



**HAL**  
open science

# Models of data and theoretical hypotheses: a case-study in Mendelian genetics

Marion Vorms

► **To cite this version:**

Marion Vorms. Models of data and theoretical hypotheses: a case-study in Mendelian genetics. *Synthese*, 2012, 190 (2), pp.293-319. 10.1007/s11229-012-0147-2 . halshs-01122296

**HAL Id: halshs-01122296**

**<https://shs.hal.science/halshs-01122296v1>**

Submitted on 14 Dec 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## **Models of Data and Theoretical Hypotheses: A Case-Study in Classical Genetics**

**Marion Vorms**

Published in *Synthese*, 2012, 190(2): 293-319

### **Abstract:**

Linkage (or genetic) maps are graphs, which are intended to represent the linear ordering of genes on the chromosomes. They are constructed on the basis of statistical data concerning the transmission of genes. The invention of this technique in 1913 was driven by Morgan's group's adoption of a set of hypotheses concerning the physical mechanism of heredity. These hypotheses were themselves grounded in Morgan's defense of the chromosome theory of heredity, according to which chromosomes are the physical basis of genes. In this paper, I analyze the 1919 debate between William Castle and Morgan's group, about the construction of genetic maps. The official issue of the debate concerns the arrangement of genes on chromosomes. However, the disputants tend to carry out the discussions about how one should model the data in order to draw predictions concerning the transmission of genes; the debate does not bear on the data themselves, nor does it focus on the hypotheses explaining these data. The main criteria that are appealed to by the protagonists are simplicity and predictive efficacy. However, I show that both parties' assessments of the simplicity and predictive efficacy of different ways of modeling the data themselves depend on background theoretical positions. I aim at clarifying how preference for a given model and theoretical commitments articulate.

## 1. INTRODUCTION

Linkage (or genetic) mapping is a representational technique, which was first designed in 1913 by Thomas Morgan's student Alfred Sturtevant. Linkage maps are graphs, constructed on the basis of statistical data obtained through Mendelian breeding experiments. These data concern the transmission of hereditary factors (genes) in particular species. The choice of such modeling of the data was driven by Morgan's adoption of a set of hypotheses concerning the physical mechanism of heredity. These hypotheses were themselves grounded in Morgan's defense of the chromosome theory of heredity, according to which chromosomes are the physical basis of genes. From the moment when this technique was designed, the construction and analysis of the genetic maps of *Drosophila melanogaster* became the object of a research program on which the geneticists' efforts concentrated for at least two decades. At the same time, debates arose in the geneticists' community about how to construct and interpret these maps.

In this paper, I focus on one of these debates, which took place in 1919-1920, between the geneticist William Castle, on the one hand, and Thomas Morgan's group (composed of Morgan and his three students Alfred Sturtevant, Calvin Bridges, and Hermann Muller), on the other hand. The debate centers on the form of the maps. Discussions focus on the way one should model the data in order to draw predictions concerning the transmission of genes. They do not concern the validity of the data themselves. Nor do they center on the hypotheses explaining these data. Castle's main arguments against Morgan's group's way of drawing maps appeal to criteria such as simplicity and predictive efficacy.

However, I will show that there is much more at stake here than simplicity and predictive efficacy. I will show that both parties' assessment of the simplicity and predictive efficacy of the maps is determined by strong theoretical commitments. Many hypotheses at different levels, concerning the physical explanation of the genetic data, underly the construction and interpretation of the maps. I will examine the map models proposed by Morgan's school, on the one hand, and by Castle, on the other hand, as well as their respective arguments in favor of their own models (Castle 1919a,b,c, Sturtevant *et al.* 1919, and Muller, 1920). Each of these scientists appears to have a different understanding of the different hypotheses at play, and of the explanatory link between these hypotheses and the

data.

My main claim is that preference for a particular way of modeling the data is driven by theoretical commitments. I aim at clarifying how the two (preference for a given map model, and theoretical commitments) articulate, in a context in which the correct theory about the underlying level is not known. Drawing from this particular case, I suggest that a detailed analysis of the epistemic attitudes of scientists towards hypotheses at different levels might shed light on the practices of data modeling, in particular in domains where no general theory is available (so-called “data-driven” domains).

In section 2, I present the theoretical context in which the debate arose. I introduce the various hypotheses at play, and I give a first statement of what is at issue in the debate. In section 3, I describe the two rival map models, respectively advocated by Morgan’s group and by Castle. Finally, in section 4, I analyze the debate and clarify the relations between the disputants’ disagreement on what seems to be a practical issue (how to model the data) and their theoretical differences.

## **2. THEORETICAL CONTEXT AND CRUX OF THE DEBATE**

The technique of linkage mapping embodies the junction of two domains, namely the Mendelian theory of heredity and cytology. Mendelian theory studies the transmission of hereditary factors — genes<sup>1</sup> — from generation to generation. Its method consists in performing breeding (hybridization) experiments on different strains of the same species, and in observing the distribution of characters among individuals of different generations, in order to draw information concerning the transmission of genes, which are supposed to be responsible for these characters. One of Mendelian theory’s typical tools is statistical analysis. Cytology, on the other hand, is the study of cellular processes. It relies on the observation of cells, and more specifically chromosomes, by means of microscopes.

It is worth emphasizing the distinction between what is usually called “classical genetics”, on the one hand, and Mendelian theory, as I have just characterized it, on the other hand. Classical genetics contains cytological hypotheses, whereas Mendelian theory

---

<sup>1</sup> The term “gene” was introduced by Willem Joannsen in 1909. But, until 1917, Morgan’s group would rather speak of “Mendelian factors”.

contains hypotheses only at the genetic level, namely at the level of analysis concerned with the transmission rules of genes, whatever the physical basis of genes may be. The birth of Mendelian theory so defined (which was called “genetics” by William Bateson in 1905)<sup>2</sup> dates back to the rediscovery of Mendel’s laws in 1900.<sup>3</sup> The birth of classical genetics<sup>4</sup> (as including cytological hypotheses) dates back to the early 1910’s, when Morgan adopted the chromosome theory of heredity, according to which the chromosomes are the physical basis of the genetic material. This is precisely the period in which linkage mapping was invented, as a result of the articulation of the genetic (Mendelian) and the cytological levels.

In the debate I am interested in, hypotheses at these different levels (genetic and cytological) are in play. Each of these hypotheses is the object of a more or less strong epistemic commitment by the different protagonists. In what follows, I briefly present this double disciplinary context. For each of the two domains, I specify what is accepted or rejected, and to what extent, by the two parties involved in the debate. For the sake of clarity, my outline of the theoretical landscape around 1910 is quite schematic.

## 2.1. Genetic level

### *The fundamental hypotheses of Mendelism*

The fundamental genetic hypothesis, which all Mendelian geneticists accept, is that the observable characters of individuals are caused<sup>5</sup> by hereditary factors, called “genes”.<sup>6</sup> Genes themselves are conceived of as unobservable entities, which are transmitted from generation to generation following statistical laws (Mendel’s laws). In multicellular

---

<sup>2</sup> See Dunn, 1965, 69.

<sup>3</sup> Note that Mendelian theory so defined does not correspond to Mendel’s theory, but rather to the theory developed by the geneticists in the 1900’s.

<sup>4</sup> In current usage, “classical genetics” and “Mendelian genetics” are synonymous expressions. However, I will restrict the use of the adjective “Mendelian” to refer to the study of the genetic phenomena, independently from their cytological basis.

<sup>5</sup> To be more precise, it is a *gene difference* that causes a *phenotypic difference* (see Waters 1994). When stating the theory of the gene in 1926, Morgan writes: “the characters of the individual are *referable* to paired element (genes)” (Morgan, 1926, 25, emphasis added).

<sup>6</sup> See footnote 1.

organisms with sexual reproduction, each individual possesses a pool of genes resulting from an equal participation of his/her parents. Therefore, each individual has an even number of genes (for each non-sex-linked character, he/she has two genes, also called “alleles”), one half of them coming from the male parent and the other half from the female parent.

Mendel’s first law (“segregation law”) states that, in the formation of germ cells (sperms and eggs), the two alleles of each gene “segregate” so that each germ cell contains only one copy (one allele) of each gene responsible for a character. Any germ cell has an equal chance of containing one or the other allele of each gene. Mendel’s second law<sup>7</sup> (of “independent assortment”) states that different pairs of genes assort independently from each other during the formation of germ cells. That means that the segregation of one pair of genes has no influence on the segregation of another pair.

The experimental method associated with this theoretical framework consists of breeding experiments, where the transmission of genes responsible for given characters is traced by observation of the distribution of these characters among individuals of different generations. From the consideration of the phenotypes of the individuals, one can deduce their genotype, by way of Mendel’s two laws.<sup>8</sup>

### *The phenomenon of partial linkage*

From 1905 on, geneticists observed a phenomenon, which seems to contradict Mendel’s second law. New data<sup>9</sup> showed that some genes tend to be inherited together, without being always so. For instance, genes of *Lathyrus odoratus* responsible for the color of the petals, on the one hand, and for the shape of the seeds, on the other, are partially linked (or “coupled”). They are not randomly redistributed, contrary to what Mendel’s second law would predict: partially linked genes are inherited together in more than 50%

---

<sup>7</sup> As Lindley Darden (1991, 139) notes, this law was not explicitly stated and distinguished from Mendel’s first law until exceptions were found (see below).

<sup>8</sup> Such deduction involves other hypotheses, such as the recessivity/dominance hypothesis, which I will not present here.

<sup>9</sup> The first case of partial linkage (or rather “coupling of traits”) is reported by Bateson, Saunders, and Punnett (1905).

of the cases. But their association is not systematic either (they are inherited together in less than 100% of the cases). Systematic association (or complete linkage) could be explained by the hypothesis that one and the same gene is responsible for these two characters.

A few years later, Morgan's work on *Drosophila melanogaster* (Morgan, 1910a) revealed the existence of sex-linkage. Sex-linkage means that some genes (for example the ones responsible for eye color and size of the wings) appear to be linked to what was assumed to be the genes responsible for sex determination.<sup>10</sup> Moreover, the various sex-linked characters appeared to be partially linked with each other. For instance, individuals that are identical to their mother regarding their eye color are also observed to be identical to her regarding the size of their wings in more than 50%, but less than 100% of the cases. The discovery of this relation between partial linkage and sex-linkage played a crucial role in Morgan's adoption of the chromosome theory of heredity.

*Genetic explanation of the phenomenon of partial linkage: the hypotheses of linkage groups and of crossing-over*

Various explanations of partial linkage were given.<sup>11</sup> I will only present Morgan's explanation, which finally prevailed. It consists of a double hypothesis. First, one assumes the existence of "linkage groups", that is, groups of genes that are linked together in some way, and therefore tend to be inherited together. Each individual would possess two copies of each linkage group.

Second, it is assumed that groups can break during the formation of germ cells, and that each group could exchange some genes with its homologous group. Such an exchange of genes is called *crossing-over*.<sup>12</sup> When two genes of the same linkage group (that

---

<sup>10</sup> In 1891 already, cytologists had identified a non-paired chromosome (a chromosome lacking its homologue), which Wilson called "X". But the hypothesis of the chromosome determination of sex was controversial until the 1910's.

<sup>11</sup> Bateson would explain it in terms of "coupling" (or "attraction") and "repulsion". Lindley Darden (1991, 122) notes that this terminology reflects Bateson's preference for physical analogies. He often refers to waves, vortices and charges (see e.g. Bateson, 1913, chap. 2 and 3). About Bateson's intellectual evolution, see (Coleman, 1970, Darden, 1977, Cock, 1983).

<sup>12</sup> This term was first proposed by Morgan and Cattell (1912).

is, genes which are inherited together in more than 50% of the cases) are not inherited together (are redistributed separately), one says that a *recombination* has occurred for these genes.

*Status of the hypotheses of linkage groups and crossing-over*

Some opponents of Morgan's theory, such as Bateson, would reject the hypothesis of crossing-over. But I will not consider them here. Suffice it to say that the double hypothesis of linkage groups and of crossing-over was accepted by the two parties to the debate I am interested in (Morgan's group and Castle). However, this double hypothesis is accepted only insofar as it is a purely *genetic* hypothesis. By "purely genetic", I mean that this hypothesis, as such, does not include any statement concerning the physical processes by which linkage is maintained and can be broken. Nor does it say anything about the nature of linkage groups, whose specification is just as neutral on the underlying physical mechanisms as the specification of genes is. Crossing-over, as accepted by Morgan and Castle, only means "gene exchange". The very term "crossing-over" certainly has a mechanical connotation, and this connotation is intended by Morgan. But Castle would reject it.<sup>13</sup> The main point of disagreement between Castle and Morgan's group concerns the physical mechanism of crossing-over.

## 2.2 Cytological level

Cytology was already a well-developed science in the late 19th century. Cytologists had quite a precise knowledge of the normal processes of cell division (mitosis). But the processes of the formation of germ cells (meiosis) were not precisely understood. Yet, geneticists conceived Mendelian segregation in terms of the formation of germ cells, and cellular processes were assumed to be the physical bases of hereditary (genetic) processes. Beyond that, though, hypotheses about the cytological bases of heredity were far from clearly understood. Nor was there any established hypothesis about what part of the cell is concerned with heredity. Therefore, the explanatory links between cytology and genetics

---

<sup>13</sup> In fact, Castle uses the term mostly to refer to Morgan's group's work. But, as we will see, he misunderstands it.



were still very shaky until the late 1920's. In the following, I will focus on one hypothesis about the cytological bases of heredity, which finally prevailed and is still at the core of today's genetics. Although it was challenged at the time of the debate, it was favored by both Morgan's group and Castle.

*The chromosome theory of heredity*

Chromosomes were identified by cytologists in the 1880's. In the early 1900's, the cytologists Montgomery and Sutton observed that pairs of similar chromosomes seem to join during the first phase of meiosis, and then disjoin just before the reductive division. Such observation obviously suggests an analogy between the laws of heredity and the behavior of chromosomes. This led some cytologists and geneticists, already in the early 1900's, to assume that chromosomes might be the physical basis of genes, and that their association and separation during meiosis might well explain Mendel's law of segregation (see in particular Sutton, 1902, Boveri 1904).

This hypothesis is called the "chromosome theory of heredity". Many geneticists, as Bateson did, rejected this theory until the 1920's. Although it might be difficult for us to understand how one could doubt it, there were many tenable considerations for such skepticism at the time. Indeed, no fine observation of the structure of chromosomes was available until the 1930's. So the hypothesis that chromosomes are the physical basis of genes could not be articulated precisely, and was controversial. It was not unreasonable to assume, like Bateson, that chromosomic phenomena were consequences, rather than causes, of deeper hereditary phenomena, in the same way as other observable characters of the living organisms.

Morgan himself, who became a fervent advocate of the chromosome theory of heredity from 1911, was still an opponent of it in 1910 (see Morgan 1910b). However, at the time of the debate with Castle, he was totally convinced that genes were located on chromosomes. So was Castle. Yet, because nothing was known about the fine structure of chromosomes, the modalities of the hypothetical "location" of genes "on" chromosomes were still the object of conjectures.

*The chiasmotypie hypothesis*

One cytological discovery played a determining role in Morgan's adoption of the chromosome theory of heredity.<sup>14</sup> In 1909, the cytologist Janssens had observed an intertwining of homologous chromosomes during meiosis (fig. 1). He conjectured that homologous chromosomes might exchange segments while intertwining. Note, however, that no such exchange was observed before the 1930's. Janssens called this putative physical exchange of segments of chromosomes "chiasmotypie". Janssens' observation and formulation of the chiasmotypie hypothesis (together with advances in the knowledge of sex-linkage) led Morgan to hypothesize that chromosomes might well be the physical basis of the linkage groups. Hence, the chiasmotypie hypothesis is, so to speak, the cytological side of the genetic hypothesis of crossing-over. However, one could admit the latter while rejecting the former.<sup>15</sup> Indeed, as I have highlighted, the hypothesis of crossing-over as such does not contain any statement about the physical processes underlying crossing-over. As far as I know, Castle did not express any judgement about the chiasmotypie hypothesis, but it is likely that he would have rejected it, at least in its quite simple mechanical representation, since it does not fit his representation of the structure of chromosomes.

*Hypotheses about the structure of the chromosomes. Status of the cytological hypotheses*

The chromosome theory of heredity, though controversial, is accepted by the two parties to the debate I shall present. However, Morgan's group and Castle disagree in their representation of the structure of chromosomes. Morgan's group describes chromosomes as threads on which genes are linearly ordered, like beads on a string (see fig. 2, which provides a mechanical model of crossing-over based on the chiasmotypie hypothesis). Castle, on the other hand, believes that chromosomes are complex chemical molecules, on which genes are arranged in a tridimensional complex fashion. He doubts whether genes *are* portions of chromosomes, and he rather conceives of genes as "attached" to the

---

<sup>14</sup> About Morgan's intellectual evolution, see (Carlson 1967, Allen, 1978).

<sup>15</sup> For instance, Richard Goldschmidt (1917) accepts both the chromosome theory of heredity *and* the hypothesis of crossing-over, while rejecting chiasmotypie (*i.e.* he refuses to correlate genes exchange with chromosomal segments exchange). He conceives of genes as attached to specific locations of chromosomes by biochemical forces, rather than as being themselves parts of chromosomes (see Allen, 1974, Dietrich, 2000). For a detailed analysis of Goldschmidt's model of crossing-over, see also Wimsatt, 1987.

chromosomes by “molecular”, rather than “mechanical” forces (Castle, 1919c, 501). He conceives of recombination as the result of a chemical reaction, rather than as the result of a mechanical process.

Beyond this difference, all agree, though, that the observational basis is insufficient to decide among alternative hypotheses about the structure of chromosomes. Because of the poor observational information one could draw from fixed preparations of chromosomes viewed from microscopes at that time, such hypotheses were doomed to be conjectural. The degree of uncertainty of any hypothesis about the structure of chromosomes was much higher than the degree of uncertainty of the chromosome theory itself, which was a reasonable and quite widely (though unconfirmed, and controversial) hypothesis. Hence, without agreeing on their representation of the structure of the chromosomes, the disputants do agree in their assessment of the poorly warranted epistemic status of any hypotheses concerning such structure that would be drawn from microscope observations. Therefore, for both Morgan’s group and Castle, such hypotheses could not, at the time, serve as premises in any decisive argument. Hypotheses about the structure of the chromosomes should rather be the conclusions of arguments based on other premises.

### **2.3. Relation between the genetic and the cytological levels. Crux of the debate.**

#### *Relation between the genetic and cytological levels*

When presenting the genetic hypotheses of linkage groups and crossing-over in section 2.1, I have insisted on their “purely” genetic status, by distinguishing them from hypotheses about the physical processes underlying linkage and crossing-over. Obviously, for people who believe in the chromosome theory of heredity, like Morgan and Castle, such explanatory hypotheses about the physical processes underlying genetic phenomena (linkage and crossing-over) are deeply related to their hypotheses concerning the way genes are arranged on chromosomes, and hence on their representation of the structure of chromosomes.

Yet, one should carefully distinguish between hypotheses concerning the physical mechanism of crossing-over that do not include cytological commitments from hypotheses

about the specific cytological basis of crossing-over. Hypotheses concerning the physical mechanism of crossing-over certainly exceed what I have called the “purely” genetic, or Mendelian level, since they go beyond statistical analysis of the results of breeding experiments, by proposing a mechanical explanation of these results. But such hypotheses about the physical mechanism of crossing-over do not imply, as such, any statement at the cytological level. Since hypotheses about the details of the cytological level are, as I emphasized, acknowledged as conjectural, the physical models explaining crossing-over are, in principle, conceptually independent from any hypothesis concerning chromosomes. For instance, one could perfectly accept Morgan’s mechanical model of crossing-over (fig. 2) while denying that the structure represented in the schema is a chromosome.

Conversely, the mere acceptance of the chromosome theory does not *ipso facto* commit one to a particular hypothesis concerning the physical basis of crossing-over. To be sure, Morgan’s formulation of the theory of crossing-over is historically inseparable from his adoption of the chromosome theory. But one could adopt the chromosome theory while rejecting such a mechanical model, as Castle did. On paper, Morgan’s group and Castle seem to agree on the most important theoretical points. They all accept the genetic hypothesis of crossing-over as an explanation of partial linkage, and the chromosome theory of heredity.<sup>16</sup> Their debate does not center on any of these hypotheses. However, their disagreement is grounded in their different conceptions of the explanatory link between the cytological and the genetic levels. Let me now schematically present the crux of this debate.

#### *Crux of the debate*

The official issue of the debate, which concerns the cytological basis of the genetic phenomena, is:

(Q1) What is the structure of chromosomes? How are genes arranged “on” chromosomes?

Another question, one that concerns the underlying mechanism of the genetic phenomena (be they chromosomic or not) is the following:

---

<sup>16</sup> This is not the case with Bateson, for instance, who rejects the chromosome theory of heredity but shares other aspects of Morgan’s views, which Castle would reject (such as the linearity hypothesis, which I present below).

(Q2) What is the mechanism of crossing-over? And hence: How should one model it in order to explain the genetic phenomena?

Finally, a third question is raised, on which the disputants in fact carry out the debate. It concerns the best way to present the data resulting from breeding experiments, in order to have a handy tool to predict the genetic phenomena (*i.e.* the inheritance patterns of different genes in given species), independently from the explanation of these phenomena:

(Q3) How should one present the statistical data concerning genes' recombination frequencies?

The disputants of the debate agree in considering that one's answer to Q1 could not, at the time, be based on cytological observations. To be sure, if more precise observations of the chromosomes were available, then such observations would offer a solid ground for answering Q1. An answer to Q1 so obtained would in turn shed light on the underlying mechanism of the genetic phenomena (Q2). But, by the time of the debate, the disputants do not feel entitled to accept any particular hypothesis concerning Q2 on the basis of their preferred hypothesis concerning Q1 — insofar as it is based on insufficient observational support, which cannot play any decisive evidential or warranting role. As a consequence, the disputants claim that their own answer to Q2 is justified by the mere analysis of the genetic data.

Let me clarify why it is important to distinguish between Q1 and Q2. One's answers to these two questions (once one accepts the chromosome theory of heredity) certainly do have to be strongly related. If any solid answer to Q1 were available (on an observational basis), then it would definitely have epistemic force, and should drive one's answer to Q2. But, because of the acknowledged uncertainty of any hypothesis concerning Q1, the answer to Q2 has to be, as the disputants claim, driven by the mere analysis of the data. As a consequence, the two parties tend to carry out their dispute about Q3, as if it was merely a question of simplicity and predictive efficacy. From their alleged theory-neutral answer to Q3, they then claim to infer what seems to be the most probable answer to Q2 (by something like inference to the best explanation), and hence to Q1.

In the following, I will show that the disputants' answer to Q3 is determined by their favored hypothesis concerning Q2. Because they do not feel entitled to rely on their beliefs concerning Q1 (based on cytological observations), I suggest that they fail to face

up their deep theoretical differences on Q2 (and Q1), and center the discussions on Q3. An open question, which will be raised but not answered, is: To what extent their answer to Q2 is influenced by, and influences their beliefs concerning Q1?<sup>17</sup>

### 3. MORGAN'S MECHANICAL MODEL OF *CROSSING-OVER* AND THE TWO RIVAL TECHNIQUES OF LINKAGE MAPPING

In section 2.1, I have presented the hypothesis of crossing-over. Insofar as it does not imply any commitment to a hypothesis concerning the physical mechanism underlying genes' exchange, it is accepted by both Morgan and Castle. However, the theory of crossing-over, as it is articulated by Morgan in 1911, says something more than the hypotheses presented above (Morgan 1911a). Beyond assuming that linkage groups can exchange genes, it provides a *mechanical model* for such an exchange. This model underlies the invention of the technique of linkage mapping, on which the debate with Castle bears. After having introduced Morgan's theory in section 3.1, I present the mapping technique, as it was designed by Sturtevant in 1913 (section 3.2). In section 3.3, I introduce Castle's rival method for constructing maps.

#### 3.1 Morgan's theory

Morgan's theory of crossing-over is embodied by the mechanical model schematized in figure 2. This theory can be reconstructed in the form of two premises and a conclusion.

*First premise: linearity.*

The first premise is the hypothesis according to which genes are ordered in a linear fashion along their linkage groups, like beads on a string. Note that this string might be the chromosome, as Morgan supposes, but that it might also not. Nothing is said, even in

---

<sup>17</sup> Indeed, one's choice of a particular answer to Q1 and Q2 might be due to one's preference for, say, mechanical models. In this case such preference might itself drive one's interpretation of the cytological observational data. On the other hand, one could argue that the mechanical model of crossing-over is itself strongly suggested by cytological images such as the one displayed in figure 1 — and that such images do warrant, to some extent, one's acceptance of the mechanical model.

Morgan's statement of the "theory of the gene" (Morgan, 1926, 25), about the physical nature of linkage groups. Nevertheless, Morgan's theory says something more than the hypothesis of linkage groups, according to which some genes are linked together in some way (see 2.1). Morgan's theory says that linkage groups are line-shaped (or thread-like).

The hypothesis of linearity was suggested by a phenomenon evinced by breeding experiments on *Drosophila*: geneticists could observe that the frequencies of recombination were additive. What does that mean? Let us consider three genes  $A$ ,  $B$ , and  $C$ , belonging to the same linkage group. Let us call  $R(AC)$  the frequency of recombination of  $A$  and  $C$ , namely the frequency of their being inherited separately (which has to be less than 50%, since they are linked). Saying that frequencies of recombination are additive means that  $R(AC)=R(AB)+R(BC)$ . This equation obviously suggests that genes are arranged in a linear fashion. This hypothesis of linearity is the one Castle rejects.

#### *Crossing-over.*

The second premise of Morgan's theory is the hypothesis of crossing-over (see 2.1). It states that there is an ordered exchange between portions of homologous linkage groups during the formation of germ cells. The consequence of this exchange is the recombination of the genes lying on these portions.

#### ***Conclusion: proportionality of the recombination frequency and of distance***

These two hypotheses together lead to the conclusion that the percentage of crossing-over between two genes (their recombination frequency) is proportional to the "distance" between these genes on the linkage group. Indeed, as appears clearly in the model of crossing-over schematically presented in figure 2, the more distant from each other two genes are, the more a break between them is likely to occur<sup>18</sup>, and hence the more they are likely to be redistributed separately.

### **3.2 Sturtevant's linkage mapping technique**

---

<sup>18</sup> Assuming that different points of the line are equally likely to break (but see footnote 17).

The technique devised in 1913 by Sturtevant consists in transforming Morgan's theory into a mapping scheme for the linkage groups. Since the recombination frequency between two genes of a given group is assumed to be a function of the distance between them, Sturtevant proposed that this frequency could be used as an index of the distance.

*Sturtevant's (1913) map. Exceptions to additivity and the hypothesis of double crossing-over.*

On the basis of the frequencies calculated from the results of breeding experiments, Sturtevant (1913) constructed the map for one linkage group of *Drosophila* (the group of the so-called "sex-linked" characters, corresponding to chromosome X). However, for high recombination frequencies (represented by long distances), some experiments (Morgan 1911b, Morgan and Cattell 1912) showed exceptions to additivity: for two genes  $A$  and  $C$  with high recombination frequency  $R(AC)$ , one finds  $R(AC) < R(AB) + R(BC)$ . Instead of rejecting the hypothesis of linearity (which, remember, was suggested by the observation of additivity), Sturtevant hypothesized that there could be more than one crossing-over occurring on the same linkage group. As is shown in figure 3, double crossing-overs would cancel the recombination of the corresponding genes. The genes located at the extremities of the linkage groups, and separated by two breaks, are, in the end, inherited together.

Drawing from this hypothesis, Sturtevant chose to construct his map by relying on the *short distances*. Long distances on the map therefore correspond to the sum of short distances, rather than to the observed recombination frequencies between the most distanced genes. This is a crucial point, which appears clearly when one considers the table of the recombination frequencies (fig. 4) and its corresponding map (fig. 5). The table displays the proportions of crossing-over for each pair of genes (*i.e.* their recombination frequency), and the corresponding percentage, which is supposed to give the distance between them. These genotypic data are themselves inferred from the phenotypical data concerning the distribution of the observable characters among individuals. Consider  $BM$ : the table says that, out of 693 cases,  $B$  and  $M$  were inherited separately 260 times, that is, 37.6% of the cases. However, on the map, the distance between  $B$  and  $M$  is 57.6 (and not 37.6). This result derives from the fact that this "distance" was calculated by adding up short distances, rather than by relying on the recombination frequencies that could be



inferred from the phenotypic data.

*The function of maps. Ambiguity of the term “crossing-over”.*

Before turning to Castle’s model, let me make a few remarks about Sturtevant’s technique. One can interpret the goal of his enterprise in at least two ways. First, maps could be thought of as (highly schematic) representations of the chromosomes. Sturtevant’s technique was indeed inspired by Morgan’s suggestion that recombination frequencies might be an index of the distance between genes on the chromosomes. In fact, Sturtevant presents his map as a “diagram of the chromosome” (1913, 6). However, even if the chromosome theory was false<sup>19</sup>, mapping would still have a value at the genetic level. In other words, even if it had turned out that “distance”, defined as a function of the proportion of crossing-overs, was used metaphorically, and did not represent any physical distance<sup>20</sup>, maps would still be useful as predictive tools at the genetic level:

It would seem, if this hypothesis be correct, that the proportion of “crossovers” could be used as an index of the distance between any two factors. Then, by determining the distances (in the above sense) between A and B and between B and C, one should be able to predict AC.  
(Sturtevant, 1913, 3)

The meaning of this statement is not straightforward. Consider, again, the map in figure 5. Since it is constructed by relying on low recombination frequencies, distances on it do not stand for actual recombination frequencies (which appear in the table in figure 4). They rather stand for the putative percentages of crossing-overs *at the physical level of the linkage group*. These “physical” (or “real”) crossing-overs may or may not have consequences on the resulting genotype of the individuals (and, hence, on their phenotype). Indeed, two crossing-overs result in some failures of gene recombination. Hence, in virtue of the

---

<sup>19</sup> Or if, as Sturtevant hypothesizes (and it turned out that he was right), chromosomes were not equally likely to break on every point, which would imply that distance is a measure of strength combined with length.

<sup>20</sup> In a work in progress, I distinguish between what I call “schematic drawings” and “diagrams”. “Schematic drawings” are spatial representations of spatial relations (*e.g.* the spatial structure of a cell). A diagram is a spatial representation of non-spatial relations (*e.g.* hierarchical, temporal, causal relations, etc.). For instance, the synthesis of proteins is usually represented by a diagram. In this terminology, linkage maps can alternatively be considered as schematic drawings of chromosomes (which conserve the topological relations, if not the metric ones), or as diagrammatic representations of the relative recombination values.

possibility of double crossing-overs, the percentage of “physical” (or “real”) crossing-overs differs from the percentage of “observed”<sup>21</sup> crossing-overs (recombinations). Therefore, the “distances” that one can predict by using Sturtevant’s scheme do not correspond to the recombination frequencies (as Sturtevant’s quote above suggests), but rather to the putative percentage of real crossing-overs at the physical level.

As is clear from the preceding analysis, the term “crossing-over” is ambiguous. For the sake of clarity, one could (and should) adopt the following convention: speaking of recombination when two genes of the same group are redistributed separately (“observed” crossing-over), and keep the term “crossing-over”, which has a strong mechanical connotation, to refer to the putative underlying exchanges of genes at the physical level (which, when they happen an even number of times, count as no recombination). However, the protagonists of the debate do not adopt any such terminological convention, and this ambiguity embodies one of the central issues of the debate.

### 3.3 Castle’s model

During the years following the invention of the mapping technique by Sturtevant, Morgan and his students, in collaboration with the cytologist Edmund Wilson, progressively construct the genetic maps for the four chromosomes of *Drosophila*. Through striking predictions at the genetic level, they adduce more and more evidence in favor of the chromosome hypothesis. Consensus on the Morgan group’s theory was not, however, reached until the early 1920’s. Until then, it was the target of many criticisms, in particular by Bateson (1916) and Goldschmidt (1917). Castle’s (1919a,b,c) criticism goes beyond attacking theoretical hypotheses advanced by Morgan’s group. It consists in proposing a rival mapping scheme, resulting in a completely different way of modeling the data.

Castle (1919a) does not question the validity of the data about genes’ recombination frequencies that Morgan’s group draws from the phenotypic data obtained through breeding experiments. In fact, he constructs his map by relying on the data used by Morgan and Bridges (1916).

---

<sup>21</sup> “Observed”, here, means “inferred from the phenotypical data”, which are themselves the result of a statistical processing of the data concerning the observable characters of the individuals.

The main difference between Castle's and Sturtevant's schemes lies in their strategy in handling the problem of non-additivity for high recombination frequencies. Unlike Sturtevant, who invokes double crossing-overs in order to keep the hypothesis of the linear arrangement of genes, Castle merely drops this hypothesis. Non-additivity, he argues, is a problem only if one assumes linearity. Dropping linearity enables one to conceive of distance as being always proportional to recombination frequency, be it high or low. Castle's scheme thus relies on the definition of distance as a linear<sup>22</sup> function of recombination frequency. Since high frequencies are not additive, the resulting map has to be multidimensional (see fig. 6). Indeed, if  $R(AC) < R(AB) + R(BC)$ , then  $A$ ,  $B$ , and  $C$  cannot lie on a unique line. Figure 7 shows Morgan and Bridges' (1916) unidimensional map for the same data, which was constructed according to Sturtevant's scheme.

Castle's map is undeniably a linkage map: it consists in converting recombination frequencies into distances on the map, as suggested by Sturtevant. Castle even makes the conversion between frequencies and map distances in a purer way. His scheme merely relies on the definition of distance as a linear function of recombination frequencies. If one focuses on the genetic level, without considering the conclusions one could conjecturally draw from it about the arrangement of genes on chromosomes, Castle's map model might seem to be merely an alternative way of presenting the same data, without any theoretical issue being at stake. However, his mapping scheme, even at the genetic level (leaving aside the cytological level), involves a drastically different conception of crossing-over and of the very meaning of the maps.

#### 4. ANALYSIS OF THE DEBATE

In order to advocate his model, Castle proposes a series of arguments against linearity, which Morgan and his students (Sturtevant *et al.*, 1919, Muller, 1920) answer systematically. William Wimsatt's (1987) diagnosis is that this debate exemplifies the opposition between a "mechanistic or realist research program" (Morgan's) and "an operationalist or instrumentalist one" (Castle's). For Wimsatt, Castle "avoid[s]

---

<sup>22</sup> One has to be cautious with the term "linear": Castle defines distance as a *linear* function of recombination frequencies, which leads him to reject the hypothesis of the *linear* (unidimensional) arrangement of genes along their linkage groups, and hence to construct a multidimensional map.

commitments to theoretical models” and “stay[s] as close to the data as possible” (Wimsatt, 1987, 39). This is true to quite a large extent, as I will show. But my analysis aims to shed a slightly different light on the debate, and to qualify Wimsatt’s judgement that Castle is a pure instrumentalist. Quite interestingly, Castle and Morgan’s group officially agree both in the hierarchy they establish among the types of arguments they invoke, and in the criteria they appeal to in order to justify their preference for a model. I will show that Castle’s choice of a multidimensional model does not merely stem from a theory-neutral attempt to present the data in a simple and efficient way. Rather, from both sides, strong theoretical commitments underlie arguments appealing to simplicity and to predictive efficacy.

#### 4.1 Different types of arguments

##### *Cytological arguments: the putative structure of chromosomes*

Castle’s first set of arguments against linearity are at the cytological level. They question the plausibility of the hypothesis that genes are linearly ordered on chromosomes (Q1). Castle strongly doubts that “elaborate organic molecule[s]”, as he assumes chromosomes to be, would have “a simple string-like form” (Castle, 1919a, 26), as suggested in Morgan’s unidimensional maps.<sup>23</sup> Sturtevant *et al.* (1919) simply reject the hypothesis that chromosomes are organic molecules.

Yet, as I already highlighted, such considerations lack observational support. They can, at most, lead one to lean towards a particular map model. Castle’s conjectures about the complex structure of chromosomes merely aim to show that the cytological level itself does not offer any evidence in favor of a linear (unidimensional) map model. Despite the title of his paper indicating that Q1 *is* the central question (“Is the arrangement of genes in the chromosome linear?”), Castle then focuses on the question of the form of the maps (Q3).

Moreover, Castle is careful not to claim that his map really represents the actual

---

<sup>23</sup> Castle acknowledges that microscope images of chromosomes *suggest* such a thread-like form. But he emphasizes that it is not clear “how much of the visible chromosomes is composed of genes and how much is something else.” Nor is there “evidence whatever that the chromosomes are genes, but only that the genes are in some way connected with them” (Castle, 1919c, 501). Hence, according to him, one is not entitled to draw conclusions about the structure of the chromosomes on the basis of the microscope images that were available at the time.

structure of chromosomes. What can be drawn from his other arguments (presented below) is only that he thinks that Morgan's unidimensional maps do *not* represent the actual structure of chromosomes. In other words, to the question stated in the title of his paper, Castle's official claim is that there exist good reasons to answer negatively, but nothing more. As Wimsatt (1987) suggests, Castle presents his map as a tool enabling one to predict further recombination frequencies (and thus to study the transmission rules of given genes in given species) on the basis of the available recombination data. He does not claim that his map provides an explanation of these data, by representing the mechanism underlying them.

Yet, the situation is not so simple. After having given two more arguments against linearity (presented below), Castle comes back to cytological considerations at the end of his paper. He suggests that the great predictive efficacy of his model is a warrant for the belief that the model represents in some way the real processes underlying the genetic phenomena. However, he claims, these processes might not be mechanical, hence spatial relations on the map might well represent non-spatial relations, such as molecular forces.<sup>24</sup> Distance on the map might not represent physical distance.

Before turning to the other arguments, let me consider Castle's argumentative strategy. Having advocated his own map model (*i.e.* having answered Q3), Castle infers an answer to Q2. His argument consists in saying that the predictive efficacy of a model provides one with good reasons to hold that it represents *in some way* the real underlying processes. But having such a (tentative) answer to Q2 is not enough, according to Castle, to infer an answer to Q1, because spatial relations on the map might well stand for non-spatial relations. *Pace* Wimsatt, I suggest that, rather than avoiding commitment to *any kind* of theoretical model, both Castle's claim that the map represents real relations *and* his reluctance to consider that these relations are spatial reveal his preference for a certain kind of explanation. He believes that the cytological level should provide a chemical, rather than

---

<sup>24</sup> "What, it might be asked, does this reconstruction signify? Does it show the actual shape of the chromosome, or at any rate of that part of it in which the observed genetic variations lie? Or is it only a symbolical representation of molecular forces? These questions we can not at present answer. A first step toward answering them will be the construction of a model which will give us a reliable information as to undetermined genetic relationships. A model which will answer questions truthfully must be a truthful presentation of actual relationships even though we do not know whether they are spatial or dynamic" (Castle, 1919a, 30).

a mechanical explanation of the genetic phenomena.

*Arguments appealing to the simplicity of the model*

According to Castle, the linearity hypothesis is costly to maintain, since it requires the assumption of double crossing-overs. The hypothesis of double-crossing over, he argues, it is an *ad hoc* hypothesis, whose introduction would be acceptable only if linearity were indispensable.<sup>25</sup> But, according to him, there is no good evidence for linearity and some evidence against it. Dropping both hypotheses altogether leads to what he considers to be a simpler map model. It is simpler in the sense that it relies only on the definition of distance as a linear function of recombination frequency, without any additional hypothesis being needed.

On the other hand, Sturtevant *et al.* (1919) argue that assuming the occurrence of double (and more generally multiple) crossing-overs is not *ad hoc*. Rather, once one admits the possibility of one break along the linkage group, the possibility of multiple breaks is a null hypothesis (one would need justification, instead, to restrict breaks to occurring only once). Hence, Morgan's group argues, double crossing-over, far from being a costly addition, leads to the construction of much simpler and more readable maps. They hold that the unidimensional form of their maps has the advantage of simplicity over what they call the "rat trap" form of Castle's maps.

Note that both Castle and Morgan's group appeal to what Kuhn (1977) would have called the same "value", namely simplicity. However, they obviously make different simplicity judgements. I suggest that their respective assessment of the simplicity of map models is a consequence, rather than a cause, of the choice of their respective models. Hence, this choice has to be guided by deeper reasons, which I will try to clarify in section 4.2.

*Arguments concerning the predictive efficacy of the model*

The third set of arguments concerns the predictive efficacy and accuracy of maps

---

<sup>25</sup> "The experimental data show that double crossing-over *must* occur, if the arrangement of the genes is linear. [...] But if the arrangement is *not* linear, double-crossing-over need not be assumed as an explanation of the observed regroupings" (Castle, 1919a, 30).

at the genetic level. The issue of predictive efficacy and accuracy is explicitly presented as the most decisive by the two parties. Maps have primarily to fit the data, and to serve as predictive tools. Each party claims the superiority of its own map model as a computational tool facilitating accurate predictions, independently of any underlying physical explanation.

Castle's argument.

According to Castle, the hypothesis of double crossing-over is costly; and nothing justifies paying the price. Quite the contrary, maintaining the hypothesis of linearity leads to what Castle judges as incoherent predictions. According to him, Morgan and Bridges' map does not fit the data. Their map displays distances of more than 50 cartographic units, one distance unit corresponding to 1% of crossing-over. But, Castle argues, 50 units should be the maximum, corresponding to the cases of independent assortment or absence of linkage (namely cases involving genes belonging to different linkage groups, therefore obeying Mendel's second law). He claims that values greater than 50 "have not been observed and are logically impossible" (Castle, 1919c, 501).

*A cross-over value greater than fifty cannot exist. For there must be either linkage or no-linkage. But no-linkage means 50% cross-overs, and linkage means less than 50% cross-overs. (Castle, 1919a, 28)*

This argument relies on a misuse of the notions of distance and of crossing-over. Castle mixes up observed crossing-overs (recombinations), whose percentage undeniably cannot exceed 50, with real crossing-overs, that is, real breaks and exchanges at the physical level. But, in Morgan's maps, distance does not correspond to the percentage of recombinations, but rather to the probability of real crossing-overs. Since there can be more than one such crossing-over at the level of the physical group, it makes perfect sense to speak of distances of more than 50 (50% of chance is the maximum for *one* crossing-over to occur).

On the other hand, since Castle rejects the hypothesis of double crossing-over, he has no reason to make such a distinction between real and observed crossing-over. This leads him to (mis)use the concept of crossing-over to refer to cases of independent assortment, when two genes of different linkage groups are redistributed separately (which,

in virtue of Mendel's laws, always has a 50% chance of occurring). In other words, for Castle, "crossing-over" entails absence of linkage (separate inheritance), whether or not the genes in question belong to the same linkage group.

If A and B assort wholly independently, without any linkage whatever, just as they would in ordinary Mendelian inheritance where no linkage exists, cross-overs and non-cross-overs will be equal, 50% each. (Castle, 1919a, 28).

But, in Morgan's conceptual framework, speaking of crossing-over (or absence thereof) between genes lying on different linkage groups makes no sense: a crossing-over corresponds to a break within a linkage group. I shall come back to Castle's misunderstanding in section 4.2. Beforehand, I wish to say a word about some aspects of Muller's answer.

Muller's answer.

After having emphasized Castle's misunderstanding of the notion of crossing-over<sup>26</sup>, Muller (1920) argues that the Morgan group's map model fits the data much better than Castle's model does.<sup>27</sup> An important argument in favor of the Morgan group's model is that it predicts the deviation from additivity, by accounting for it as a function of map distance, too. Indeed, in their scheme, the probability of double crossing-over is expected to increase with map distance. By contrast, Muller argues, Castle's model would predict the opposite: in Castle's scheme, three arbitrary factors are more likely to fall on a straight line the further they are separated. Hence, Muller shows that the two models are not empirically equivalent, and that the unidimensional model is empirically more adequate than Castle's tridimensional model, and yields better predictions.

Note that Muller claims that the predictive superiority of Morgan's model can be

---

<sup>26</sup> "[I]t has never been claimed, in the theory of linear linkage, that the per cents of crossing-over are actually proportional to the map distances: what has been stated is that the per cents of crossing-over are *calculable* from the map distances" (Muller 1920, 98).

<sup>27</sup> Muller's (1920) reply to Castle consists of many strong arguments, which I will not present here. In addition to showing that Castle's model does not fit the available data, he also provides a thorough criticism of Castle's use of the data. In particular, he shows that Castle should not treat the data in the same way for map construction and for testing the hypothesis of linear arrangement.



shown without any consideration of the underlying physical processes. His argument is that the mathematical apparatus enabling one to accurately predict the recombination frequencies relies on a representation of linkage groups, in which each gene is bound with only two others in a linear fashion. According to him, the very analysis of the recombination data justifies a unidimensional representation of the arrangement of genes.

Now, from this argument concerning Q3, which could be called, in Wimsatt's terminology, an "operationalist" or "instrumentalist" argument, he draws that the mathematical relations in the model do stand for real relations. However, in a somewhat ironical way, he pretends to concede to Castle that the "line" formed by the genes in the model must not correspond to a physical line:

Whether or not we regard the factors as lying in an actual material thread, it must on the basis of these findings be admitted that the forces holding them together — be they physical, "dynamic" or transcendental — are of such nature that each factor is directly bound, in segregation, with only two others — in bipolar fashion — so that the whole group, dynamically considered, is a chain. This does not necessarily mean that the spatial relations of the factors accord with these dynamic relations, for it is conceivable *a priori* that factor A might be far off from B, in another part of the cell, and that they might nevertheless attract each other, during the segregation division, by some sort of chemical or physical influence. [...] no implication as to the physical arrangement of the genes is intended when the terms "linear series", "distance", etc., are used; these will refer only to the relations existing between the points in the linear map, which may be regarded merely as a mathematical mode of representation of the data themselves. (Muller 1920, 100-101)

He nevertheless adds that, in absence of any strong reason to doubt it, one would be "fanciful" in denying that the genes are *really* ordered on a "physical line":

when the various conditions which have to be fulfilled at segregation are taken into consideration, any other explanation for these peculiarly linear linkage findings than an arrangement of the genes in the spatial, physical line proves to be hazardously fanciful. (Muller, 1920, p. 101)

Muller's argument is that the mathematical model enabling one to predict the data seems in

such a striking way to describe a spatial mechanism that one has no good reason to reject its spatial interpretation. Muller thus presents his realist commitment toward the mechanical model of crossing-over, according to which genes are arranged spatially on a material line, as a consequence of his searching for the best account of the genetic data.<sup>28</sup>

Muller's considerations can be described as a rigorous attempt to articulate the inductive patterns of an inference to the best explanation. Although his position might be considered as better warranted than Castle's one — and although it turned out that Morgan's group was right —, it was still necessary, by the time of the debate, to look for more evidence in order to be fully satisfied with Muller's spatial-mechanical explanation. In the next section, I shall try to clarify why a skilled scientist like Castle could still reject this explanation, without being merely stubborn or unprofessional. I will show that neither simplicity judgements nor considerations of predictive efficacy could settle the debate to either side's satisfaction, because such judgements and considerations were too theory-dependent to be persuasive. Indeed, the disputants' commitments to a physical explanation of heredity and their search for a predictive tool (at the genetic level) are more intricate than what they suggest.

#### 4.2 Maps as predictive tools, and theoretical commitments

Castle's misunderstanding of the concept of crossing-over is quite surprising. One cause of his misconception might be the ambiguous use of the term "crossing-over" by Morgan's group itself. As we have seen, the distinction between the two meanings of crossing-over ("real" crossing-over and genetic recombination) relies on the hypotheses of linearity and double crossing-over, which together embody a mechanical explanation of partial linkage. This distinction is never made explicit by Morgan and his students, who use the term "crossing-over" in both senses. One reason for this lack of terminological precision might be that the distinction is pretty clear for anyone who has a mechanical representation of crossing-over. The term "crossing-over" does strongly suggest this

---

<sup>28</sup> Interestingly, in his 1913 paper (thus in a different argumentative context), Sturtevant presents the fact that the best mathematical model of recombination frequencies is a linear one as a piece of evidence in favor of the chromosome theory of heredity, assuming that chromosomes are thread-like: "[These results] form a new argument in favor of the chromosome view of inheritance, since they strongly indicate that the factors investigated are arranged in a linear series, *at least mathematically*" (p. 16, emphasis added).

mechanical representation. But then, it is quite surprising that Morgan's group keeps using this term to refer to gene recombination, once the hypothesis of double crossing-over has been made, and recombination has been distinguished from physical crossing-over.

Still, highlighting Morgan's group's ambiguous use of the term "crossing-over" is certainly not sufficient to explain Castle's misunderstanding. His mistake should rather be considered a symptom of his commitment to another theoretical model, which prevents him from making the adjustments he should have made in understanding the system of concepts underlying Morgan's scheme. Indeed, the very meaning of Morgan's maps, as well as the rules one has to follow in order to draw predictions from them, are underlaid by the mechanical model. The recombination data displayed in the table (fig. 4) are not directly retrievable from the map, since mapping distance does not always correspond to recombination frequencies. On the other hand, Castle claims that "from [his] model one may by *direct measurement* ascertain what other undetermined linkage values are likely to be" (Castle, 1919b, 32, emphasis added).<sup>29</sup> In order to retrieve data concerning the recombination frequencies and to draw predictions concerning such frequencies from Morgan's maps, one needs to assume that distances on the map *have a spatial meaning*.

To be sure, one could hold Castle's views and still *understand* the maps. But what I wish to highlight here is that hypotheses that cannot be drawn from the mere analysis of the data underlie the Morgan group's use of the concept of spatial distance and the mapping scheme designed by Sturtevant. The adoption of Morgan's model is justified only if one considers that maps represent something more than mere recombination frequencies of genes, namely the mechanical explanation of these recombinations. Far from being mere graphical presentations of the statistical data contained in the corresponding table, these maps are theoretical representations, whose format embodies strong hypotheses. The mechanical model of crossing-over underlies the extraordinary battery of concepts and mathematical tools developed by Morgan's group in order to

---

<sup>29</sup> From that perspective, it is true that Morgan's scheme is more theory-dependent than Castle's (as suggested by Wimsatt's calling Castle an instrumentalist). Nevertheless, Castle's misunderstanding forces qualification on Wimsatt's claim that Castle "avoid[s] commitments to theoretical models". Although one has to be cautious with such psychological considerations, it seems hard to understand how Castle could make such a mistake without having some deep prior commitment *against* Morgan's mechanical model.

account for what *prima facie* appeared as recalcitrant data.<sup>30</sup> As a consequence, if one focuses on the recombination data, and if one considers maps as merely *predictive* tools — and not as *explanatory* models —, as both Muller and Castle claim to be doing in their respective arguments, then the debate between disputants holding different theories cannot be settled. Indeed, considerations of predictive efficacy and of simplicity are themselves theory-dependent.

Let me briefly consider Morgan's maps as merely predictive tools. For someone who does not first accept the mechanical model represented in figure 2, Morgan's maps are far from being handy predictive tools. If one does not first admit the mechanical explanation of linkage, and the whole battery of concepts that comes with it (distance, double crossing-over, and the distinction between recombination and real crossing-over), the Morgan group's mapping scheme seems unjustifiably complex. Maps do not only *suggest* a mechanical explanation of the genetic phenomena; they are practically unusable — or at least unnecessarily complex — for someone who would not first admit such explanation. What appears to be simple here thus depends on theoretical commitments.

My purpose is not to defend Castle's position in an anti-whiggish spirit. Morgan's group already had, in 1919, many good reasons to consider that the best explanation of the genetic data was to be found in a spatial-mechanical model. The fact that Morgan's theory was not fully justified — because it still lacked evidential support<sup>31</sup> — does not mean that their position was not better warranted than Castle's, already in 1919 (independently from the fact that it turned out to be true). The rules for inference to the best explanation are difficult to state, and such epistemological issue is far beyond the scope of this paper. Let me just incidentally note that cases in the history of science such as the one I have presented here provide one with good examples of the complexity of epistemic warrants in science, which are worth studying in detail.

Wimsatt holds that the debate between Morgan's group and Castle “provides a

---

<sup>30</sup> An example is Muller's (1916) mathematical work on the notion of coincidence, aimed at accounting for the phenomenon of interference (see Wimsatt 1992).

<sup>31</sup> Such evidence was available a few years later, when striking correspondence of genetic phenomena with cytological observations was established in the 1930's. See, in particular, the works of Creighton and McClintock (1931) and of Painter (1934). Note that Castle abandoned his reluctance to accept Morgan's model already in the early 1920's.

clear example of the superiority of a mechanistic or realist research program over an operationalist or instrumentalist one” (1987, 39). It is true that Morgan’s mapping scheme is more theory-dependent than Castle’s one, whose position can be described as stemming from a tendency to stay “as close to the data as possible”. Nevertheless, I suggest that the difference between the two parties does not so much lie in a difference in their epistemic attitudes towards their own models. Rather, the two parties (superficially) agree in giving priority to the predictive efficacy of maps at the genetic level. But Morgan’s group’s commitment toward the mechanical model of crossing-over, on the one hand, and Castle’s representation of chromosomes as complex organic molecules, on the other hand, did drive their preference for a particular way of modeling the data. In the absence of more evidence in favor of one or the other underlying explanatory model, considerations of predictive efficacy and simplicity could not settle the debate. I finally suggest that this case might rather be taken as an example of the epistemic superiority of a position over a less articulated and warranted one, in the absence of full justification of any of them.

## 5. CONCLUSION

My analysis of the debate between Castle and Morgan’s group has shown that their respective choices of a particular way of modeling the data were indissociable from their theoretical commitments to a particular model explaining these data (mechanical, chemical). The meaning of their maps, and their use as predictive tools, derive from these theoretical positions. Maps are more than “models of data” containing no theoretical hypothesis; their very format embodies a theoretical model of the mechanism of heredity. Given the terms of the debate, which tended to center more on predictive accuracy and simplicity than on the deep theoretical differences in modes of explanation, it was unlikely that either side was going to persuade the other, in the absence of more evidence. Indeed, both sides could claim predictive adequacy if one just focuses on the recombination data, and what appeared to be simple depended on background theoretical positions. Drawing from this case, I suggest that data modeling might always be, to some extent, theory-dependent, and that judgements of computational tractability might never be totally theory-neutral.

## ACKNOWLEDGMENTS

I wish to thank Anouk Barberousse and Jean Gayon, for having supervised the PhD research from which this paper was drawn, Chris Pincock, as well as two anonymous referees for useful criticisms and suggestions, and above all Tyler Burge for thorough and insightful comments on successive drafts of this paper.

## REFERENCES

- ALLEN, G. (1974). Opposition to the Mendelian-chromosome theory: The physiological and developmental genetics of Richard Goldschmidt. *Journal of the History of Biology*, 7:49-92.
- ALLEN, G. (1978). *Thomas Hunt Morgan*. Princeton: Princeton University Press.
- BATESON, William, E.R. SAUNDERS and R.C. PUNNETT (1905). Further Experiments on Inheritance in Sweet Peas and Stocks: Preliminary Account. *Proceedings of the Royal Society*, B, LXXVII.
- BATESON, W. (1913). *Problems of Genetics*. New Haven: Yale University Press.
- BOVERI, Theodor (1904). *Ergebnisse über die Konstitution der chromatischen Substanz des Zellkerns*. Jena: G. Fischer.
- CARLSON, E. A. (1967). *The Gene: A Critical History*. Philadelphia: W.B. Saunders, reprinted by the University of Iowa Press.
- COCK, A. (1983). William Bateson's rejection and eventual acceptance of chromosome theory. *Annals of Science*, 40:19-60.
- COLEMAN, W. (1970). Bateson and chromosomes: Conservative thoughts in science. *Centaurus*, 15:228-314.
- CREIGHTON, H. B. and MCCLINTOCK, B. (1931). A correlation of cytological and genetical crossing-over in *Zea Mays*. *Proceedings of the National Academy of Sciences of the USA*, 17:492-497.
- CASTLE, W.E. (1919a). Is the arrangement of the genes in the chromosome linear? *Proceedings of the National Academy of Sciences of the USA*, 5(2):25-32.
- CASTLE, W.E. (1919b). The linkage system of eight sex-linked characters of *Drosophila virilis* (data of Metz). *Proceedings of the National Academy of Sciences of the USA*, 5(2):32-36.
- CASTLE, W.E. (1919c). Are genes linear or non-linear in arrangement? *Proceedings of the National Academy of Sciences of the USA*, 5(11), 500-506.

- BATESON, W. (1916). Review of *The Mechanism of Mendelian Heredity*. *Science*, 44.
- DARDEN, L. (1977). William Bateson and the promise of Mendelism. *Journal of the History of Biology*, 10:87-106.
- DARDEN, L. (1991). *Theory Change in Science: Strategies from Mendelian Genetics*, Oxford University Press.
- DIETRICH, M. R. (2000). From gene to genetic hierarchy: Richard Goldschmidt and the problem of the gene. In BEURTON, P., FALK, R. and RHEINBERGER, H.-J. (eds). (2000). *The Concept of Gene in Development and Evolution. Historical and Epistemological Perspectives*. Cambridge: Cambridge University Press, 91-114.
- DUNN, L.C. (1965). *A Short History of Genetics*. New York: McGraw-Hill.
- GOLDSCHMIDT, R. (1917). Crossing-over ohne Chiasmotypie? *Genetics*, 2:82-95.
- JANSSENS, F. A. (1909). Spermatogénèse dans les batraciens. v. la théorie de la chiasmotypie. Nouvelles interprétations les cineses de maturation. *La Cellule*, 25: 387-411.
- KUHN, T. (1977). Objectivity, value judgement, and theory choice, in *The Essential Tension*, Chicago University Press, 320-339.
- MORGAN, T. H. (1910a). Sex-limited inheritance in drosophila. *Science*, 32:1.
- MORGAN, T.H. (1910b). Chromosomes and heredity. *American Naturalist*, 44:449-296.
- MORGAN, T.H. (1911a). Random segregations versus coupling in Mendelian inheritance, *Science*, 34, 873, 384.
- MORGAN, T.H. (1911b). An attempt to analyze the constitution of the chromosomes on the basis of sex-limited inheritance in drosophila. *Journal of Experimental Zoology*, 13:79.
- MORGAN, T.H. and CATTELL, E. (1912). Data for the study of sex-limited inheritance in drosophila. *Journal of Experimental Zoology*, 13:79.
- MORGAN, T.H., STURTEVANT, A. H., MULLER, H.J. and BRIDGES, C.B. (1915). *The Mechanism of Mendelian Heredity*, Henry Holt and Company.
- MORGAN, T.H. and BRIDGES, C.B. (1916). Sex-linked inheritance in drosophila. *Carnegie Institute Washington Publications*, 345-359.
- MORGAN, T.H. (1926). *The Theory of the Gene*, Yale University Press.
- MULLER, H. J. (1916). The mechanism of crossing-over. *American Naturalist*, 50 (I, II, III, IV):193-221, 284-305, 350-366, 421-434.
- MULLER, H.J. (1920). Are the factors of heredity arranged in a line?. *American Naturalist*, 54: 97-121.

- PAINTER, T. P. (1934). A new method for the study of chromosomes aberrations and the plotting of chromosome maps in *Drosophila melanogaster*. *Genetics*, 19:175-88.
- STURTEVANT, A.H. (1913). The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. *Journal of Experimental Zoology*, 14: 43-59.
- STURTEVANT, A.H., BRIDGES, C.B. and MORGAN, T.H. (1919). The spatial relations of genes. *Proceedings of the National Academy of Sciences*, 5:168-173.
- SUTTON, W.S. (1902). On the morphology of the chromosome group in *Brachystola magna*. *Biological Bulletin*, 4, 39.
- WATERS, C.K. (1994). Genes Made Molecular. *Philosophy of Science* 61: 163-185.
- WIMSATT, W. (1987). False models as means to truer theories. In NITECKI, M. and HOFFMAN, A., *Neutral Models in Biology*, Oxford University Press, 23-55.
- WIMSATT, W. C. (1992). Golden generalities and co-opted anomalies: Haldane vs. Muller and the drosophila group on the theory and practice of linkage mapping. In *The Founders of Evolutionary Genetics. A centenary reappraisal*, pp. 101-166. Dordrecht: Kluwer Academic Publishers.





Figure 1. Intertwining of chromosomes during meiosis (Janssens, 1909).

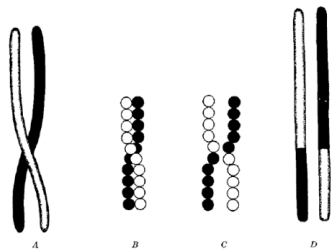


Figure 2. Mechanical model of crossing-over (Morgan *et al.*, 1915).

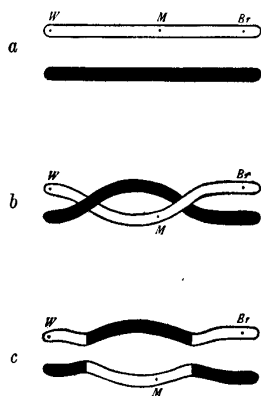


Figure 3. Double crossing-over (Morgan *et al.*, 1915).

Factors concerned	Proportion of crossovers	Percent of crossovers
BCO	$\frac{193}{16287}$	1.2
BO	$\frac{2}{373}$	0.5
BP	$\frac{1464}{4551}$	32.2
BR	$\frac{115}{324}$	35.5
BM	$\frac{260}{693}$	37.6
COP	$\frac{224}{748}$	30.0
COR	$\frac{1643}{4749}$	34.6
COM	$\frac{76}{161}$	47.2
OP	$\frac{247}{836}$	29.4
OR	$\frac{183}{538}$	34.0
OM	$\frac{218}{404}$	54.0
CR	$\frac{236}{829}$	28.5
CM	$\frac{112}{333}$	33.6
B(C, O)	$\frac{214}{21736}$	1.0
(C, O) P	$\frac{471}{1584}$	29.7
(C, O) R	$\frac{2062}{6116}$	33.7
(C, O) M	$\frac{406}{898}$	45.2
PR	$\frac{17}{573}$	3.0
PM	$\frac{109}{405}$	26.9

Figure 4. Table of recombination frequencies (Sturtevant 1913).

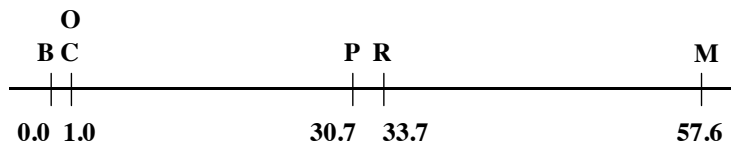


Figure 5. Genetic map corresponding to the table in figure 4 (Sturtevant 1913).

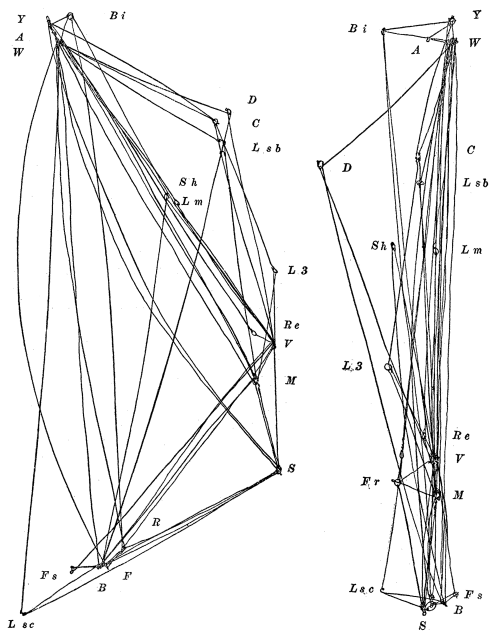


Figure 6. Castle's "rat trap" model under two different perspectives (Castle 1919a), for the same data as Morgan and Bridges' (1916) map (fig. 7).

Models of Data and Theoretical Hypotheses: A Case-Study in Classical Genetics

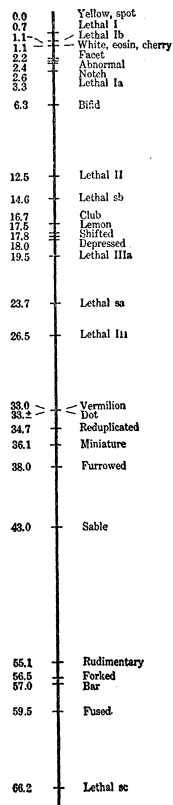


Figure 7. Morgan and Bridges' (1916) map, reproduced by Castle (1919a).