many factors in one package and thus reducing the number of packages which had to be assorted correctly to transmit each factor in mitosis. The linear chromosomal organization of factors gave a simple mechanical solution to the requirement of getting a complete set of genetic factors without having to take care of each and every factor as a special case. The Goldschmidt model would have restored the need for special forces and arrangements for each factor.

False models and adaptive design arguments:

One other common use of false models would seem to be special to evolutionary biology and to other areas like economics and decision theory where optimization or adaptive design arguments are common. Where various forms are possible, and one seeks an explanation for why one is found rather than any of the others, one may build models of the alternatives and specify fitness functions of their structural properties such that their fitnesses can be estimated and compared. *When the fitnesses of the other alternatives are significantly less than that of the type found in nature, we have a prima facie explanation for why they are not found--at least if they are assumed to arise only as variants of the given type, or it from them.* **Technically,** we do not have an explanation for the occurrence of the given type: we have shown only that it would be adaptive if it *did* arise, not that it would do so). Nor do we necessarily explain the absence of the other types (if they can arise other than in selective competition with the given type). And other factors could independently prevent their evolution or render these variants maladaptive. Nonetheless, such thought experiments can illuminate.

Consider a simple thought experiment that provides an explanation for why there are no species with three (or more) sexes. (This topic was the subject of a paper in the early 1970's by A. D. J. Robertson, which I read at that time but that I have been unable to find since and remember at best unclearly. Thus, some, all, or none of this argument may be due to Robertson.)

We must first distinguish two things that could be meant by more than two sexes. It might mean several sexes of which any two could have fertile offspring, but mating was bi-parental. This system could be called "disjunctive," since there are a variety of possible types of successful mating. Alternatively, it might mean that 3 (or **n**) different types of individuals are all required to produce fertile offspring. This system could correspondingly be called "conjunctive," since only one mating type is successful, and this requires a conjunction of **n** individuals. The first type of system is uncommon, but is found occasionally. The wood rotting fungus, *Schizophyllum communes* has a variety of (bi-parental) mating types, and almost any two types will do. There are two genetic complexes, A and B, with 40-50 variants at each "locus." To be compatible types, two individuals must have different variants at each of the two loci (see King, 1974). The second type of system does not exist in nature. Various considerations could explain why this is so. I consider only one.

I will assume the simplest kind of interaction among members of a species, in which matings are modeled on collisions (or on linear reaction kinetics.) Suppose that an individual has a probability **p** of encountering another reproductively potent individual during the time when it can mate successfully. This probability is a partial measure of fitness, which is presumably a product of it with several other factors. With a two-sex system, if the sexes are equally frequent, there is a probability of **p/2** that an individual can have a pairing which will produce offspring--one with the opposite sex. With a disjunctive **n**-sex system (where any two can play, if they are of different sexes), the probability of finding a mate becomes **((n-1)/n)p**, since an individual can mate with any other sex, and all other sexes constitute a proportion **(n-1)/n** of the individuals it could meet. This quantity is greater than **p/2** for **n > 2**, which shows that systems with more sexes have an advantage. This could be one factor explaining the multiplicity of types in *Schizophyllum communes*, and raises the interesting question why it is not more common. Presumably there are other constraints in addition to those considered in this model.

A "conjunctive" multi-sex system represents an even more interesting kind of thought experiment, yielding an explanation of why such systems have *not* been found in nature. If one sex of each type must get together during their reproductive period, the probability of this occurring is $\mathbf{(p/n)^{n-1}}$, since an individual would be expected to meet individuals of each other sex in the relevant period with a probability of **(p/n)** per sex, and there are **(n-1)** other sexes. (I assume that meetings are statistically independent. Note also that if the sexes are not equally frequent, the probability of the right conjunction occurring is even smaller, so that in this respect, this is a "best case" analysis.) $\mathbf{(p/n)^{n-1}}$ is always less than **p/2** for **n > 2,** and very much smaller if **p** is much less than 1. If **p =.5**, an individual of a 2-sex species has a chance of 1/4 of finding a mate, an individual of a conjunctive 3-sex species has a chance of 1/36 of finding its 2 mates of the right sexes, and an individual of a conjunctive 4-sex species, has a chance of only 1/512 of finding its necessary 3 mates of all the right sexes. Note that slight extensions of these arguments could be deployed against more complex mixed "disjunctive-conjunctive" systems, such as one in which any **k** out of **n** sex types could reproduce conjunctively, since whatever

the value of **n, k** = 2 is selectively superior to any larger value of **k**. A classic puzzle of sexuality is also preserved in this model, since an asexual species obviously does better in this respect than any of these more complex systems, including the familiar 2-sex system.

This brutally simple model thus serves to explain why there are no conjunctive 3-or-more sex mating systems. Of course, there are other reasons: the classic work of Boveri (1902) on the catastrophic consequences of simultaneous fertilizations of a single egg by two sperm shows that other constraints point in the same direction. Without substantial re-engineering of the mitotic and meiotic cycles, it is hard to imagine how a conjunctive 3 or more sex system could ever evolve from a 2-sex system. The mitotic and meiotic cycles are deeply "generatively entrenched"--they are effectively impossible to modify because so many other things depend upon their current forms. This is a paradigm of the kind of developmental constraint I discussed in the third section above, and in more detail in my 1986b. See also Arthur, (1984), Rasmussen (1987), and Schank and Wimsatt (1988).

This almost playful example illustrates a point which is of broad importance in evolutionary biology, and equally applicable in feasibility or design studies in engineering. In the latter case, much effort is expended to model the behavior of possible systems to determine whether they should become actual. If they are found wanting, they never get beyond design or modeling stages because the proposals to build them are rejected. The models describing their behavior are "false" because the conditions of their true applicability are never found in nature, not because they make approximations or are idealizations which abstract away from nature (though they are surely that!), but because their idealizations represent *bad designs* which are maladaptive.

Obviously false models have an important role here. It is partially for this reason that much good work in theoretical evolutionary biology looks more than a little like science fiction. This is nothing to be ashamed of: thought experiments have a time-honored tradition throughout the physical sciences. The main advantage that evolutionary biology has is that the greater complexity of the subject matter makes a potentially much broader range of conditions and structures the proper subject matter of biological fictions.

Such biological fictions are interesting because one cannot always tell if nothing like the model is found in nature, whether this is because of the idealizations required to get models of manageable complexity, or because of selection for superior alternatives. An example is provided by discussions of the importance of chaotic behavior in the analysis of ecological systems which I reviewed literature in (1980a). I argued that the absence or rarity of chaotic behavior in ecological systems did *not* show that avoidance of chaos was not a significant evolutionary factor. It seemed likely that there were evolutionarily possible responses that could act to control or avoid chaotic behavior. If so, we would expect selection to incorporate these changes because chaotic behavior is generally maladaptive. There is not only the possibility of at least local extinction, but also the fact that major fluctuations in population size, sex-ratio, and effective neighborhood can easily generate much smaller effective population sizes and lead to substantially reduced genetic variation (Wimsatt and Farber, unpublished analysis, 1979). Reduced genetic variation could lead "chaos-prone" populations either to increased probability of extinction in changing environments, or to their more rapid evolution via mechanisms suggested in Wright's "shifting balance" theory (Wright, 1977).

After many years in which the common wisdom has seemed to be that chaotic behavior was rare in nature, more recent work (e.g., by Schaffer, 1984) suggests that chaotic behavior may be much more common than we suspect, and can be seen with the

right tools. The kinds of biases discussed in Wimsatt 1980a would tend to hide chaotic behavior, even in cases where we have the good data over an extended period of time necessary to apply Schaffer's analysis. Thus it may be more common than even he suggests.

Another use of false models not discussed here is the Richard Levins' suggestion (1966, 1968) that constructing families of idealized and otherwise false models could allow the search for and evaluation of "robust theorems"--results that were true in all of the models and thus independent of the various false assumptions made in any of the models. (This is the motivation for his comment that "Our truth is the intersection of independent lies.") I have discussed this approach elsewhere (Wimsatt, 1980a, 1981). Taylor (1985) provides a study in depth of its use in ecological modeling, and a more sophisticated analysis of the strengths and limitations of different alternative approaches.

Summary:

Neutral models in biology represent "baseline models" or "null hypotheses" for testing the importance or efficacy of selection processes by trying to estimate what would happen in their absence. As such they often represent the deliberate use of false models as tools to better assess the true state of nature. False models are often used as tools to get at the truth. In this paper I have analyzed and illustrated the many ways in which this can be done, using cases from the history of genetics, drawing on the development of the "linear linkage" model of the Morgan school, and the different ways in which this model was used in countering contemporary attacks and competing theories. It is often complained by philosophers of science that most models used in science are literally false. Some of them then go on to argue that "realist" views of science in which arriving at a true description of nature is an important aim of science are fundamentally mistaken. I have tried to illustrate the variety of ways in which such deliberate falsehoods can be productive tools in getting to the truth, and thus to argue that these philosophers have despaired too quickly.[1]

[1] I thank James Crow and Bruce Walsh for useful feedback on the talk at the symposium, and discussions and references on modern theories and models of interference phenomena. Nils Roll-Hansen and another anonymous referee made suggestions on organization and content, and Matt Nitecki helped to clarify and streamline sticky prose. Janice Spofford and Ed Garber helped me find data on "non-standard" mating systems. I also thank a decade's students in my biology class, "genetics in an evolutionary perspective," who have acted as sounding boards for some of these ideas and as guinea pigs for the computer labs which grew out of them. Bill Bechtel, Jim Griesemer, Ron Laymon, Bob Richardson, Sahotra Sarkar, Leigh Star, Leigh van Valen, and Mike Wade all gave useful comments at earlier stages. Even though her *How the Laws of Physics Lie* is mostly in service of a different conclusion, Nancy Cartwright's beautiful study of model building in physics provides many further examples of the sort discussed here, demonstrating that this analysis is not limited to biology. Support from the National Science Foundation and the Systems Development Foundation, for time and computing equipment, made parts of the task substantially easier.