

THE NUMERICAL RELATIONS IN THE CROSSING OVER OF THE GENES, WITH A CRITICAL EXAMINATION OF THE THEORY THAT THE GENES ARE ARRANGED IN A LINEAR SERIES

H. S. JENNINGS

Zoölogical Laboratory of the Johns Hopkins University, Baltimore, Maryland

Received March 8, 1923

TABLE OF CONTENTS

	PAGE
INTRODUCTION	393
The system of relations for which a theory is required to account.	394
Basic propositions.	397
The "linear" theory of crossing over.	400
Obvious coincidences and discrepancies between the relations required by the linear theory, and those shown by the ratios.	401
Relations required by the linear theory.	403
The rules for crossing over when the occurrence of one break does not influence the occur- rence of others (no interference).	403
The number of breaks to be expected between any two genes, in crossing over without interference.	411
Comparison of the condition resulting from serial arrangement of the genes, if there were no interference, with those found in nature.	417
The hypothesis of interference.	422
Crossing over with interference, in case the genes are arranged in a series.	423
Complete interference.	425
Partial interference.	426
The general relations in crossing over with interference.	439
CONCLUSION AND DISCUSSION.	451
SUMMARY.	453
LITERATURE CITED.	456

INTRODUCTION

A large number of ratios showing the proportions of crossing over that occur between different pairs of genes have been determined experimentally for a particular organism, *Drosophila melanogaster*, and published by MORGAN and his co-workers. These ratios show among themselves a number of extraordinary relations, which demand explanation from some unified point of view. The explanation proposed by the workers on *Drosophila*,—the so-called "Linear Theory,"—is that the genes are arranged in the chromosomes in serial order, and that crossing over occurs by the breaking of these linear series, and exchange of parts by the two chromosomes at the time they pair.

The present author (JENNINGS 1918) examined the numerical relations involved in a theory of crossing over proposed by GOLDSCHMIDT (1917) and showed that these are inconsistent with the observed relations. I then undertook a similar study of the numerical relations necessarily implied by the "linear" theory. At the same time a similar study was carried out by HALDANE, who published in 1919 a paper in which some of the mathematical relations required by the linear theory were set forth. I therefore did not at that time complete nor publish my own study. HALDANE'S paper is, however, incomplete; it does not bring out many of the most important relations demanded by the linear theory, nor does it portray the complex system of relations presented by the observed ratios; so that its author did not enter upon a full comparison of the required and the actual conditions.

The present paper therefore attempts to analyze with some thoroughness the conditions required by the linear theory, in order to see how far the ratios and relations found in nature are consistent with and demanded by them. If the linear theory is to be looked upon as a correct account of the conditions in nature, certainly a full exposition of the mathematical relations which it involves is needed. And such an exposition and comparison will be found to furnish an illuminating, perhaps crucial, test of the theory. As will appear, there is a complex system of relations of a really astonishing character shown by the known ratios, and any general theory is almost bound to betray its truth or falsity when confronted with this system. Our analysis will lead us to deal incidentally also with the other theories of crossing over.

The time appears ripe for such an examination. The recent extensive study of the characters connected with the second chromosome of *Drosophila*, by BRIDGES and MORGAN (1919), and the still more recent summary for the three large linkage groups of *Drosophila* by BRIDGES (1921), have, with the earlier paper on the first-chromosome group by MORGAN and BRIDGES (1916), and the many papers of more limited scope, by MORGAN and his co-workers, made available a mass of most valuable data for such an examination.

THE SYSTEM OF RELATIONS FOR WHICH A THEORY IS REQUIRED TO ACCOUNT

Crossing over is of course a phenomenon of experimental breeding; it may be dealt with quite independently of any theory of the arrangement of genes or of their relations to the chromosome. So considered, it consists in the following: When to form an individual there unite two gametes,

one carrying two genes AB that follow the same chromosome, the other gamete carrying two alternative genes ab , we find that the gametes later formed by this new individual have mainly these same combinations, AB or ab . But in a certain proportion of the gametes we find the new combinations Ab and aB ,—so that A has “crossed over” to join with b , and a to join with B . The proportion of all cases in which these new combinations are found is called the crossover ratio,¹ and it varies for different pairs of genes from less than one-half of one percent to about fifty percent.

It is the fact that the ratios for different pairs of genes differ, although there is a relative constancy in the ratio for any given pair, and the fact that these ratios when compared show a number of striking and peculiar interrelations, that yield the basis for theories of the matter. These interrelations may be illustrated as follows:

Among the genes connected with the first or X chromosome of *Drosophila*, that have been most fully studied, are the nine given in the following table. If we examine the crossover ratios of each of these with all the others (as given by MORGAN and BRIDGES 1916), we find that they yield the following table, in which the order is simply alphabetical. (The ratios are given as percentages.)

TABLE 1

Table of best-known crossover ratios in chromosome I of *Drosophila*. Genes arranged alphabetically. (From MORGAN and BRIDGES 1916).

	Bar	Bifid	Club	Miniature	Rudimentary	Sable	Vermilion	White	Yellow
Bar.....	?	?	20.5	2.3	13.8	23.9	43.6	47.9
Bifid.....	?	?	30.6	42.7	?	31.1	5.3	5.5
Club.....	?	?	?	?	?	18.8	14.3	17.7
Miniature.....	20.5	30.6	?	17.9	6.7	3.1	33.2	34.3
Rudimentary.....	2.3	42.7	?	17.9	14.3	24.1	42.4	42.9
Sable.....	13.8	?	?	6.7	14.3	10.1	41.2	42.9
Vermilion.....	23.9	31.1	18.8	3.1	24.1	10.1	30.5	34.5
White.....	43.6	5.3	14.3	33.2	42.4	41.2	30.5	1.1
Yellow.....	47.9	5.5	17.7	34.3	42.9	42.9	34.5	1.1

In this merely alphabetical table, one cannot fail to notice certain peculiar features. The ratios of white and of yellow with the other genes,

¹ The above is the sense in which the terms crossover and crossover ratio have hitherto been used in the literature. I understand that the workers on *Drosophila* are considering the substitution of “recombination” and “recombination ratio” for these terms, employing the term “crossover” in another sense.

for example, are parallel throughout with the values for yellow a little above those for white. A similar relation, though less completely carried out, is seen on comparing the ratios for miniature and vermilion, or those for bar and rudimentary. Further it will be observed that the ratios for certain genes with the rest run higher than for certain others. Thus, for miniature the highest ratio found is 34.3, and the average is 20.9, while for yellow the highest is 47.9 and the average is 28.35. These and other relations discoverable by study suggest that there is some system in these ratios. It appears worth while to try arranging them in order with respect to particular genes, allowing the rest to take the places into which this order forces them, in order to see if any system is thus revealed. We may select for this purpose the gene having the lowest mean ratios (miniature) or that showing the highest (yellow); or any other. Various selections give results of interest; that obtained by selecting as the basic gene the one with the highest mean ratios (yellow) gives however the most illuminating results. When we so arrange the other genes as to give the ratios with yellow in order from lowest to highest, letting the remaining ratios fall where they must, we obtain table 2.

TABLE 2

*Table of best-known crossover ratios among genes of chromosome I of Drosophila, arranged in series with reference to that gene having the highest mean ratios (yellow).**

	Yellow	White	Bifid	Club	Vermilion	Miniature	Sable	Rudimentary	Bar
Yellow.....	1.1	5.5	17.7	34.5	34.3	42.9	42.9	47.9
White.....	1.1	5.3	14.3	30.5	33.2	41.2	42.4	43.6
Bifid.....	5.5	5.3	?	31.1	30.6	?	42.7	?
Club.....	17.7	14.3	?	18.8	?	?	?	?
Vermilion.....	34.5	30.5	31.1	18.8	3.1	10.1	24.1	23.9
Miniature.....	34.3	33.2	30.6	?	3.1	6.7	17.9	20.5
Sable.....	42.9	41.2	?	?	10.1	6.7	14.3	13.8
Rudimentary.....	42.9	42.4	42.7	?	24.1	17.9	14.3	2.3
Bar.....	47.9	43.6	?	?	23.9	20.5	13.8	2.3

If this table be examined, by following the columns from above downward, or the rows from left to right, it is at once evident that, with slight irregularities, the entire set falls into a system, although we have arranged

* In the case of miniature (34.3) and vermilion (34.5) the ratios with yellow are so nearly identical that the difference (0.2 percent) lies far within the probable error, and there is consequently no direct clue to the order in which these two genes should be placed. I have employed the order which relates them most naturally to the discoverable system of relations. If they be interchanged the irregularities of the system will be slightly increased.

it in order only with reference to the gene yellow. The ratios with white form a series parallel to those with yellow, but each a little lower in value, while the ratios with bar form a similar series, but in the reverse order of magnitude. For the other genes the ratios fall into two partial series,—one with magnitudes increasing as in that for yellow, the other in the reverse order. A similar table will be given if we arrange the ratios between the best-known genes of the second chromosome of *Drosophila* in a table in their order of magnitude with relation to the gene having the highest mean ratio. According to BRIDGES and MORGAN (1919), the eight best-known genes of the second chromosome are black, curved, dachs, purple, speck, star, streak and vestigial. If the reader will take the ratios given by BRIDGES and MORGAN and arrange them in series of magnitudes with relation to star or speck, he will obtain a table showing the same peculiarities as does table 2.

Careful examination of such a table reveals general relations, which we will formulate under the following numbered basic propositions:

Basic propositions

1. The ratios vary from below one-half percent to about fifty percent. (In *Drosophila*, ratios as low as four-tenths percent are recorded, and I am informed by Dr. MORGAN that ratios at least as low as two-tenths percent occur.)
2. If two genes show with one another a low crossover ratio, they show nearly the same ratio, low or high, with any other gene. (Thus yellow and white respectively show with rudimentary the ratios 42.9 and 42.4; with club the ratios 17.7 and 14.3, with bifid 5.5 and 5.3, etc.)
3. If two genes have together a high ratio (as do yellow and bar), their ratios with particular other genes are very diverse.
4. The greater the ratio of two genes together, on the whole the more diverse their ratios with a given other gene.
5. If any two genes, *A* and *B*, have together a certain ratio, and *B* has with a third gene *C* a certain ratio, then the ratio of the first (*A*) with the third (*C*) is either a little *less* than the *sum* of the ratios *A-B* and *B-C*, or a little *more* than their *difference*.

Thus, the ratio of yellow to club is 17.7; of club to vermilion 18.8. Their sum is 36.5, while the ratio of yellow to vermilion is 34.5. The ratio yellow-bifid is 5.5; of bifid-miniature 30.6; their sum is 36.1, while the ratio of yellow to miniature is 34.3.

Again, the ratio white-vermilion is 30.5; vermilion-club 18.8; their difference is 11.7, while the ratio white-club is 14.3.

6. If for the gene having the highest mean ratios, as yellow, we arrange in ascending order the series of its ratios with the other genes,—allowing the position of all the genes to be thus determined,—then all the ratios fall into a systematic table (as table 2), having the following characteristics:

7. Genes having a *low* ratio with this basic gene (e.g., white) show with the others a similar ascending series of ratios, but with each ratio slightly *less*.

8. Genes having a higher ratio with this basic gene show a similar ascending series of ratios with genes having a still higher ratio than its own with the basic gene (though with the ratios much decreased in value);

9. But a *reverse* series with all the genes having a lower ratio than its own with this basic gene. (Cf. the ratios of vermilion with others.)

10. In a table so arranged, if in any column (or row) we take for a gene *A* its two ratios with *B* and with *C*, these lying on the opposite sides of the diagonal, then the *sum* of these two ratios gives approximately, but somewhat *more* than, the ratio between *B* and *C*.

Thus if *A* is vermilion, *B* is white and *C* is miniature, then ratio *A-B* is 30.5, ratio *A-C* is 3.1; their sum is 33.6, which is a little more than ratio *B-C*, which is 33.2.

The difference between the sum and the ratio *B-C* becomes greater as *B* and *C* are farther apart in the table.

11. But if *B* and *C* are so chosen that the ratios *A-B* and *A-C* lie on the same side of the diagonal, then the *difference* between *A-B* and *A-C* is approximately, but a little *less* than, the ratio *B-C*.

Thus, if *A* is vermilion, *B* is white and *C* is club, the ratio *A-B* is 30.5; ratio *A-C* is 18.8; their difference 11.7 is a little less than the ratio *B-C*, which is 14.3.

To these relations, evident from the table, must be added certain others that have been established by experiment:

12. When there occurs a crossover between any two genes of the series (shown in the upper row of table 2), there occurs also as a rule a crossover between any gene lying to the right of these two and any other lying to the left of these two.

Thus, if there occurs a crossover between vermilion and sable, as a rule there occurs also a crossover between club and sable, club and rudimentary, club and bar, white and sable, white and bar, etc., etc.

13. But in such a case there is as a rule no crossover between two genes in the series on the same side of the two genes that show a crossover.

Thus, if there is a crossover between vermilion and sable, there is as a rule no crossover between any of the genes yellow-white-bifid-club-vermilion; these all remain in the same grouping as before.

14. That is, such series as are represented in the rows or columns of table 2 usually in crossing over separate into two groups at a certain place in the series, all the members of one group showing a crossover with all the members of the other.

15. But in relatively rare cases, such a series separates at two points in such a way that all the genes between the two points show crossovers with all the genes outside the two points. Still more rarely, the series separates at three points.

16. In a considerable proportion of the cases (30 to 50 percent), no crossing over occurs between any of the genes of the chromosome. (Thus, in the X chromosome of *Drosophila*, MULLER found that there is no crossing over between any of the genes from yellow to bar, in 54.4 percent of the cases.)

17. When a crossover occurs between two genes, *A* and *B*, having together a low ratio, no crossover occurs between one of these genes (as *A*), and any other gene *C* having a low ratio with *A* and lying in the series on the opposite side of *A* from *B*. Or: such a series as that of table 2 does not break at two points that are close together ("interference").

These relations are shown not alone for the 9 genes given in our table. If we take the other genes of the X chromosome, they fall, as far as the data on them goes, into this same system. Or if we select the most thoroughly studied genes of the second chromosome of *Drosophila*, we find that they yield a table similar to table 2, and the propositions set forth above apply equally to them.

It is important to observe that all these relations (table 2, and propositions 1 to 17) are purely experimental results of breeding; they do not depend in the least upon any theory of what crossing over is or of the arrangement of genes. So far as these relations have come to the consciousness of investigators not engaged in work on *Drosophila*, it appears to be the impression that they are simply postulates of some particular theory. On the contrary they are statements of observed fact.

It is of course true that these relations hold only approximately, so far as precise figures are concerned. But when one considers the extremely long and complex process of experimentation required for obtaining each of the ratios; the great number of extraneous factors coming in to disturb the precision of the results, and the number of conditions known to

modify the crossover ratios, it must appear marvelous that the system reveals itself so unmistakably as it does in tables arranged as in table 2. To come out so clearly from beneath such a load of extraneous conditions the system must be of some extremely well-defined character.

It is clear then that there is a system underlying the ratios of crossing-over; *es deutet das Chor auf ein geheimes Gesetz*. What is the law that yields this remarkable system? It is idle to propose theories of crossing over that ignore or conflict with this system; it is for these relations that a theory is called upon to account. In view of the numerous, complex and extraordinary relations shown (set forth in propositions 1 to 17), it would appear improbable that any great number of diverse theories could equally well explain the results; apparently a correct theory should show itself unmistakably.

THE "LINEAR" THEORY OF CROSSING OVER

The so-called "linear" theory holds that the system set forth above is due to the arrangement of the material genes in a serial order; that the genes either are the particles seen under the microscope arranged in serial order in the chromosomes at the stages when the latter pair in the formation of the germ cells, or that at least the genes have the same arrangement as these particles. Crossing over, it is held, is the breaking and interchange of parts of the two chromosomes while they are paired side by side. Naturally, it is set forth, genes that are far apart become separated by breaks more frequently than do those close together; and it is to this that the system described above is due. There results a series of diverse ratios, from high to low, as found in experimental breeding,—these corresponding to genes varying in distance, from contiguity to wide separation.

To judge of this theory, we require an examination of the relations that must hold if the genes were so arranged, for comparison with the relations found in the breeding experiments. Would there result from a serial arrangement the relations illustrated by table 2, and formulated in propositions 1 to 17 above? Further, would there result the particular numerical values actually recorded in the experimental studies? Under this second question we have specifically to inquire:

(1) Whether there would result the crossover ratios found in nature,—limited in the way that they are;

(2) Whether the distances between genes measured in the standard units of length (each such as to give 1 percent of crossing over), shown

in the current diagrams of the chromosomes (e.g., BRIDGES 1921) would result;

(3) Whether the numbers of breaks in the chromosome, that observation, interpreted on the linear theory, shows to occur, would result from the presuppositions of that theory.

OBVIOUS COINCIDENCES AND DISCREPANCIES BETWEEN THE RELATIONS
REQUIRED BY THE LINEAR THEORY, AND THOSE
SHOWN BY THE RATIOS

On the linear theory the genes in a chromosome are comparable to a string of beads. Two such strings (in the known pairing of the chromosomes) lie side by side and intertwine or partly fuse; they break at times and change partners; a portion of each string reunites with a portion of the other to form a new string. If this were true, what would be the rules and ratios of crossing over?

The obvious strong point of the linear theory is that it accounts directly for the facts set forth in our propositions 12 to 15,—according to which crossing over occurs between continuous groups of genes forming the series of table 2. If before crossing over the genes of the two chromosomes form the two series

$$\begin{array}{cccccccc} A & B & C & D & E & F & G & H \\ a & b & c & d & e & f & g & h \end{array}$$

then after a break occurring between the genes *C* and *D*, with resulting change of partners, we have

$$\begin{array}{cccccccc} A & B & C & d & e & f & g & h \\ a & b & c & D & E & F & G & H \end{array}$$

And by the occurrence of two or more breaks we obtain such results as

$$\begin{array}{cccccccc} A & B & C & d & e & f & G & H \\ a & b & c & D & E & F & g & h \end{array} \quad \text{or} \quad \begin{array}{cccccccc} A & B & c & d & E & F & g & h \\ a & b & C & D & e & f & G & H \end{array}$$

But does the linear theory account for the remaining relations? If we hold the genes *A* to *H* to be separated by equal distances, say ten units each, and if we then tabulate their distances from each other, beginning with *A*, we obtain a table having many of the properties shown in table 2 for the crossover ratios. Such a table of distances is given in table 3.

Examining table 3, we find that it shows most, but not *all*, of the properties set forth in our propositions 1 to 17, above. In detail, the identities and differences are as follows. The table of distances shows, if we substitute "distances" for "ratios," precisely the relations shown in propositions 2, 3, 4, 6, 7, 8, 9; and the assumption underlying it accounts, as before remarked, for proposition 12, 13, 14 and 15. But with regard to

propositions 5, 10 and 11, while the relations are analogous in the two cases, there is a systematic difference. In the table of distances, the relations formulated in propositions 5, 10 and 11 are exact; instead of saying in proposition 5 that the ratio of A to C is "a little less than the sum or a little more than the difference" of ratios $A-B$ and $B-C$, we have to say for the distances that the distance $A-C$ is precisely the sum or precisely the difference of the distance $A-B$ and $B-C$ (where A , B and C may represent any of the letters A to H). Similarly, in proposition 10, the sum of distances $A-B$ and $A-C$ gives, not more than $B-C$, but just $B-C$; and in proposition 11, the difference between $A-B$ and $A-C$ is not less than, but exactly $B-C$. This systematic "more than" or "less than" in the table of ratios requires accounting for. Further we find in the table of distances nothing related to or accounting for the facts set forth in propositions 1, 16 and 17.

TABLE 3

Table of distances between nine serial points lying ten units apart on a line. For comparison with the relations shown in the table of crossover ratios.

	A	B	C	D	E	F	G	H	I
A	..	10	20	30	40	50	60	70	80
B	10	..	10	20	30	40	50	60	70
C	20	10	..	10	20	30	40	50	60
D	30	20	10	..	10	20	30	40	50
E	40	30	20	10	..	10	20	30	40
F	50	40	30	20	10	..	10	20	30
G	60	50	40	30	20	10	..	10	20
H	70	60	50	40	30	20	10	..	10
I	80	70	60	50	40	30	20	10	..

It is of course a striking fact that such a table of linear distances shows a considerable number of the curious relations shown by the crossover ratios; this gives the linear theory a certain *prima facie* probability. Strictly, however, these coincidences do not directly support the linear theory, for we are dealing in the one case with ratios, in the other with distances. What we require is to know whether the ratios themselves would give such a system, if they are brought about in the way the linear theory sets forth. We require to know whether they would show the same coincidences with the system found in the genes as are shown by the table of linear distances. And further, would they manifest the same discrepancies with the relations shown by the distances, that we have just noted in connection with propositions 5, 10 and 11? Can the linear theory account for the "less than" and "more than" of these propositions? Does

it account for the relations set forth in propositions 1, 16 and 17, which are not shown by the table of distances? If the linear theory is correct it must yield and indeed require all the relations set forth in our 17 propositions.

More than this, the linear theory, if correct, must be consistent with the actual values found for the crossover ratios, in their relations one to another, and with the frequency of occurrence of single, double, and triple breaks, and of no-breaks, in the series given in table 1.

Can it fulfil all these requirements?

RELATIONS REQUIRED BY THE LINEAR THEORY

We now proceed to an examination of the assumptions and requirements of the linear theory.

If, as it sets forth, the serial strings of genes break and cross over at times, either the breaks must occur as frequently between any two contiguous genes as between any other two, or they occur more frequently at certain points than at others. Further, conditions might be such that (as is commonly held to be the case), the occurrence of a break at a given joint would interfere with the occurrence of another near by. As a basis for the treatment, it is necessary to examine first the generalized case in which breaks are equally frequent at all joints, and there is no interference; then the modifications required if these conditions do not hold may be set forth.

THE RULES FOR CROSSING OVER WHEN THE OCCURRENCE OF ONE BREAK DOES NOT INFLUENCE THE OCCURRENCE OF OTHERS (NO INTERFERENCE)

On the linear theory, in a particular chromosome, a crossover will have occurred between any two genes, A and M (close together or far apart), when a single break has occurred between them; for then the two that were formerly in the same chromosome will be in diverse chromosomes. But, as shown by the diagrams on page 401, if just two breaks have occurred between the two genes, they still remain in the same chromosome (though a piece between them has been exchanged), so that in experimental breeding no crossover would be found to have occurred so far as A and M were concerned. And in general, consideration shows that, on this theory, whenever the number of breaks between two genes, A and M , is odd, they will have passed into different chromosomes, so that a crossover has occurred for the pair $A-M$; while whenever there is an even number of breaks between them, or no break whatever, A and M are still in the same chromosome, so that no crossover has occurred for the pair

A-M. (There may have been crossing over for pairs between these, but this will not be discovered in breeding experiments limited to observation of the genes *A* and *M*).

In other words, on the linear theory crossing over occurs for any particular pair of genes when there is an odd number of breaks between them, and not otherwise, and the law of the frequency of crossing over for them is the law of the frequency of the occurrence of odd numbers of breaks between them. So far as the particular pair is concerned, the effect is the same whether the odd number is 1 or 101 or any other. Our problem therefore resolves itself into that of discovering the frequency of odd numbers of breaks between two genes; it is the same problem as discovering the frequency of odd numbers of breaks between two given beads of a string, when it is pulled, if all breaks occur simultaneously and are just as likely to occur in any one place as in any other.

It is a simple matter to discover the rule for this. Let us suppose that a break occurs between any two adjacent beads in a certain proportion, *b*, of the cases; say in $\frac{1}{10}$ of the cases. Then the proportion of cases in which a break does not occur (which we may call *a*) will be $1-b$; here $\frac{9}{10}$ of the cases. Consider any single adjacent pair: there is possibility of but 1 break between them; for convenience we may say there is but one joint between them. This joint will then break in $\frac{1}{10}$ of the cases, not in $\frac{9}{10}$. We can tabulate as follows:

Breaks	0	1
Proportion	$a(=.9)$	$b(=.1)$

Now, if we take three consecutive genes or beads, *A-B-C*, there are two joints between them. At the first joint there will be 0 breaks in $\frac{9}{10}$ of the cases; at the second, 0 breaks in $\frac{9}{10}$; so there will be 0 breaks at both in $\frac{9}{10} \times \frac{9}{10}$ of the cases. There will be 0 breaks at the first, 1 break at the second, in $\frac{9}{10} \times \frac{1}{10}$ of the cases; 1 break at the first, 0 at the second, in $\frac{1}{10} \times \frac{9}{10}$; 1 break at the first, 1 at the second, in $\frac{1}{10} \times \frac{1}{10}$ of the cases. Thus there will be a total of 0 breaks in $(\frac{9}{10})^2$ (or a^2) cases; a total of 1 break in twice $(\frac{9}{10} \times \frac{1}{10})$ (or in $2ab$) of the cases; a total of 2 breaks in $(\frac{1}{10})^2$ (or b^2) of the cases. Tabulating, we have

Breaks	0	1	2
Proportion	$a^2 + 2ab + b^2 = (a+b)^2$		

If we apply this reasoning to the case of three joints, we find the proportions for the different possible number of breaks, beginning with 0, to be $a^3 + 3a^2b + 3ab^2 + b^3$; and in general for any number, *n*, of joints we find the proportions of the diverse numbers of breaks from 0 to *n* to be given by the expansion of the binomial expression $(a+b)^n$.

Between two genes separated by n joints, the proportion of crossovers is, as we have seen, given by the proportion of *odd* numbers of breaks; and these odd numbers are obviously given by the alternate terms of the expansion of $(a+b)^n$, beginning with the second; that is, those terms containing odd powers of b . If we can find the sum of these alternate terms, we shall obtain the total proportion of crossovers.

This can be done as follows. Take as an example the case of four breaking joints:

Number of breaks	0	1	2	3	4
Proportions	$a^4 + 4a^3b + 6a^2b^2 + 4a^2b^3 + b^4$				

This expression gives the sum of the proportions of the crossovers and the non-crossovers; and since $a+b=1$, any power of $a+b$ is likewise equal to 1; so that the sum of the proportions of crossovers and non-crossovers is 1. If now we can likewise get the difference between the proportion of crossovers and that of non-crossovers, we can of course find the value of each. The difference between the two is obtained by subtracting the terms representing the crossovers,—those with odd powers of b ,—from the sum of the others; that is, by giving the minus sign to the alternate terms containing odd powers of b . If we let N represent the proportion of non-crossovers, C the proportion of crossovers, this then gives, in our example:

$$N - C = a^4 - 4a^3b + 6a^2b^2 - 4ab^3 + b^4$$

But this is the expression for $(a-b)^4$, and for any value of n , it is the expression for $(a-b)^n$, so that $N - C = (a-b)^n$.

Now, a is equal to $1-b$, as we have seen, so that $a-b = 1-2b$, and we can substitute this for $a-b$ in the above expression. We have therefore obtained the two equations:

$$N + C = 1$$

$$N - C = (1 - 2b)^n$$

Subtracting the latter from the former and dividing through by 2, we obtain:

$$C = \frac{1}{2} - \frac{1}{2} (1 - 2b)^n \dots \dots \dots (1)$$

- Where C = the proportion or ratio of crossovers
- b = the frequency of a break at any joint
- n = the number of joints separating the two genes, for which the crossover ratio is sought.

Formula (1) gives at once the value of the crossover ratio C , in case there is no interference, and a break occurs as frequently at any joint as at any other. By its use we can compute at once the value of the crossover ratio for genes separated by any given number of joints and with

any given proportion b of breaks at each joint. Thus, if there is a break between contiguous genes in one case out of 20, and two genes are separated by 10 joints, we have by (1):

$$C = \frac{1}{2} - \frac{1}{2} \left(\frac{9}{10}\right)^{10} = 35.57 \text{ percent}$$

This formula (1) shows at once certain important relations, and others are readily developed from it. It will be necessary to set forth these relations both for their own sake, and because these form the key to our later development of the relations in crossing over with interference.

1. Formula (1) shows at once that if there were no interference, our first basic proposition must necessarily hold; that is, the crossover ratio C would vary from a very low value up to close to 50 percent as a limit, but would never reach or exceed 50 percent. For if a break occurs at a given joint less frequently than no-break, b is less than $\frac{1}{2}$, so that $1 - 2b$, and hence $(1 - 2b)^n$, is always a positive quantity; $\frac{1}{2}(1 - 2b)^n$ then always has a positive value. But by (1), $C = \frac{1}{2}$ minus this positive value; it is therefore always less than $\frac{1}{2}$, i.e., less than 50 percent.

But $(1 - 2b)$ is of course always less than 1; it is a proper fraction; and as higher powers are taken of a proper fraction, its value becomes less; by taking a sufficiently high power, its value may be made as near 0 as one may desire. Hence with high values of n (large numbers of joints), the expression $\frac{1}{2}(1 - 2b)^n$ approaches 0 in value, so that C , or $\frac{1}{2} - \frac{1}{2}(1 - 2b)^n$ approaches $\frac{1}{2}$, or 50 percent, in value. Hence, the crossover ratio could approach 50 percent so closely as to differ from it by less than any given value, but could never reach it.

2. It is obvious that to account for the facts in *Drosophila* on the linear theory, whether there is or is not interference, b must be less than a , and thus less than $\frac{1}{2}$; in other words, that at a given joint a break is less frequent than no-break. For otherwise every chromosome would show many breaks,—many crossovers; it would not be true, as set forth in proposition 16, that a large proportion of the chromosomes show no crossover between any of the genes.

3. For later comparison with conditions resulting from interference, we must bring out a further consequence that would follow if, when there is no interference, b were equal to or greater than a . If in formula (1), b is given the value $\frac{1}{2}$, then whatever the number of joints n , C will obviously be $\frac{1}{2}$; that is, the crossover ratio would be 50 percent for all pairs of genes. If b is given a value above $\frac{1}{2}$, then the expression $(1 - 2b)$ becomes negative in value, and as such a negative quantity is raised to successively higher powers, it is of course alternately negative and positive in value. It will therefore be found that the crossover value

C will be alternately above and below $\frac{1}{2}$, but gradually approaching $\frac{1}{2}$ as n becomes higher. For all genes with odd numbers of joints between them ($n = 1, 3, 5$, etc.), C will be above $\frac{1}{2}$; for all genes separated by even numbers of joints, C will be below $\frac{1}{2}$. Thus, if the frequency of a break at any given joint were .75, the value of C would be successively, for genes separated by 1, 2, 3, 4, joints, 75, 37.5, 56.5, 46.75 percent, and so on.

Thus, if a break occurred more frequently at a given joint than no-break, the order of the genes would not be given by taking the ratios of the gene at the end with other genes in either ascending or descending order. Suppose that a break occurred at a given joint in .75 of the cases, then if there were five genes in the order $A B C D E$, the ascending order, used in formulating the work on *Drosophila*, would yield the series $A C E D B$, while the descending order would yield $A B D E C$,—both quite diverse from the actual series. But, as we have seen, in *Drosophila* b cannot be greater than a , so that this situation is not met.

From the formula (1) for C we may readily derive a formula showing directly what number of joints (and therefore of genes) is required to obtain any crossover ratio, for any given value of b (any given frequency of breaks at each joint) provided there is no interference. When there is but one joint, b is equal to C ; that is, the crossover ratio is the frequency of a break at that joint. Now, for any two diverse values of b , we require of course diverse values of n (number of joints) to give the same crossover ratio. Let these be n_1 and n_2 . Then obviously for the two cases:

$$C = \frac{1}{2} - \frac{1}{2}(1 - 2b_1)^{n_1}$$

and

$$C = \frac{1}{2} - \frac{1}{2}(1 - 2b_2)^{n_2}$$

which yield, by transposition of terms:

$$(1 - 2b_1)^{n_1} = 1 - 2C$$

$$(1 - 2b_2)^{n_2} = 1 - 2C$$

And since C is the same in both cases

$$(1 - 2b_2)^{n_2} = (1 - 2b_1)^{n_1}$$

But if diverse powers of two different quantities are equal, it is a well-known relation that the two exponents are inversely proportional to the logarithms of the two quantities. Hence in this case,

$$\frac{n_2}{n_1} = \frac{\log (1 - 2b_1)}{\log (1 - 2b_2)} \dots \dots \dots (2)$$

Taking now the case that $n_1 = 1$; as remarked above, b_1 is now C ; substituting these two values in the above equation, (and dropping the subscript of b_2 , since it is now unnecessary) we obtain:

$$n = \frac{\log (1-2C)}{\log (1-2b)} \dots \dots \dots (3)$$

Equation (3) gives directly, for the case of no interference, the number *n* of joints required to give any crossover ratio *C*, for any frequency *b*, of breaks at each joint.² Thus: How many joints are required to give 20 percent crossing over if there is no interference and the frequency of a break at any joint is $\frac{1}{2}$ percent? Here *C* = .2 and *b* = .005. Hence:

$$n = \frac{\log (1-.4)}{\log (1-.01)} = \frac{\log .6}{\log .99} = \frac{-.2218487}{-.0043648} = 50.83$$

In the chromosomes we of course know neither the value of *b* (frequency of a break at a given joint), nor the number of joints, *n*, that yield a given crossover ratio *C*. For determining general relations, therefore, we must obtain formulae that are independent of the particular values of *b* and *n*. In experimental breeding the workers who employ the linear theory have attempted to do this by taking as the unit of measurement that length of the chromosome that yields 1 percent of crossovers; this length is then laid off as a foot-rule on the chromosome (by processes of breeding that we need not go into here), the distances between specific genes, and the distances required to give particular crossover ratios, being thus determined in terms of this unit. This permits comparison of the distances required to give diverse values of the crossover ratio. It is found thus, for example, that genes 30 units apart do not yield a crossover ratio three times that of two genes 10 units apart, but definitely less; and similarly of other cases. To determine for the ideal linear theory without interference these relations, independently of the values of *b* and *n*, the following must be taken into consideration. In formula (1), *n* is the distance (number of joints) from one of the two genes to the other. With any given value of *b*, two different crossover ratios, *C*₁ and *C*₂, are produced, of course, by different powers (values of *n*), as *n*₁ and *n*₂. Hence we have:

$$C_1 = \frac{1}{2} - \frac{1}{2}(1-2b)^{n_1}$$

$$C_2 = \frac{1}{2} - \frac{1}{2}(1-2b)^{n_2}$$

By transposition of terms, these equations yield

$$(1-2b)^{n_1} = 1-2C_1$$

$$(1-2b)^{n_2} = 1-2C_2$$

² It is to be noticed in using logarithms of decimal fractions for such purposes as are employed in this paper,—and particularly in cases where the exponents are fractional,—that the logarithms of decimals smaller than unity are minus quantities, not the plus quantities (mantissas) given in the usual tables of logarithms. They are obtainable from the logarithms taken from the tables by subtracting the latter from 10 and giving the result the minus sign. Thus for .2 the logarithm obtained from the table is 9.301030; subtracting from 10 and using the minus sign, this yields -.6989700; for .002 the logarithm is -2.6989700, and so on.

Now, it is a well known general relation that when we have two diverse powers of a single quantity (in this case $1 - 2b$), the two exponents are proportional to the logarithms of the resulting powers: so that in this case

$$\frac{n_2}{n_1} = \frac{\log (1 - 2C_2)}{\log (1 - 2C_1)} \dots \dots \dots (4)$$

By this equation we can therefore readily find for the case of no interference the proportional distances required to produce, with any given value of b , any two crossover ratios. If it be asked: How many times as great a length is required to produce 40 percent crossovers as to produce 20 percent, we have

$$\frac{n_2}{n_1} = \frac{\log (1 - .80)}{\log (1 - .40)} = \frac{\log .2}{\log .6} = \frac{-.6989700}{-.2218487} = 3.1507$$

That is, the distance between two genes that is required to yield 40 percent crossing over would be, if there were no interference, 3.1507 times that required to give 20 percent.

If now we wish to use the distance required to produce some particular crossover ratio (as 5 percent) as a unit of measurement, we have but to let $n_1 = 1$, and give to C_1 this particular value; we may then find for any other crossover value C_2 the number of times this unit distance that is demanded, by an equation derived thus directly from equation (4). Letting "the number of times this unit distance" be represented by r , and replacing in (4) the left side of the equation by this, we have

$$r = \frac{\log (1 - 2C_2)}{\log (1 - 2C_1)} \dots \dots \dots (5)$$

For comparison with the measurements given by the workers on *Drosophila*, we have but to let C_1 in (5) be 1 percent or .01; then $1 - 2C_1 = .98$; this will give us the number of units, each sufficient to give 1 percent crossing over, that are required to produce the crossover ratio C_2 . We may replace the letter r of formula (5) by the letter U , signifying units; dropping the subscript of C_2 , as now unnecessary, we have for the case of no interference:

$$U = \frac{\log (1 - 2C)}{\log .98}, \text{ or } = \frac{\log (1 - 2C)}{-.0087739} \dots \dots \dots (6)$$

For example, in case there were no interference, how many units of distance are required to produce 49.9 percent crossing over? Here

$$U = \frac{\log .002}{-.0087739} = 307.617$$

Again, how many units of distance are required to produce $\frac{1}{2}$ percent crossing over?

$$U = \frac{\log .99}{\log .98} = .497 \text{ (of one unit)}$$

Certain other numerical relations require elucidation:

Suppose we have a certain crossover ratio C_1 , for a certain length. What will be, if there is no interference, the crossover ratio C_2 for a stretch r times as great? Here we have but to arrange equation (5) so as to solve it for the value of C_2 . Equation (5) yields directly

$$\log (1 - 2C_2) = r \log (1 - 2C_1)$$

Now, if the logarithm of $(1 - 2C_2)$ is r times that of $(1 - 2C_1)$, the quantity $(1 - 2C_2)$ is of course the r th power of $(1 - 2C_1)$; that is,

$$1 - 2C_2 = (1 - 2C_1)^r$$

Whence, by transposition, etc.,

$$C_2 = \frac{1 - (1 - 2C_1)^r}{2} \dots\dots\dots (7)$$

Thus, suppose a certain stretch on the chromosome gives 20 percent of crossing over. What crossover ratio will be given by a stretch three times as long?

Here

$$C_2 = \frac{1 - (1 - .4)^3}{2} = .392 = 39.2 \text{ percent}$$

Suppose that we know the distance apart of two genes in our standard units (each yielding a 1-percent ratio). What is their crossover ratio? For this we have but to solve formula (6) for C ; it may be transformed, on the principles employed above, into

$$C = \frac{1 - .98^U}{2} \dots\dots\dots (8)$$

Thus, suppose $U = 20$. Then

$$C = \frac{1 - .98^{20}}{2} = 16.62 \text{ percent}$$

When there is no interference, what crossover ratio is given by the sum of several stretches, the crossover ratio for each of which is known? Call the partial ratios C_1, C_2, C_3 , etc., and the ratio for the entire stretch, C . We could find the value of the ratio C for the total stretch by finding by (6) the values of U for each partial ratio, adding these values of U , and determining by (8) the corresponding value of C . Performing these operations algebraically, we find that they yield the formula

$\log (1 - 2C) = \log (1 - 2C_1) + \log (1 - 2C_2) + \log (1 - 2C_3)$, etc. (9)
which is equivalent to

$$1 - 2C = (1 - 2C_1) (1 - 2C_2) (1 - 2C_3)$$
, etc.

Whence

$$C = \frac{1 - (1 - 2C_1) (1 - 2C_2) (1 - 2C_3), \text{ etc.}}{2} \dots\dots\dots (10)$$

EXAMPLE: Four crossover ratios for successive stretches are, in chromosome I of *Drosophila*: 1.1, 5.3, 31.1, and 10.1 percent. What should be the crossover ratio of the two genes at the extremes if there were no interference?

$$C = \frac{1 - (1 - .022) (1 - .106) (1 - .622) (1 - .202)}{2} = 36.8 \text{ percent}$$

For the common case of the crossover ratio to be expected from the sum of two contiguous regions, the ratio for each of which is known, formula (10) yields for the case of no interference the well known formula

$$C = C_1 + C_2 - 2C_1 C_2 \dots\dots\dots (11)$$

In formula (8) we have the crossover ratio determined for a given number, *U*, of units, each of such a length as to yield a 1-percent crossover ratio. Obviously, it makes no difference in the validity of the formula whether these units are all of the same absolute length or not; whether all contain the same number of joints or not. If the frequency of a break is greater at certain joints than at others, then certain of these standard units will contain more joints than others; but the effect of all the units on the crossover ratio will be the same. That is, with formula (8) we have emancipated ourselves from the assumption with which we began, that breaks are equally frequent between any two adjacent genes. All formulae in which these standard units are employed are similarly independent of the relative frequency of breaks at different joints; the consequences flowing from them are valid, whether this frequency is or is not the same for different joints. This applies practically, in fact, as will be seen, to our entire exposition.

Before proceeding by the use of the above formulae to a comparison of the relations required by the linear theory with those found in experimentation, one other matter requires formulation. This concerns the number of breaks to be expected in an entire chromosome, or between any two genes having a given crossover ratio.

The number of breaks to be expected between any two genes, in crossing over without interference

As before noted, on the linear theory the crossover ratio for any two genes is given by the frequency of odd numbers of breaks between them. The number of breaks might run up into the hundreds; if the number were odd, a crossover would result, exactly as if there were but one break.

But the conditions in inheritance do not admit of large numbers of breaks. If we interpret on the linear theory the facts set forth in our propositions 12 to 16, page 398, we are forced to assume that in a large proportion of the cases there is, for a particular chromosome, no break whatever (proposition 16, page 399); that for most other cases there is but 1 break (propositions 12 to 14); that two breaks occur only very rarely, and that more than two occur with extreme rarity.

The question arises whether these assumptions are consistent with the fundamental postulates of the linear theory. If the genes are in series and crossing over occurs by breaks in this series, would the numbers of breaks in the chromosomes be those which the breeding results, interpreted on the linear theory, require?

We have seen on page 404 that the proportions of the cases showing the different possible numbers of breaks from 0 to n are given by the successive terms of the binomial $(a+b)^n$. The exponent of b in each term indicates the number of breaks of which the total value of that term gives the proportional number. For any given number of breaks d , the proportion will be that shown by the value of the term containing b^d . The term containing b^d is of course $\frac{n! a^{(n-d)} b^d}{d!(n-d)!}$. If we let B_d represent the proportion of cases in which just d breaks occur between two genes separated by n joints, then (if we substitute for a its value $1-b$) we have:

$$B_d = \frac{n! (1-b)^{n-d} b^d}{d! (n-d)!} \dots \dots \dots (12)$$

For our later purposes, this and similar factorial expressions are better written in such a way as to make the factors in the numerator and denominator as few as possible. The expression, $\frac{n!}{d!(n-d)!}$ is equivalent to an ex-

pression in which the numerator consists of d factors, beginning with n and descending by differences of 1, while the denominator is factorial d . Methods of writing in a formula " d factors beginning with n and descending by differences of 1" vary, and all those that I have observed are cumbersome. There appears to be no ambiguity involved in writing this simply $n^{!d}$, and I shall therefore employ this method. Thus $23^{!4} = 23 \times 22 \times 21 \times 20$. Using this method, formula (12) becomes

$$B_d = \frac{n^{!d} (1-b)^{n-d} b^d}{d!} \dots \dots \dots (13)$$

(Where $d!$ = factorial d , and $n^{!d} = d$ factors beginning with n and descending by differences of a single unit.)

This formula gives us the proportion of cases having any number d of breaks when the value of n and of a and b are known. But in practice these are not known, except that b is less than .005. What we desire to know is the proportion of chromosomes having each particular number of breaks for a given crossover ratio. In other words, we wish to know the value of B_d in terms of C , the crossover ratio.

For this purpose, we may begin with B_0 , the proportion of chromosomes that have no break. This is, as we have seen $(1-b)^n$.

Now, by formula (3), page 408, the number of joints n is given by the formula:

$$n = \frac{\log(1-2C)}{\log(1-2b)}$$

We can therefore substitute this for n in the expression $B_0 = (1-b)^n$; in other words the value of B_0 is given for any value of C by raising $1-b$ to the power represented by $\frac{\log(1-2C)}{\log(1-2b)}$. To do this, we may of course multiply the logarithm of $(1-b)$ by $\frac{\log(1-2C)}{\log(1-2b)}$, and this will give us the logarithm of B_0 . Performing this operation (and changing the order of the factors), we obtain:

$$\text{Log } B_0 = \log(1-2C) \cdot \frac{\log(1-b)}{\log(1-2b)} \dots \dots \dots (14)$$

Now, as we have seen, b is necessarily less than .005. It will be found that for $b = .005$, the value of the expression $\frac{\log(1-b)}{\log(1-2b)}$ is almost exactly $\frac{1}{2}$ (it is .49874), and that as b is made smaller, this expression approaches $\frac{1}{2}$ as a limit. For all of the possible values of C it may be therefore taken as $\frac{1}{2}$, so that our equation becomes

$$\text{Log } B_0 = \frac{1}{2} \log(1-2C) \dots \dots \dots (15)$$

But of course if $\log B_0$ is $\frac{1}{2}$ of $\log(1-2C)$, it follows that B_0 is the square root of $1-2C$. We obtain then the following

$$B_0 = \sqrt{1-2C} \dots \dots \dots (16)$$

That is, if there is no interference, for two genes showing a given crossover ratio C , the proportion (B_0) of cases in which there is no break between the genes is the square root of $(1-2C)$. Thus if the crossover ratio is 10 percent ($C = .1$), the proportion of cases in which there is no break between the two genes is $\sqrt{.8}$, which is 89.44 percent. If $C = .40$, $B_0 = \sqrt{.2} = 44.72$ percent.

We have then the value of B_0 expressed entirely in terms of C , without regard to the precise value of n and b . To obtain corresponding expressions for B_1, B_2 , and in general for B_d , we may note the following: The value of B_0 is $(1-b)^n$, and this is, therefore, as we have just seen, equivalent, in the cases with which we are dealing, to $\sqrt{1-2C}$. We may therefore substitute $\sqrt{1-2C}$ for $(1-b)^n$ in formula (13) for B_d . Noting that

$$(1-b)^{(n-d)} = \frac{(1-b)^n}{(1-b)^d}, \text{ we thus obtain:}$$

$$B_d = \frac{n^{!d} b^d}{d! (1-b)^d} \cdot \sqrt{1-2C} \dots \dots \dots (17)$$

In this formula, the expression $n^{!d}$ signifies d factors, beginning with n and descending by differences of 1. For our purposes, as we shall see, only low values of d have to be considered,—from 0 up to 5, or thereabouts. Furthermore, as before remarked, we know that the value of b is less than .005. We wish to determine whether, as n is increased, there is a fixed or limiting value for the proportion of cases having any particular number of breaks, as 0, 1, 2, 3, etc.

Now, for high values of n , with low values of b and d , we can substitute n^d for $n^{!d}$ in formula (17), without making any appreciable difference in the result, as will be seen if b is given the value of .005, d any value up to 10; and n any value above 20. In no case will there be a difference in the result greater than a unit in the fourth decimal place, which is, of course, negligible. For the conditions in crossing over, therefore, we can make this substitution, and with slight change in the arrangement of the terms we obtain from (17) the following:

$$B_d = \left(\frac{nb}{1-b} \right)^d \cdot \frac{\sqrt{1-2C}}{d!} \dots \dots \dots (18)$$

Now, as we have seen earlier, when $n=1, b=C$, so that in this case

$$\frac{nb}{1-b} = \frac{C}{1-b}$$

For any higher value of n , with the same value of C , we have by equation (3),

$$n = \frac{\log (1-2C)}{\log (1-2b)}$$

From this equation we can obtain the values of b and of $1-b$, in terms of n and C . The steps may be indicated as follows. We have directly

$$\log (1-2b) = \frac{\log (1-2C)}{n}$$

Now, if $\log(1-2b)$ is equal to $\log(1-2C)$ divided by n , it of course follows that $1-2b$ is the n th root of $1-2C$; that is

$$1-2b = \sqrt[n]{1-2C}$$

whence

$$b = \frac{1 - \sqrt[n]{1-2C}}{2}$$

and

$$1-b = \frac{1 + \sqrt[n]{1-2C}}{2}$$

From which it follows that

$$\frac{nb}{1-b} = \frac{n(1 - \sqrt[n]{1-2C})}{1 + \sqrt[n]{1-2C}}$$

and therefore

$$B_d = \left(\frac{n(1 - \sqrt[n]{1-2C})}{1 + \sqrt[n]{1-2C}} \right)^d \cdot \frac{\sqrt{1-2C}}{d!} \dots \dots \dots (19)$$

Now, in the expression $\sqrt[n]{1-2C}$, the quantity under the radical is necessarily a proper fraction (since C is less than $\frac{1}{2}$). As we extract successively higher roots of a proper fraction, it becomes greater, approaching 1 as a limit. Thus, for high values of n , the expression $\sqrt[n]{1-2C}$ approaches 1 as a limit. Hence, in the first member of the above value for B_d , the denominator, $1 + \sqrt[n]{1-2C}$, approaches 2 as a limit, as n becomes large.

In the case of the numerator, $n(1 - \sqrt[n]{1-2C})$, if the numerical values of this expression are worked out for a given value of C with successively higher values of n , it becomes obvious that it is approaching a limit. Now the limit of the expression $n(\sqrt[n]{x}-1)$ is the natural logarithm of x . Our expression is of the same form, with $(1-2C)$ taking the place of x , and with the signs reversed. That is, the limit of $n(1 - \sqrt[n]{1-2C})$ is the *minus* natural logarithm of $(1-2C)$.* Since $1-2C$ is itself a proper fraction, the minus natural logarithm is itself a positive quantity.

Substituting the two limits we have thus obtained, in formula (19), we get:

$$B_d = \frac{[-\text{nat} \log(1-2C)]^d}{2^d} \cdot \frac{\sqrt{1-2C}}{d!} \dots \dots \dots (20)$$

We have now obtained an expression for the value of B_d corresponding to any given value of C , that is quite independent of the particular values

* I am indebted to Professor L. S. HULBURT of the Mathematical Department of THE JOHNS HOPKINS UNIVERSITY, for pointing this out to me. The natural logarithm of any number is of course obtained from its common logarithm by multiplying the latter by 2.30258509. In practice, the natural logarithm of a decimal fraction is obtained by taking its logarithm from a table, subtracting 10 from it, and multiplying the remainder by 2.30258509.

of n and of b , provided that n is large and b is small (as we know to be the case).

An example will make clear the use of formula (20). If the crossover ratio is 25 percent, in what proportion of the cases are there three breaks between the two extreme genes? Here $d=3$, while $C=.25$; so that the proportion of cases yielding B_3 is:

$$B_3 = \frac{(-\text{nat log } .50)^3}{2^3} \cdot \frac{\sqrt{.50}}{3!}$$

$$= \frac{.69314719^3}{8} \times \frac{.70710678}{6} = .004906, \text{ or } .49 \text{ of } 1 \text{ percent}$$

The values given by this formula are the limiting values in case n is a large number, and this condition is fulfilled when the frequency of a break between two contiguous genes is small; that is, when b is small. We know that b is certainly smaller than .005. The lowest possible values of n will therefore be higher than those corresponding to the value of .005 for b ; these may be obtained for any value of C by equation (3), page 408. If now we work out by formula (13), page 412, the values of B_0, B_1, B_2, B_3 , etc., for $b=.005$, we find that these do not differ, for any possible value of C , or for any number of breaks from 0 to 5, by any appreciable amount from the values given by the general formula (20); in no instance is the difference as great as one-tenth of one percent. For example, for 25 percent crossing over, if we take the value of b to be .005, and determine by (3) the number of joints necessary to produce this crossover ratio, then work out directly by the binomial formula (13) the proportion of cases yielding 0, 1, 2 and 5 breaks, we obtain the result given as (A) below, as compared with the proportions in (B), worked out by the general formula (20):

Number of breaks	0	1	2	5
(A) from binomial formula (13).....	.70671	.24507	.042177	.0003955
(B) from general formula (20).....	.70710	.24594	.042467	.0004251

The values given at (A) are lower than the lowest possible limits, those at (B) are the highest possible limits, for the proportions. Such minute differences are obviously negligible, so that the values obtained from formula (20) may be taken as the correct proportions of cases showing the different number of breaks. In table 4 are given these proportions for a series of values of the crossover ratio C .

Table 4 shows in general that for moderate values of the crossover ratio, the number of breaks to be expected is small, much as appears to be the

case in nature, if interpreted on the linear theory. More precise comparisons are found in the following section.

TABLE 4

Linear theory; no interference. Showing for typical diverse values of the crossover ratio C , the distance apart of the genes (in units each of a length to give a crossover ratio of 1 percent), computed by formula (6), and the percentages of the cases in which there would be 0, 1, 2, 3, 4 or 5 breaks between the two genes yielding the given ratio C computed from formula (20).

C	DISTANCE IN UNITS OF 1 PERCENT	PERCENTAGE THAT SHOWS THE INDICATED NUMBERS OF BREAKS, FROM 0 TO 5, BETWEEN THE TWO GENES					
		0	1	2	3	4	5
1	1.00	98.997	0.01
5	5.22	94.868	4.998	0.132	0.002	0.00003
10	11.05	89.443	9.979	0.557	0.021	0.0006
15	17.66	83.666	14.928	1.330	0.079	0.0035	0.000001
20	25.29	77.428	19.858	2.497	0.205	0.0124	0.000006
25	34.31	70.711	24.507	4.247	0.491	0.0425	0.000029
30	45.36	63.199	29.074	6.615	0.992	0.1104	0.000106
35	59.60	54.795	33.046	9.905	1.963	0.2885	0.000361
40	79.67	44.721	35.988	14.480	3.884	0.7814	0.001258
45	113.97	31.623	36.407	20.958	8.043	2.315	0.005330
49	193.64	14.141	27.598	26.930	17.518	8.547	0.033743
49.9	307.61	4.472	13.928	21.590	22.362	17.372	0.107957

COMPARISON OF THE CONDITION RESULTING FROM SERIAL ARRANGEMENT
OF THE GENES, IF THERE WERE NO INTERFERENCE, WITH THOSE FOUND
IN NATURE

By the use of the formulae and tables set forth above, it becomes a simple matter to determine how far the conditions demanded by the linear theory, would, if there were no interference, agree with those found in experimental breeding,—as summarized in propositions 1 to 17, pages 397–399. It will be well to examine these relations here, for comparison later with the more complex and less perspicuous case in which there is interference. For this purpose it will be helpful to tabulate the crossover ratios for a number of genes, A to H , that are successively 10 units apart (each unit of length to yield a one-percent crossover ratio), on the assumption that there is no interference. This is of course readily done by the use of formula (8). The results are shown in table 5.

Table 5 shows that if there were no interference, all of the general relations set forth in our fundamental propositions 1 to 17 (pages 397–399) as requiring explanation would be fully accounted for, except that num-

bered 17. We have already seen that the linear theory accounts immediately for proposition 1 (page 397) and propositions 12 to 15 (page 398). We saw also that the *distances* between objects in serial arrangement show the relations numbered 2, 3, 4, 6, 7, 8, 9, but do not show accurately those numbered 5, 10 and 11 (see page 397, and table 3). Examination of table 5 shows that when we take the required crossover ratios, in place of the distances, all these relations (propositions 1 to 11) are fully accounted for. In the case of propositions 5, 10 and 11, which the distances did not correctly fit, the required ratios (table 5) show precisely the conditions found in the ratios discovered in breeding. The theory requires

TABLE 5

Crossover ratios for eight genes, A to H, if they are separated by successive stretches each 10 units in length (when there is no interference).

	A	B	C	D	E	F	G	H
A	9.2	16.8	22.9	27.5	31.6	35.2	37.7
B	9.2	9.2	16.8	22.9	27.5	31.6	35.2
C	16.8	9.2	9.2	16.8	22.9	27.5	31.6
D	22.9	16.8	9.2	9.2	16.8	22.9	27.5
E	27.5	22.9	16.8	9.2	9.2	16.8	22.9
F	31.6	27.5	22.9	16.8	9.2	9.2	16.8
G	35.2	31.6	27.5	22.9	16.8	9.2	9.2
H	37.7	35.2	31.6	27.5	22.9	16.8	9.2

precisely that, as in proposition 5, the ratio of gene A to gene C should not be just equal to the sum or the difference of ratios A to B and B to C, but should be either somewhat less than the sum, or somewhat more than difference; this will be seen by examination of the ratios in table 5. In the same way, the theoretical ratios of table 5 show precisely the peculiar relations of the actual ratios that are formulated in propositions 10 and 11. Further, table 4 shows that, as set forth in proposition 16, a considerable proportion of the chromosomes would be found that show no breaks whatever even between genes lying far apart (as indicated by the high crossover ratio they yield).

It remains to inquire, as set forth page on 403, how far the requirements of the linear theory, without interference, are consistent with the actual numerical values of the ratios discovered in the experimental studies, and with the actual numbers of breaks in the chromosome that observation, interpreted on the linear theory, shows to occur. We shall at the same time inquire how far the distances apart of the genes, represented on the maps constructed by the students of *Drosophila*, are consistent with this form of the linear theory.

With regard to the observed numerical values of the ratios, we find, by application of formula (10), page 411, that the actual ratios for genes that are, on the linear theory, far apart, are systematically too high in comparison with the ratios for the partial distances of which these longer stretches are made up, so that they could not be accounted for on the linear theory without interference. For making these comparisons, it is necessary to employ only ratios that have been determined with sufficient accuracy to be fairly dependable; we will arbitrarily select as such those based on statistics from at least 5000 cases. We find determinations based on numbers above 5000 for several stretches of chromosome I of *Drosophila* given by MORGAN and BRIDGES (1916), and certain additional ones are obtainable from the data of PLOUGH (1917). For the shorter stretches of chromosome I we find the following ratios in MORGAN and BRIDGES: yellow-white, 1.1; white-vermilion, 30.5; vermilion-miniature, 3.1; miniature-rudimentary, 17.9.

If now we compute from these the ratios for longer stretches required by the linear theory if there is no interference, we find these required ratios to be the following, in comparison with the observed ratios:

	<i>Required</i>	<i>Observed</i>
Yellow-miniature.....	31.6	34.3
White-rudimentary.....	38.2	42.4

Again, we find another set of satisfactorily determined shorter stretches, as follows: White-vermilion, 30.5; vermilion-sable, 10.1; sable-bar, 13.8; bar-fused, 2.5.

Computing from these the required ratios for the longer stretches, and comparing with the observed, we have:

	<i>Required</i>	<i>Observed</i>
White-bar.....	38.7	43.6
Vermilion-bar.....	21.1	23.9
Vermilion-fused.....	22.5	25.8

From PLOUGH (1917) we have the following additional data: sable-garnet, 1.4; garnet-forked, 12.0; giving:

	<i>Required</i>	<i>Observed</i>
Vermilion-forked.....	20.0	22.5

For chromosome II, BRIDGES and MORGAN (1919) give the following ratios, each based on more than 5000 cases: star-black, 37.9; black-curved, 22.7; curved-speck, 30.2; yielding for longer stretches:

	<i>Required</i>	<i>Observed</i>
Star-speck.....	47.4	48.3
Star-curved.....	43.4	45.9

With the additional ratio black-purple 6.2, we obtain:

	<i>Required</i>	<i>Observed</i>
Star-purple.....	39.4	43.7

We have also the ratios dachs-black, 17.8; black-vestigial, 17.8, which yield:

	<i>Required</i>	<i>Observed</i>
Dachs-vestigial.....	29.3	29.6

Further, from the ratios black-arc, 42.6; arc-morula, 7.9; purple-curved, 19.9; and curved-speck, 30.2, we obtain the comparisons:

	<i>Required</i>	<i>Observed</i>
Black-morula.....	43.8	46.6
Purple-speck.....	38.1	45.7

For the third chromosome, GOWEN (1919) gives the following ratios (each based on 31,456 cases):

Sepia-dichaete, 11.5; dichaeete-spineless, 9.9; spineless-sooty, 10.6; sooty-rough, 20.3.

For the longer stretches these yield the following comparison of required and observed ratios:

	<i>Required</i>	<i>Observed</i>
Sepia-rough.....	35.5	39.1
Dichaete-rough.....	32.0	36.1
Sepia-sooty.....	25.7	26.2

Thus, it is a fact that the crossover ratios between genes that are (on the linear theory) far apart, are found to be regularly greater than is possible on the linear theory without interference, when we consider what the ratios are for shorter intervening stretches taken separately. It is therefore clear that the linear theory without interference cannot account for the actual numerical relations found among the ratios.

As to the actual numbers of breaks occurring between given distant genes of a chromosome, determinations, based on the theory of the linear arrangement of the genes, have been made by MULLER (1916), PLOUGH (1917), WEINSTEIN (1918) and GOWEN (1919), for the first, second and third chromosomes of *Drosophila melanogaster*. Certain unpublished determinations for the first chromosome of *Drosophila virilis* have been furnished me, further, by WEINSTEIN. To compare the proportions of these observed numbers of breaks with the proportions required by the linear theory without interference, the observed proportions having no break (B_0) must be taken as the point of reference, since these (as set forth later) are not altered by interference. From equations (16) and (20)

a formula is readily obtained for determining the proportion (B_d) of cases showing any given number of breaks d , corresponding to the proportion with no breaks (B_0); it is:

$$B_d = \frac{(-\text{nat log } B_0^2)^d \cdot B_0}{2^d \cdot d!} \dots\dots\dots(20a)$$

In order to avoid repetition, the results of this comparison are tabulated on a later page, in connection with those of a similar comparison for different extents of interference, in table 13, page 449. As will be found on examination of this table, in every case the proportion of cases with one break is in nature much too high, while the proportion with more than one break is much too low, as compared with what is required by the linear theory without interference. For example, in GOWEN'S study of the numbers of breaks in the third chromosome of *Drosophila*, based on 31,456 cases, with 54.6 percent showing no break, the proportion required by the theory is 33.1 percent with one break, while the observed proportion is 38.9 percent. On the other hand, for two breaks the theory requires 10.0 percent, while only 6.3 are found; theory requires 2.0 percent with three breaks, but only 0.3 percent are found; and these results are typical.

Thus the linear theory without interference will not account for the proportions having the different numbers of breaks actually observed; the latter show an excess of those with but 1 break; a deficiency of those with 2 or more breaks.

We may inquire finally as to the consistency of the "map distances" between the genes, given by the students of *Drosophila*, with the linear theory without interference. These map distances are of course not matters of direct observation, but they are determined by such careful methods that we may accept them as approximately what must be the actual relations, if the genes are arranged serially. Table 6 shows the "map distances" given by BRIDGES (1921) for the more important genes of chromosomes I and II of *Drosophila*, in comparison with the distances required by the linear theory without interference. The map distances are of course given in units, each of such length as to yield by itself a 1-percent crossover ratio. For chromosome I, I have used the distances from the gene white instead of the gene yellow, since the crossover ratios from the former have been much more accurately determined. These ratios are given in table 2. The ratios for the second chromosome, from the end gene, star, are those given by BRIDGES and MORGAN (1919). From these ratios the number of units required by the linear theory without interference are computed from formula (6).

As table 6 shows, if there were no interference, the distances (measured in units of length to give 1 percent crossing over) required to give the observed crossover ratios would be much greater than those determined by BRIDGES in the chromosomes of *Drosophila*.

In sum, therefore, we find that if the genes were arranged in linear series, and crossing over took place by breaking and exchange of partners, and if a break at a given joint did not interfere with a break elsewhere, then all of the general relations found to hold among the observed ratios, summarized in our basic propositions 1 to 17 (page 397), would necessarily hold, except the last one (number 17).

TABLE 6

Comparison of the map distances given by BRIDGES (1921) with those required by the linear theory without interference.

CHROMOSOME I			CHROMOSOME II		
Gene	Map distances BRIDGES 1921	Required (no inter- ference)	Gene	Map distances BRIDGES 1921	Required
White.....	0.0	0.0	Star.....	0.0	0.0
Bifid.....	5.8	5.5	Streak.....	15.4	18.2
Club.....	15.2	16.7	Dachs.....	29.1	42.5
Vermilion.....	31.5	46.6	Black.....	46.5	70.2
Miniature.....	34.6	54.0	Purple.....	52.7	102.8
Sable.....	41.5	86.0	Curved.....	73.5	123.8
Rudimentary.....	53.0	93.2	Speck.....	105.3	167.4
Bar.....	55.5	101.8			

But in addition to this discrepancy as to proposition 17, we find three discrepancies as to the numerical relations involved. These are:

(1) The observed numerical values of the ratios could not occur; the observed ratios for distant genes are regularly greater than they would be on the assumptions set forth above.

(2) The proportions of the chromosomes showing the different numbers of breaks, from 0 to 3 or 4, would not be those actually found; the latter show a marked excess in the proportion of those with 1 break, and a corresponding deficiency of those with 2 or 3 or more breaks.

(3) The "map distances" of the genes that have been determined in the chromosomes of *Drosophila* are much less than are required by the linear theory without interference.

THE HYPOTHESIS OF INTERFERENCE

To account for the relation set forth in proposition 17, namely, that when a crossover occurs between two genes having together a low crossover

ratio, no crossover occurs between one of these genes and any other gene having a low ratio with it and lying on the same side of the crossing-over break, in the series given by such a table as table 2, we know that the students of *Drosophila* have proposed the hypothesis of interference. This asserts that the occurrence of a break at a certain joint in the linear series interferes in some way with the occurrence of another break at any joint near by. Various suggestions have been made as to how such interference might be brought about; with these we are not concerned.

The hypothesis of interference was proposed only to account for the relation set forth in our proposition 17, not at all with reference to the three numerical discrepancies above set forth. We have now to examine whether it would in fact do away with these three discrepancies; whether, in other words, interference would (1) increase the crossover ratios of distant genes in proportion to those of genes close together; (2) increase the proportion of cases with a single break at the expense of those with more than one break; (3) decrease the "map distances" in the chromosomes. Would it have these effects to just the extent required to bring the facts into consonance? And of course it might turn out that interference would disturb some of the other relations, set forth in basic propositions 1 to 16: this would be fatal.

It is obvious that the fulfilment of these conditions imposes a severe test on the theory of interference. To this test we now turn.

CROSSING OVER WITH INTERFERENCE, IN CASE THE GENES ARE ARRANGED IN A SERIES

We next examine the conditions which must hold if crossing over occurs in linear series of genes in the method set forth in the linear theory, but the occurrence of a break at a certain joint interferes with the occurrence of another break at any joint within a certain distance.

When there is no interference, the proportions of the chromosomes showing the diverse numbers of breaks, from 0 up, between genes separated by n joints, are given, as we have seen by the successive terms of the binomial $(a+b)^n$, where b is the frequency of a break at any joint, and $a = 1 - b$.

Giving, for example, the value 4 to n , we have the following proportions:

Number of breaks	0	1	2	3	4
Proportions	$a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4 = 1$				

Now, in case there is interference, it is evident that this formula suffers the following transformations:

(1) The proportion showing 0 breaks (a^n) remains unaltered, for interference is effective only after the chromosome concerned has a break.

(2) All those that would give a single break if there were no interference—that is, the second term, $n.a^{(n-1)}b$,—still give a single break, since interference does not come into action until after there is one break.

(3) The proportions that would give respectively 2, 3, etc., breaks,—that is, the terms beyond the second—are now reduced, some of their component parts giving in case of interference a smaller number of breaks.

Thus, those chromosomes which would have 2 breaks (the proportion of which is $6a^2b^2$ in the above instance) will some of them, if there is interference, have but 1 break. All those in which the two breaks would be within the distance over which interference occurs, will now have but one break; the rest will retain two breaks.

Similarly, for the proportion (fourth term) that would otherwise have 3 breaks; if all these three breaks are within the distance over which interference is effective, there will be but 1 break; if but two are within this distance, these two will be reduced to 1, while the third will remain; so that the chromosome will have but 2 breaks in place of 3. In the same way, some that would have 4 breaks will now be reduced to 3, 2 or 1 break; and so of all the other terms.

(4) That is, the proportion that has just one break will in case of interference, be increased by contributions from all the succeeding terms. The proportion for any greater number, as 2 breaks, will be decreased by the contributions which it makes to those having but 1, but increased by contributions from all the terms for higher numbers of breaks.

It is thus obvious that among the general results of interference as compared with the results if there were no interference, are the following: (a) the proportion of chromosomes with no breaks is unaltered; (b) the proportion with a single break is increased. This, as we saw on an earlier page (421), is what is required to bring the results as to numbers of breaks into consonance with the linear theory. Whether it will do this in the correct proportions, and whether it will, as required, increase the crossover ratios between distant genes in comparison to those between genes that are near together, and decrease the map distances; also whether it will in any way disturb the consonance of the observed relations with the linear theory,—these remain to be determined.

To determine these things, we require formulae giving the crossover ratios and the numbers of breaks occurring, in case there is interference; we can then compare the results with those from crossing over without interference. A general formula for the numbers of breaks occurring in

case of interference would be obtained from the binomial expression $(a+b)^n$, by determining what part of each of the terms is to be transferred to the terms preceding it. What part of those that without interference would show two breaks, will now show but one? And so of the remaining terms.

Complete interference

We shall deal first with the relatively simple possible case that a single break in the chromosome prevents the occurrence of another break in this entire chromosome. As before, we let b represent the frequency of a break at any particular joint *so long as interference has not come into operation*, while a represents the frequency of no-break at any particular joint, under conditions of no interference. The total number of joints in the chromosome we designate n ; interference extends over the entire n joints.

When there is no interference, as we have seen, the relative numbers of chromosomes showing the different possible numbers of breaks, from 0 up, are given by the successive terms of the binomial $(a+b)^n$. If, when there has occurred a single break, no other occurs, it is obvious that all the chromosomes will now show a single break, except those which, in case there were no interference, would show no break. The proportion of these latter is given by the first term of the binomial $(a+b)^n$; it is of course a^n . The proportion of the chromosomes showing a single break is therefore in the case of complete interference, $1-a^n$; and this of course will be the crossover ratio, as well as the value of B_1 (the proportion of cases having a single break). We therefore have for the case of complete interference for the entire length of the chromosome the formula:

$$C = B_1 = 1 - a^n \dots \dots \dots (21)$$

From this formula it is clear that, if interference is complete throughout the length of the chromosome, C may have any value between 0 and 1. Crossover ratios above 50 percent may occur as readily as ratios below 50 percent. For whatever value be given to a (the proportion of cases in which a given joint is not broken, so long as interference has not come into action), if n be taken sufficiently great a^n will become less than $\frac{1}{2}$, so that C will be greater than $\frac{1}{2}$ or 50 percent. Thus, if for b we take the super-maximal value of .005, so that $a = .995$, then for $n = 138.28$, $C = \frac{1}{2}$ or 50 percent; for $n = 277$, $C = 75$ percent; for $n = 455$, $C = 90$ percent. In general the number of joints n , between the two extreme genes, necessary to give for them any particular crossover ratio C is given by transforming formula (21) as follows:

$$1 - C = a^n$$

Whence

$$n = \frac{\log(1-C)}{\log a} \dots\dots\dots (22)$$

In formula (22) of course n is the total number of joints in the chromosome, and C the crossover ratio for the two end genes.

In the case of complete interference, if any given joint shows a break as frequently as any other, the actual mean frequency of a break (which we may call β , to distinguish it from b , the frequency when there is no interference) is, of course:

$$\beta = \frac{1-a^n}{n} \dots\dots\dots (23)$$

And for genes separated by any number, m , of joints the crossover ratio is the frequency of a single break in the total m joints, so that:

$$C = B_1 = \frac{m(1-a^n)}{n} \dots\dots\dots (24)$$

(Where n is the total number of joints in the chromosome.)

In this case, for comparing the crossover ratios of two stretches of the chromosome, it is evident from (24) that their ratios are directly proportional to the number of joints they contain. That is, if C_1 is the ratio for m_1 joints, then for C_2 , the ratio for m_2 joints, we have

$$\frac{C_2}{C_1} = \frac{m_2}{m_1}, \text{ whence } C_2 = \frac{m_2 C_1}{m_1} \dots\dots\dots (25)$$

And of course the crossover ratio for the sum of two or more contiguous stretches of the chromosome is the sum of the ratios for the several stretches taken separately. The value of the crossover ratio between any two genes would be the same as the number of units (each yielding 1 per cent of crossing over.)

No cases are thus far known of complete interference; the observations show that, interpreted on the linear theory, some of the chromosomes show more than a single break. We proceed therefore to examine:

Partial interference

In partial interference, the occurrence of a break at a given joint interferes with another break within a certain distance from that joint, but this distance is not so great as to include the entire length of the chromosome. Hence, two or more breaks may occur in the chromosome, depending on its length.

The generalized problem of partial interference is complicated by the fact that the results differ slightly, depending on the order in which the fate of the different joints of the chromosome is determined. This may be illustrated by a simple example. Suppose that when a break occurs, this prevents the occurrence of another break for three joints in each direction. Then if in the 6 joints shown below, there would without interference occur three breaks (b), at joints 1, 3 and 5,

$$\begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \\ & & b & & b & \\ & & & & & b \end{array}$$

as shown, the number of breaks remaining in case there is interference will depend upon which of the three breaks is produced first. If number 1 is produced first, this will interfere with number 3, but not with number 5, so that there will remain two breaks. But if number 3 is produced first, this will interfere with both the other two breaks (since these are both within 3 joints of it), so that there would now remain but a single break. The order of determination affects in other ways also the proportions showing particular numbers of breaks.

If, however, the linear theory holds for the chromosomes, it is extremely probable that the determination of the fate of the joints occurs in serial order, from one end of the chromosome to the other. For the cytological evidence indicates that when the conjugation of the chromosomes occurs, in the very long thread-like stage, the union begins at one end and progresses toward the other; and the separation of the two united chromosomes appears to take place in a similar manner. The determination of the fate of the joints must apparently occur either when the two chromosomes are uniting, or when they are separating.

It appears therefore that a solution of the problem for determination of the fate of the joints successively in linear order, is what we require. In any case, however, it appears that when the frequency of breaks is small, and the number of joints is large, as we know to be the case in the chromosome (if the linear theory holds), the difference in the results due to different orders of determination would be so minute as to be entirely negligible. A number of typical particular diverse orders worked out gave results differing only in the fourth or fifth decimal place or beyond.³

³ If it be held that the first break occurs as frequently at any one joint as at any other, and that any given order of determining the fate of the entire n joints occurs as frequently as any other; then there would be $n!$ equally frequent orders of determination. By transforming, in accordance with these principles, the binomial $(a+b)^n$ into an expression giving the mean result for all these $n!$ orders of determination, in case of interference extending for any given number of joints, a solution under these conditions would be obtained. Apparently such a solution would not differ in numerical results appreciably from that given in the text.

We assume then that the determination of whether each joint is to show a break or no break occurs in serial order, from one extremity of the chromosome to the other. Thus in the series

1 2 3 4 5 6 7 8

the fate of number 1 is first determined, then that of number 2, and so on. (Determination proceeding serially from the middle in both directions, as might occur in V shaped chromosomes, would give essentially identical results.)

We shall, as in the case of no interference, first develop formulae in which b , the frequency of a break at any given joint, and n , the number of joints, are assumed to be known; from these we shall then obtain general formulae showing the relations even when b and n are not known.

It is convenient to designate as k the number of joints whose fate is determined (in one direction, of course, if the order of determination is serial), when a break occurs at a given joint, including in k the joint at which a break occurs. Thus, if, when a break occurs, this prevents a break in the four succeeding joints, the value of k is 5.

The problem of the results of interference in a series of n joints, in case n is greater than the distance of interference, may be attacked directly, in the following manner.

The chromosome has n joints between its $n+1$ genes. At each joint, so long as interference does not come into operation, a break occurs in the fraction b of all cases; no-break in the fraction $a (= 1-b)$. Whenever the fate of a certain joint is determined as no-break, this has no effect on the fate of succeeding joints. But when a certain joint receives a break, then a certain number ($k-1$) of the succeeding joints are determined as yielding no break, so that at any break a total of k joints is determined.

For convenience of illustration, we will designate a break by the letter b ; no-break (when independently determined) by the letter a ; these will also represent the proportions of cases in which these conditions occur. We may further designate a "no-break" which is determined by a foregoing break, by the Greek letter α . If then, for example, $n=8$, $k=4$, and the first break occurs at the third joint, while no other occurs, we represent the condition as follows

$a a b \alpha a a a a$

And the proportion of cases in which such a result will occur is given by the product of the a 's and the b 's; so that in this case the proportion is a^4b (that is, in .06561 of all cases, if $a=.9$, $b=.1$). The α 's do not have to be included in the product, since they occur in all cases after a b has

occurred at the given point; the value of each a as a factor would be 1, thus not altering the result.

Now, the problem which we have to solve is this: in what proportion of the cases would no b , $1b$, $2b$, etc., occur in n joints, if a b , on its occurrence, is inevitably followed by $(k-1)$ a 's (no-break)? For convenience, let us represent, as before, the proportions having no breaks, 1 break, etc., by B_0 , B_1 , B_2 ; or in general, let B_d represent the proportion of the chromosomes showing (within the stretch of n genes) any required number of breaks d .

We must distinguish two cases for the occurrence of the b 's followed by a 's:

(1) If a break b occurs at a sufficient distance from the (latest determined) end of the chromosome, it is followed by $(k-1)$ a 's. Thus there occur blocks of k joints, in all of which joints the fate is determined by the occurrence of b . For example, if k is 4, these blocks are each $b a a a$. The chromosome would show such combinations as:

$b a a a a a a a$ or $a a a a b a a a a$ or $a a a a b a a a$ or $b a a a b a a a$, etc.

(2) But if a break b occurs very near the (latest determined) end of the chromosome, it cannot be followed by $(k-1)$ a 's, but by some lesser number (from 0 to $k-2$). We shall then (if $k=4$) have such combinations as:

$a a a a a a a b$ or $a a a a a b a a$ or $b a a a a a b a$, and the like.

We will distinguish the former case as that in which all the k -blocks are complete, while in the second case some of the k -blocks are incomplete. These two cases must be dealt with separately.

(1) *All k -blocks complete.* If the number of joints n is large in proportion to k , (the distance over which interference occurs), this first group will contain by far the larger proportion of the crossovers; and this is the situation that we find in the chromosomes. The proportion of these having any given number d of breaks is determined as follows. There will be d complete k -blocks, such as $b a a a$. These will occupy of course dk of the n joints, and there will of course be left $n-dk$ joints to be occupied by the independently determined a 's, as in the following 13 joints:

$a b a a a a b a a a a a$

In such a set of n joints, there are obviously d b 's and $(n-dk)$ independent a 's, so that the frequency of each is $a^{(n-dk)}b^d$. But the k -blocks can occupy different positions, and the frequency of d complete k -blocks will be given by determining the number of diverse positions in which d such blocks may occur.

We may look upon each k -block as a single unit. Our problem is reduced to general terms by considering these k -blocks as units of one kind; the a 's as units of another kind. Since there are d of the k -blocks, $(n-dk)$ of the a 's, the total units are $n-dk+d$. Our problem then is: How many different arrangements of $n-dk+d$ things can be made, if $n-dk$ are of one kind, d of another kind? The number is of course

$$\frac{(n-dk+d)!}{(n-dk)!d!}$$

(see text-books of probability; e.g., Carpenter's "Choice and chance".)

As each occurs with the frequency $a^{(n-dk)}b^d$, the total proportion of cases with d complete k -sets is

$$\frac{(n-dk+d)! a^{(n-dk)}b^d}{(n-dk)!d!}$$

(2) *Some of the k -blocks not complete* (since a b occurs within the distance $(k-1)$ of the termination). It is obvious that there will be $k-1$ possible cases in which a b occurs within the distance $k-1$ of the end; these will be followed by a 's in numbers ranging from 0 to $k-2$. To determine the proportions of these in which there are d breaks, we may deal first with the set in which the final b is just $k-1$ joints from the end, so as to be followed by $(k-2)$ a 's. Then of the n joints, $(d-1)k$ will be taken up by complete k -blocks, while $k-1$ are occupied by the terminal incomplete block, making $dk-1$ that are occupied by the k -blocks. Thus there are left $(n-dk+1)$ independently determined a 's. The incomplete k -block is fixed at the end of the chromosome, so that we have $(d-1)$ blocks of k each, that can be shifted about among the $(n-dk+1)$ a 's. The total number of units to be considered is therefore again $n-dk+d$, and we require the number of diverse arrangements that can be made of these if $n-dk+1$ are of one kind; $d-1$ of another. This (with the total frequency, in view of the number of a 's) is:

$$\frac{(n-dk+d)! a^{(n-dk+1)}b^d}{(n-dk+1)!(d-1)!}$$

Similar reasoning applies to the remaining ones of the $(k-1)$ sets in which b is less than k joints from the end. If we work these out, and add them to those already obtained, we obtain the complete formula for the frequency of any specified number of breaks, B_d , in case of interference, with the order of determination serial. This formula, fundamental for crossing over with interference, is given below:

Let n = the total number of joints dealt with.

k = the number of joints in one direction over which interference extends, including in k the joint at which the break occurs.

b = the frequency of a break at any one joint, when the conditions for interference are not present.

a = the frequency of no-break at any one joint when the conditions for interference are not present (so that $a = 1 - b$).

d = the number of breaks of which the relative frequency is required.

B_d = the frequency of just d breaks, in case there is interference extending over k joints.

Then the value of B_d is found to be made up of b^d times the sum of k terms, as follows:

$$B_d = b^d \left[\frac{(n - (d - 1)k + d - 2)!}{(n - (d - 1)k - 1)! (d - 1)!} a^{(n - (d - 1)k - 1)} + \frac{(n - (d - 1)k + d - 3)!}{(n - (d - 1)k - 2)! (d - 1)!} a^{(n - (d - 1)k - 2)} + \frac{(n - (d - 1)k + d - 4)!}{(n - (d - 1)k - 3)! (d - 1)!} a^{(n - (d - 1)k - 3)} + \dots \right]$$

..... to $(k - 1)$ terms, the $(k - 1)$ th term being:

$$\frac{(n - dk + d)!}{(n - dk + 1)! (d - 1)!} a^{(n - dk + 1)} + \frac{(n - dk + d)!}{(n - dk)! d!} a^{(n - dk)} \dots \dots \dots (26)$$

It will be observed that, of the terms within the brackets, for the first $(k - 1)$ terms the numerator (factorial) of the fraction and the first term of the denominator (factorial) are decreased, from the first term on, successively by 1, while the exponent of a is likewise diminished throughout successively by 1. In case n is small in proportion to k , not all the terms will be realized; as soon as a term is reached in which the numerator or first term in the denominator becomes a minus quantity, that and all succeeding terms are to be omitted. It is important to remember, however, that when the numerator or a term of the denominator becomes 0!, this has the value of 1 (since in general $0! = 1$).

The last term of the formula represents the frequency of the complete k -blocks; the other terms the frequencies of the incomplete k -blocks at the terminus of the chromosome (or of the part of the chromosome considered).

It may be observed that this formula (26) is an entirely general formula for the successive terms when in a series of things the occurrence of a

certain chance result for one unit determines also its results for a number of the contiguous succeeding units, the order of determination being serial. If we let $k = 1$, so that no other units are determined, we have the case of no interference, and formula (26) yields the usual binomial expression for the expansion of $(a + b)^n$.

For B_0 and B_1 (the proportion of cases having no break, and 1 break, respectively), formula (26) simplifies directly into the following expressions:

$$B_0 = a^n \dots \dots \dots (27)$$

$$B_1 = a^{(n-k)}b(1 + a + a^2 \dots \dots \dots a^{(k-1)}) + (n - k) a^{(n-k)}b \dots \dots \dots (28)$$

Since $b = 1 - a$, the first member of this expression for B_1 is equal to $a^{(n-k)}(1 - a^k)$ (see text-books of algebra, e.g., HAWKES'S Advanced (1905), page 22). Substituting this we have for B_1 the more convenient expression:

$$B_1 = a^{(n-k)}(1 - a^k) + (n - k) a^{(n-k)}b \dots \dots \dots (29)$$

Formulae (26) to (29) give the proportions for the different possible numbers of breaks between two genes, in terms of n , k and b . If these three values are known, the crossover ratio C can be discovered by their aid, since it will be the sum of all the cases in which there are odd numbers of breaks between the two genes separated by n joints; that is, it will be the sum of B_1, B_3, B_5 , etc., up to $\frac{n}{k}$ (plus 1 for any remainder above an integral quotient).

But for the conditions in inheritance we know neither n , k nor b , so that it is desirable to obtain formulae for B_0, B_1, B_2 , etc., and for the crossover ratio C , that shall be independent of the precise values of n , k and b , depending merely upon certain general limits to be assigned them, as we did for the case of no interference.

In this case as in that of no interference, we know that the frequency of a break at a single joint is less than .005.

With regard to k , we know, not the actual number of joints over which interference extends, but the extent of the chromosome, measured in percentage of crossing over, over which it extends. According to BRIDGES (1921), in chromosome I of *Drosophila*, interference is practically complete over a distance which itself yields about 15 percent of crossing over; in chromosome II near the ends this distance gives about 20 percent of crossing over, while in the middle region it is about 10 percent.

For our purposes, consider the case that two genes are separated by just k joints. In this case there can be but 1 break between them. The frequency of no break whatever between them will of course be a^k (since

$n = k$), so that the frequency of a single break will be $1 - a^k$ (since all cases that do not show no-break must show a single break).

Now, suppose that we know the percentage of crossing over yielded by the distance over which interference occurs (as we approximately do in *Drosophila*). This percentage of crossing over is due to a single break (since only 1 break can occur in that region of interference). It is therefore equal to $1 - a^k$. Let this percentage be called P , and for convenience let $1 - P$ be called A . Since $P = 1 - a^k$, A or $1 - P = a^k$. Thus, if P be .20 (as in the second chromosome of *Drosophila*), then $A = .80$, and this we know to be equal to a^k .

Now, if $A = a^k$, of course $a = \sqrt[k]{A}$, and since $b = 1 - a$, $b = 1 - \sqrt[k]{A}$. We now have the values of a and b transmuted into terms, one of which at least (A) is approximately known, experimentally. If now we can in some way replace k and n by known quantities, we shall be on the way to a solution of our problem.

Consider the case that n , although unknown, is some definite number of times k . Thus, if k be 20 units, each of which gives 1 percent crossing over, we may readily deal with a distance of 25 or 30 or 40 or 100 of such units; that is, distances in which $n = 1.25 k$, $1.5 k$, $2 k$ or $5 k$. This is of course done in practice in *Drosophila*. In general, let $n = mk$.

Now substitute these values in formula (26). That is, for a substitute $\sqrt[k]{A}$, or $A^{1/k}$; for b , substitute $1 - \sqrt[k]{A}$; for n , substitute mk .

We may first undertake this for the values of B_0 and B_1 , in formulae (27) and (29).

From formula (27), we obtain

$$B_0 = (\sqrt[k]{A})^{mk}, \text{ which of course yields directly} \\ B_0 = A^m \dots \dots \dots (30)$$

Making the proposed substitutions in formula (29), for B_1 , we obtain directly (with a slight change in the arrangement of the factors of the second member)

$$B_1 = A^{(m-1)}(1 - A) + (m - 1)A^{(m-1)} \cdot k(1 - \sqrt[k]{A})$$

In the first member of this expression, $1 - A$ is of course P , the extent of interference. In the second member we find the expression $k(1 - \sqrt[k]{A})$, and this we recognize again (as on page 415) to approach, as k becomes large, the limit given by the minus natural logarithm of A . Substituting these, we obtain the following formula (31) for B_1 , depending entirely on the observationally knowable quantities A , P and m :

$$B_1 = A^{(m-1)} \cdot P + (m - 1)A^{(m-1)} \cdot (-\text{nat log } A) \dots \dots \dots (31)$$

For example, suppose that interference extends over a distance that yields 25 percent crossing over, what is the crossover ratio between two

genes lying twice this distance apart? Here the total crossover ratio will be due to the value of B_1 . We have $P = .25$, $A = .75$, $m = 2$. Hence: $C = B_1 = .75 \times .25 + .75 \times (-\text{nat log } .75) = 40.33$ percent.

In the case of the general formula (26), for the value of B_d (that is, for the proportion giving any designated number of breaks), the conditions are of course more complex. It will be observed that the expression within the parenthesis in formula (26) consists of k terms, each containing a coefficient composed of factorials, and a certain power of a , so that the form of each term is

$\frac{p!}{(p-q)! q!} a^r$. Now, such a factorial expression is a definite function of p and q , which can be written in the simpler form, more convenient for our purposes:

$\frac{p^{!q}}{q!} a^r$, in which $p^{!q}$ signifies q factors, beginning with p , and descending by successive differences of 1.

We may then write the terms of formula (26) in this manner, at the same time making the substitutions proposed. For reasons that will appear later, we also multiply the numerator and denominator by k^d ; this of course does not alter its value. We thus obtain:

$$\begin{aligned}
 B_d = & \frac{k^d(1 - \sqrt[k]{A})^d}{k^d} \left[\frac{((m-d+1)k+d-2)^{!(d-1)}}{(d-1)!} A^{(m-d+1-1/k)} \right. \\
 & + \frac{((m-d+1)k+d-3)^{!(d-1)}}{(d-1)!} A^{(m-d+1-2/k)} \text{ etc., to } k-1 \text{ terms;} \\
 & \left. \text{then } + \frac{((m-d)k+d)^{!d}}{d!} A^{(m-d)} \right] \dots\dots\dots (32)
 \end{aligned}$$

This formula can be simplified for practical working purposes, and the unknown quantity k gotten rid of, without sensibly altering the numerical results which it yields, by the following series of considerations, and changes based on them:

(1) The numerator of the first term of the expression for B_d is the d th power of $k(1 - \sqrt[k]{A})$. In this latter again we find an expression the limit of whose value, as k is increased, is the minus natural logarithm of A (see page 415). We know that k is a large number; if b is taken as .0025 (which is beyond doubt an extreme upper limit for it), to yield 20 percent interference k would have to be 89+, while if b is less than this, k would be still greater. We may therefore replace the numerator of the first

member of the expression (32) for B_d by the d th power of the minus natural logarithm of A .

The quantities within the large brackets that enclose the second member of the expression for B_d (formula (32)) are a series of k terms, of powers of A ,—the coefficients and exponents of A changing according to definite laws. The exponents and coefficients must be dealt with separately.

(2) For the first $k-1$ terms, the exponents of A form a series decreasing successively by a difference of the very small quantity $\frac{1}{k}$, and the total difference between the first and last term of these $k-1$ terms is 1. (The first exponent is $m-d+1-\frac{1}{k}$; the last is $m-d+1-\frac{k-1}{k}$). As the difference between the extremes of this long series of slightly differing powers of A is so slight, it will make little difference in the total value of the expression if we substitute for each exponent the mean of all, which is $\frac{2m-2d+1}{2}$.

(Whether this substitution does make an appreciable difference will of course be tested later.)

(3) We have thus brought all the terms within the brackets except the last one to a common power of A . If now we can add together their coefficients, we shall obtain a single expression for the total value of these $k-1$ terms. The coefficients are, for successive decreasing numbers, the

functions represented by the form $\frac{p^{1q}}{q!}$. The first coefficient of formula (32) is $\frac{((m-d+1)k+d-2)^{(d-1)}}{(d-1)!}$ while for the $(k-1)$ th term it is

$\frac{((m-d)k+d)^{(d-1)}}{(d-1)!}$. Now, the sum of a series of successive terms of the form

$\frac{p^{1q}}{q!}$, beginning with a value of 1 (in which case p and q are equal), so that

the series runs $\frac{q^{1q}}{q!} + \frac{(q+1)^{1q}}{q!} + \frac{(q+2)^{1q}}{q!}$ up to $\frac{p^{1q}}{q!}$, is $\frac{(p+1)^{(q+1)}}{(q+1)!}$. Re-

placing p and q by their values in the coefficients of formula (32), we find that the total sum of all the coefficients, from the value 1, up to the highest value given in the first term, is $\frac{((m-d+1)k+d-1)^d}{d!}$.

But the lowest of these $(k-1)$ terms does not have the value 1, but instead the value $\frac{((m-d)k+d)^{(d-1)}}{(d-1)!}$. Hence, from the total sum, just

given, must be subtracted that part of the sum given by the series from the value 1 up to $\frac{((m-d)k+d-1)^{(d-1)}}{(d-1)!}$. This part will, according to the rule just given, be $\frac{((m-d)k+d)^{1d}}{d!}$. The actual sum of the $(k-1)$ coefficients of $A^{\frac{2m-2d+1}{2}}$ is therefore:

$\frac{((m-d+1)k+d-1)^{1d}}{d!} - \frac{((m-d)k+d)^{1d}}{d!}$; so that the total value of the first $(k-1)$ terms within the large brackets is

$$\left[\frac{((m-d+1)k+d-1)^{1d}}{d!} - \frac{((m-d)k+d)^{1d}}{d!} \right] A^{\frac{2m-2d+1}{2}}$$

There remains the k th or last term of the expression in the brackets of (32): this is $\frac{((m-d)k+d)^{1d}}{d!} A^{(m-d)}$. We add this, and at the same time we divide each term within the large brackets by k^d , thus removing k^d from the quantity outside the brackets. This gives us as the present condition of the formula for B_d the following:

$$B_d = (-\text{nat log } A)^d \left[\left(\frac{((m-d+1)k+d-1)^{1d}}{d! k^d} - \frac{((m-d)k+d)^{1d}}{d! k^d} \right) A^{\frac{(2m-2d+1)}{2}} + \frac{((m-d)k+d)^{1d}}{d! k^d} A^{(m-d)} \right]$$

From this we can obtain an approximate formula that shall not include the unknown quantity k , and that will give results differing from the complete formula by only a negligible amount (as will be shown). This results from the following considerations:

The first coefficient $\frac{((m-d+1)k+d-1)^{1d}}{d! k^d}$ as k becomes larger approaches $\frac{(m-d+1)^d}{d!}$ as a limit. Thus, for the values $m=4$ and $d=2$ the value of this limiting expression $\frac{(m-d+1)^d}{d!}$ is 4.5. If now we make $k=40$, the value of $\frac{((m-d+1)k+d-1)^{1d}}{d! k^d}$ is 4.5375; if k is 100, its value is 4.515. Similarly the second and third coefficients within the parenthesis each approaches $\frac{(m-d)^d}{d}$ as k becomes large. We may therefore substitute

these values for the three coefficients. We then obtain as our final approximation formula for B_d the following which is to be employed in computing the crossover ratio, C , in case there is interference:

$$B_d = (-\text{nat log } A)^d \left[\frac{(m-d+1)^d - (m-d)^d}{d!} A^{\frac{2m-2d+1}{2}} + \frac{(m-d)^d}{d!} A^{(m-d)} \right] \dots \dots \dots (33)$$

In formula (33):

d = the number of breaks between the two genes in question.

B_d = the proportion of cases having any number d of breaks.

$A = 1 - P$, where P is the average number of standard units of length (each unit yielding 1 percent of crossing over), over which interference extends.

m = the distance between the two genes for which the crossover ratio is sought, in units, each of which is P . Thus, if the distance between the two genes is 1.25 times the mean distance over which interference extends, $m = 1.25$.

C , the crossover ratio, is obtained from this formula (33) by taking the sum of the values of B_d for odd values of d . As it turns out, usually, the only odd values of d that have an appreciable value are 1 and 3, so that as a rule $C = B_1 + B_3$. The value of B_1 is most readily obtained from formula (31); that of B_3 from (33).

It is to be noted that there may be d breaks and so a real value for B_d , only if m is greater than $d - 1$. Thus if the chromosome has length just twice that over which interference extends, it is not possible that there should be three breaks within that distance. But if it extends beyond twice the interference distance, then there can be three breaks. If now m is greater than $d - 1$, but less than d , a number of terms in the above expression (33) disappear, and the expression for B_d becomes simplified. It is perhaps unnecessary to give in detail the steps of that simplification; it will be found that in this case the formula takes the form given in the following (34); that is:

When the value of m lies between d and $d - 1$,

$$B_d = (-\text{nat log } A)^d \left(\frac{(m-d+1)^d}{d!} A^{\frac{m-d+1}{2}} \right) \dots \dots \dots (34)$$

The justification for making the above approximations of course requires test. Such test was fully supplied in the course of this investigation, through the fact that, as may be supposed, these changes,

approximations and simplifications, did not suggest themselves at first, but came after much labor and many divagations. At first it appeared necessary to discover the effects of interference by employing the fundamental formula for B_d , in terms of n , k and b , given in formula (26). A maximum value was assigned to b , using for this .005, since as before remarked, b is certainly less than this. Then if we assume a certain extent of interference P , we can determine the corresponding value of k through the relation $P = 1 - a^k$, set forth on page 433; from this it follows that

$$k = \frac{\log(1-P)}{\log(1-b)}, \text{ or } = \frac{\log A}{\log a}$$

When b is put at .005, so that $a = .995$, we find that for interference extending over a distance yielding 20 percent crossing over, $k = 45$; for 25 percent, $k = 58$; for 30 percent, $k = 72$, and so on. Taking now values of n that are a certain number of times k , we work out the crossover ratios by adding the values of the odd numbers of breaks, $B_1 + B_3 + B_5$, etc. The value of B_1 is obtained from formula (29); those for B_3 , B_5 , etc., are gotten with great labor from formula (26). It is found that the results do not differ sensibly from those given for the same extent of interference and same proportion of n to k by our simple general approximation formula (33). The following are typical comparative results of the two methods, for certain values of m , when the interference extends over a distance yielding a crossover ratio of 25 percent.

TABLE 7
Crossover ratio, percent.

m	COMPLETE FORMULA (26) ASSUMING $b = .005$	APPROXIMATION FORMULA (33)
1	25.00	25.00
1.5	34.08	34.11
2	40.27	40.32
2.5	44.22	44.26
3	46.69	46.77
4	49.02	49.06
5	49.49	49.78

As table 7 shows, the approximation formula (33) is quite accurate enough for all purposes, even for so high a value of b as .005; for the actually smaller values it is of course still more accurate. It even gives results accurate to the third decimal place for values of B_d when there is

no interference. In this case the value of k is 1, and $m=n$, while $P=b$ and $A=a$. Letting $b=.005$, we find by (33) that $B_3=.012007$, while by the simple binomial formula it is .012430.

Throughout the remainder of this paper, therefore we employ the results yielded by formula (33) and its modifications. Thus, the crossover ratio C in case of interference is:

$$C = B_1 + B_3 + B_5, \text{ etc.} \dots \dots \dots (35)$$

The values of B_1, B_3, B_5 , etc. are of course found by (31), (33) and (34).

The general relations in crossing over with interference

In formula (33) with its modifications, formulae (30), (31) and (34), we have an instrument by which we can determine the crossover ratios for any degree of interference, with the rules and general relations of crossing over when there is interference, discovering whether interference alters any of the general relations found to hold for linear crossing over without interference; how far there is agreement with our 17 basic propositions; and how far the conditions resulting agree with those found in nature.

1. We may first examine a table of crossover ratios between genes ten units apart, with a degree of interference comparable to that held to be found in *Drosophila*. Selecting for this purpose interference extending over a distance that itself yields 20 percent crossing over, we obtain table 8.

TABLE 8

Crossover ratios for genes A to L, if they are separated by successive stretches each 10 units in length, and there is interference extending for 20 units (worked out from formula (33)). For comparison with the results when there is no interference (table 5), and with the conditions in nature (table 2).

	A	B	C	D	E	F	G	H	I	J	K	L
A	10.0	20.0	27.9	33.9	38.3	41.5	43.8	45.6	46.8	47.7	48.4
B	10.0	10.0	20.0	27.9	33.9	38.3	41.5	43.8	45.6	46.8	47.7
C	20.0	10.0	10.0	20.0	27.9	33.9	38.3	41.5	43.8	45.6	46.8
D	27.9	20.0	10.0	10.0	20.0	27.9	33.9	38.3	41.5	43.8	45.6
E	33.9	27.9	20.0	10.0	10.0	20.0	27.9	33.9	38.3	41.5	43.8
F	38.3	33.9	27.9	20.0	10.0	10.0	20.0	27.9	33.9	38.3	41.5
G	41.5	38.3	33.9	27.9	20.0	10.0	10.0	20.0	27.9	33.9	38.3
H	43.8	41.5	38.3	33.9	27.9	20.0	10.0	10.0	20.0	27.9	33.9
I	45.6	43.8	41.5	38.3	33.9	27.9	20.0	10.0	10.0	20.0	27.9
J	46.8	45.6	43.8	41.5	38.3	33.9	27.9	20.0	10.0	10.0	20.0
K	47.7	46.8	45.6	43.8	41.5	38.3	33.9	27.9	20.0	10.0	10.0
L	48.4	47.7	46.8	45.6	43.8	41.5	38.3	33.9	27.9	20.0	10.0

A study of table 8 in connection with the fundamental assumptions as to the serial order of the genes, shows that interference extending for 20 units does not alter any of the general relations already found to hold for crossing over without interference, except of course our basic proposition 17, which it satisfies. That is, linear crossing over with interference of about 20 units requires and accounts for all of the general relations set forth in our 17 basic propositions of pages 397-399. To this very remarkable fact we return later.

2. For full understanding of the relations just mentioned and of those to follow, it is necessary to note the effect of interference on the actual length of the standard "units" in which the chromosome is measured, and on certain other quantities. A unit is the length (or number of joints) necessary to give one percent of crossing over. Assuming that before the conditions of interference arise, any given joint shows a break on the average in a certain proportion b of cases, then with the same value of b , *as interference increases, the length of the unit increases.*

With a given value of b , if k is the number of joints over which interference extends, then for genes separated by just k joints the crossover ratio C is given by the simple formula $C = 1 - a^k$ (where $a = 1 - b$). (This is derived directly from formula (29), replacing n by k , and remembering that in this case $C = B_1$.) From this again we can of course obtain the value of k for any extent of interference P , by replacing C by P in the above formula, and solving for k . This yields:

$$k = \frac{\log(1-P)}{\log(1-b)}$$

The number of joints in a unit is then given by dividing the value of k by P (considering the latter as integral percents). An example will make this clear. Let the frequency b of a break at any joint be .002, so that $1 - b = .998$. It is required to know how many joints there are in k if interference extends over a distance that yields 20 percent crossing over.

Here $C = P = .20$, so that $k = \frac{\log .8}{\log .998} = 111.455$.

If 111.455 joints are required to give 20 percent crossing over, to give 1 percent 5.573 are required (since within the stretch k the values are simply additive), so that a unit is 5.573 joints.

Working out the value of a unit for a number of different extents of interference by this method, we obtain table 9.

3. In case there is interference, it is necessary to distinguish carefully between the frequency of a break at any joint before the conditions for

interference have arisen (this frequency we have called b), and the actual average frequency of a break at any joint, in view of the fact that after one break has occurred interference does arise. This actual average frequency of a break at any joint, in case of interference, we may designate β . With constant b , the value of β decreases as interference increases. This may be obtained for the distance k , over which interference extends, by dividing the value P ($=C$ for k joints) by k . The value of β for different extents of interference when $b=.002$, is given in the third row of table 9.

TABLE 9

Number of joints in a single unit giving 1 percent crossing over, for certain different values of interference, in case b , the frequency of a break at any joint before the conditions for interference arise, is .002.

Extent of interference, percent	0	20	25	30	35	40	50	75	90
Number of joints for 1 unit	5.04	5.57	5.75	5.94	6.15	6.38	6.94	9.23	12.78
Average frequency of a break at any joint ($=\beta$).	.002	.00179	.00174	.00168	.00163	.00157	.00144	.00108	.00078

4. If interference extends for not more than 30 units, the crossover ratio satisfies the first of our basic propositions (page 397),—namely, that the ratio rises to about 50 percent, but does not go appreciably above 50 percent.

5. But if interference extends for a distance greater than 30 units, then as the distance between the genes becomes greater, the ratio rises above 50 percent. At interference of 30 units it rises theoretically to a maximum of 50.6 percent (at a length of 120 units); for still higher values of interference it rises still higher, as shown in table 10. If interference is sufficiently extensive it may rise to any value less than 100 percent.

It is to be kept in mind, of course, that for the different grades of interference the units are of diverse lengths (table 9), so that the third column of table 10 does not give the relative actual lengths at which the maximum ratio is reached. These may be obtained by multiplying the values in the third column of table 10 by the corresponding lengths of the units in the second row of table 9. If we adopt the lowest of these values as a unit, the relative lengths are shown in the fourth column of table 10.

6. When as a result of interference the maximum crossover ratio rises above 50 percent, then as the distance between the genes is still farther increased, the ratio undergoes oscillatory changes illustrated in the curves of figure 1. Beyond the maximum the ratio begins to decrease, falling to 50 percent, then below 50 percent. The amount by which it thus goes below 50 percent is somewhat less than the amount by which, at its highest, it exceeded 50 percent. As the distance between the genes is

TABLE 10

Limiting values of the crossover ratio, with different extents of interference (measured in units each of which gives 1 percent crossing over); with the number of units distance at which the ratio is at the maximum.

EXTENT OF INTERFERENCE UNITS	LIMITING CROSS-OVER RATIO IN PERCENT	APPROXIMATE DISTANCE AT WHICH THE MAXIMUM VALUE IS ATTAINED, IN UNITS	APPROXIMATE RELATIVE ACTUAL LENGTHS AT WHICH MAXIMUM IS ATTAINED
0	50
20	50
25	50	175	1.75
30	50.6	120	1.24
35	52.4	95	1.02
40	54.9	90	1.00
50	60.6	85	1.03
75	77.9	90	1.45
90	90.0	90	2.00

still farther increased it turns again and climbs once more above 50 percent, but does not again reach the maximum previously attained. It then falls again, and continues to oscillate about 50 percent, the divagations above and below that value becoming less and less, until it finally settles down at approximately 50 percent, the oscillations becoming imperceptible. As the curves of figure 1 show, the dips below 50 percent are, in the case of high interference, often less in amount than the later rise above 50 percent, so that the oscillations above and below the line of 50 percent are not symmetrical.

These oscillations in the value of the crossover ratio are due to the fact that odd numbers of breaks give crossing over, while even ones do not. At the maximum value of the ratio the prevailing number of breaks is 1; this then decreases, while the proportion with 2 breaks increases, thus decreasing the ratio; and so on. For example, when the interference is 75 percent, the maximum value of the ratio (77.9 percent), is likewise the percentage with a single break, while now only 3.2 percent have two

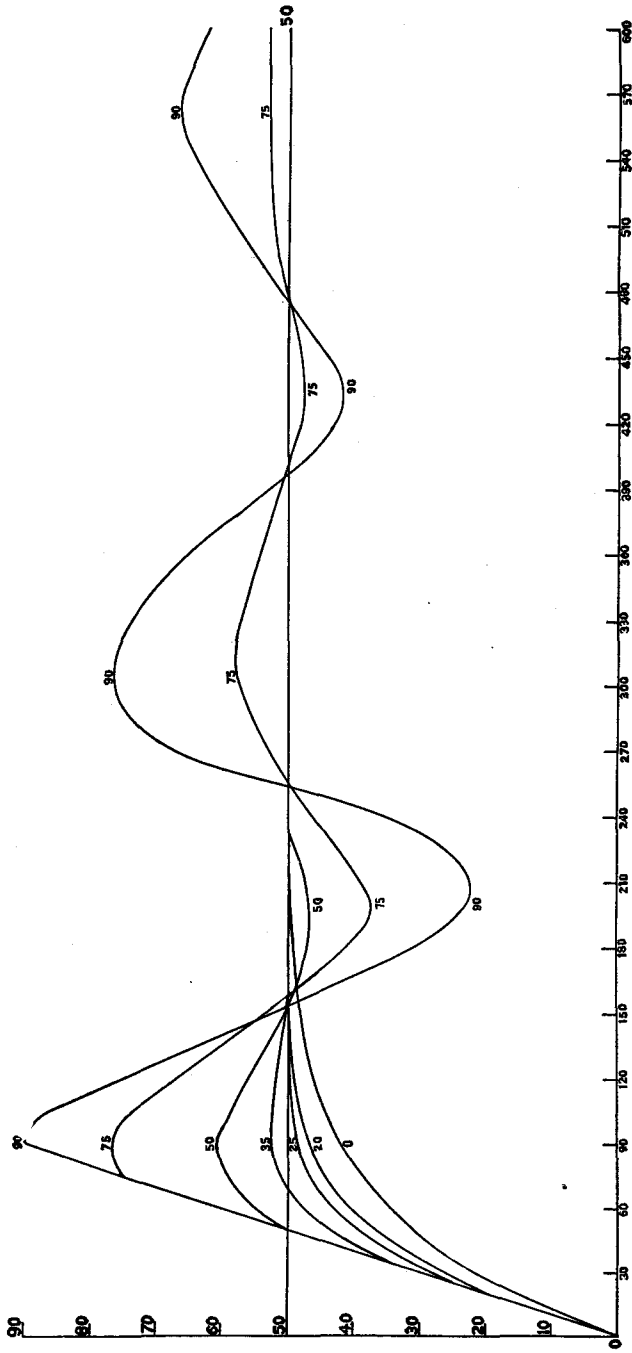


FIGURE 1.—Curves showing for a number of different amounts of interference the changes in the crossover ratios as the distance between the genes increases. The vertical scale gives the crossover ratios, in percentages. The horizontal scale gives the distances apart of the genes, in standard units (each unit being a distance sufficient to give 1 percent of crossing over). The figures on the curves show the amount of interference represented by that curve. The curves for no interference and for interference extending 20 and 25 units do not rise above 50 percent; those for 35, 50, 75 and 90 units rise above 50 percent, descend as the distance between the genes becomes greater, and oscillate below and above 50 percent, as shown.

breaks. At the minimum value (38.8 percent), there are but 29.4 percent with 1 break and 9.4 percent with 3 breaks, while 58.8 percent have two breaks (the remainder, 2.4 percent having no breaks).

Figure 1 shows sketches of comparative curves of the crossover ratios for different amounts of interference, from 0 to 90 percent. The lengths are measured in the standard units of length yielding 1 percent crossing over. It is to be remembered of course that these units are of different lengths for different degrees of interference, as shown in table 8.

Examination of the curves of figure 1 shows that for different degrees of interference the curves diverge more and more widely until the genes are about ninety units apart; here the divergence between the different curves is at a maximum. Then the curves approach, and for genes distant apart about 150 units the crossover ratios are nearly the same (in the neighborhood of 50 percent) for all degrees of interference. For still greater distances the divergences increase and decrease in an oscillating manner.

7. It is obvious that for high extents of interference, in which the crossover ratio goes above 50 percent, the serial order of the genes would not be given by the increase of the crossover ratio measured from the end gene. Genes farther from the end gene would often have a lower ratio than those nearer. Genes near the end gene would give with it a low ratio, this gradually increasing with the distance, to above 50 percent, then decreasing with greater distance, again increasing, and so on. A chromosome map constructed, like those in use in *Drosophila*, on the basis that increased crossover ratios meant greater distances would give incorrect and inconsistent results, the genes seeming to be in different orders when different basic genes were used from which to measure the ratios of the rest.

Even in such cases, however, a map giving the genes in correct order could be obtained by considering successively genes having low crossover ratios (say 10 percent or less). The order of 3 genes whose ratios all fall within 10 to 20 percent would be first determined, then the order of two of these with a fourth that was near; by a concatenation of such successive determinations, in the manner now practically employed by the students of *Drosophila*, the correct order would finally be obtained for the entire chromosome.

None of these difficulties or inconsistencies are normally met in the case of *Drosophila*, so far as present knowledge goes; since interference does not extend so far as to give crossover ratios above 50 percent. A case giving results akin to these was however described briefly by MULLER at the meeting of the AMERICAN SOCIETY OF ZOÖLOGISTS in December, 1921.

In a particular instance a gene was present which interfered with a break in the chromosome for a longer distance than usual; for genes more than twice that distance apart two breaks were more frequent than 1, so that the crossover ratio became less for genes farther apart than for those nearer together.

At the same meeting IBSEN (1922) described in the guinea-pig certain genetic results that could be explained by assuming that 75 percent crossing over had occurred in parents of both sexes. This possibility had not been subjected to a critical test. If substantiated, it would follow that in the guinea-pig the relative frequency of crossing over between genes would not indicate the serial order of the genes in the chromosome: this serial order would be determined only by dealing successively with genes having low crossover ratios.

8. In comparing the computed relations with those found in experimental breeding, it is to be noted that our curves and computations are for certain definite *average* extents of interference. The general results would be similar whether this average represented a single invariable distance over which interference extends, so that no second break ever occurred within that distance, while beyond that distance there was no interference; or whether it is an average of the different distances over which interference extends in different cases. In *Drosophila* it is clear from the breeding data that the distance over which interference acts is not invariable; at certain distances from a given break another break is merely less frequent than would be the case if the given break did not exist. The distance over which interference extends is not given by the students of *Drosophila* as an average, but simply as the distance over which interference is practically complete. It appears clear that the average distance over which interference extends would be somewhat greater than this distance over which interference is practically complete.

According to BRIDGES (1921, p. 127), this distance over which interference is practically complete is (from the results of experimental breeding) in the first or X chromosome of *Drosophila* about 15 units or 15 percent; for the second and third chromosomes of *Drosophila* it is about 20 percent in the end regions, 10 percent in the middle regions. These would correspond to averages somewhat higher, perhaps from 20 to 25 percent. In none of these cases therefore should there occur crossover ratios above 50 percent, and according to BRIDGES (1921, p. 128) no ratios above 50 percent have yet been met in *Drosophila*. The serial order of the genes would be indicated in all such cases by the increasing crossover ratios with the end gene.

9. On page 422 it was shown that there are four serious discrepancies between the crossover relations found in nature and those which must occur if crossing over occurs as the linear theory holds, but without interference. One of these discrepancies consisted of the very facts for explaining which the hypothesis of interference was introduced. The three others were numerical discrepancies, not held in view (nor even known) when the hypothesis of interference was proposed. It remains to determine whether these discrepancies would be removed if interference occurs. We will deal with the three in succession.

10. The first numerical discrepancy was, that the observed values of the ratios are inconsistent with each other; if crossing over occurs without interference, the observed ratios for distant genes are regularly greater than they should be as compared with those for genes close together. That is, if we compute from the shorter stretches of the chromosome the necessary crossover ratios for the longer stretches, made up of the shorter ones, we find that these ratios computed on the basis of no interference are always less than those actually observed (see page 419).

Examining the results of interference, we find that this does increase the crossover ratios for distant genes, as compared with those between genes close together, thus causing the required conditions to approach those actually found. For example, if we take the crossover ratio between genes 10 units apart as that for genes "close together," and employ this as unity, then table 11 shows what number of times this ratio is yielded by genes separated by greater distances,—in the case of certain degrees of interference.

TABLE 11

INTERFERENCE	DISTANCES BETWEEN THE GENES, IN UNITS OF 10 PERCENT					
	10	20	30	60	100	160
0	1	1.83	2.48	3.84	4.74	5.25
20	1	2.00	2.79	4.15	4.77	4.97
25	1	2.00	2.90	4.37	4.91	4.99
30	1	2.00	3.00	4.60	5.04	5.03

It will be observed in this table that for distances up to 100 units (where the crossover ratio reaches 47 to 49 percent) the more the interference, the greater is the proportion of the longer stretches as compared to the shorter ones. At very long distances, however, the two become equalized, and finally the proportion is greater for no interference. Most of the crossover ratios usually dealt with fall in the intermediate region.

To determine how closely the theoretically required ratios approach those found in nature, we may compute from the ratios observed for short distances those that are required for longer distances, on the basis of different degrees of interference. For the short distances, we will employ the well-established ratios, each based on more than 5000 cases, that are set forth on page 419. In table 12 are given for the longer stretches: (a) the ratios required if there is no interference; (b) those required by 20 percent interference; (c) those required by 25 percent interference; (d) the observed ratios.

TABLE 12

Comparison of the crossover ratios for the long stretches between the genes named, as computed from the component shorter stretches given on page 419, for different extents of interference (a, b and c), with those observed (d).

Chromosome I

	(a) NO INTERFERENCE	(b) INTERFERENCE 20 PERCENT	(c) INTERFERENCE 25 PERCENT	(d) OBSERVED RATIOS
Yellow-miniature.....	31.6	32.8	33.3	34.3
White-rudimentary.....	38.2	40.0	40.5	42.4
White-bar.....	38.7	40.9	41.6	43.6
Vermilion-bar.....	21.1	23.4	23.9	23.9
Vermilion-fused.....	22.5	25.3	26.2	25.8
Vermilion-forked.....	20.0	23.0	23.5	22.5

Chromosome II

Star-speck.....	47.4	47.9	49.0	48.3
Star-curved.....	43.4	44.1	45.4	45.9
Star-purple.....	39.4	39.8	40.6	43.7
Dachs-vestigial.....	29.3	31.4	32.9	29.6
Black-morula.....	43.8	44.3	44.8	46.6
Purple-speck.....	38.1	39.4	40.9	45.7

Chromosome III

Sepia-rough.....	35.5	39.1	41.2	39.1
Dichaete-rough.....	32.0	34.3	36.0	36.1
Sepia-sooty.....	25.7	29.2	30.5	26.2

Table 12 is made by the aid of extensive and detailed tables computed in the course of this investigation, showing the crossover ratios corresponding to many different numbers of units of length, for nine different extents of interference, ranging from 0 to 90 percent.

As table 12 shows, in all cases, the observed values for the ratios are greater than those permissible in case there were no interference. No exact coincidence of the required with the observed values is, of course, to be expected, owing to the numerous sources of irregularity and inaccuracy in the observed ratios; but the systematic diversity in one uniform direction is significant. With 20 percent interference the required ratios approach more nearly the observed ones, but still remain below them (in all except two of the fifteen cases). With 25 percent interference the required values approach still more closely the observed ones; now in nine of the fifteen cases the required ratios are either practically identical with the observed ones (vermillion-bar, star-curved, dichaete-rough) or are a little above them. It is clear that with the addition of interference of about 20 or 25 percent, our first numerical discrepancy disappears; the observed ratios for genes close together become consistent with those for genes far apart.

11. The second numerical discrepancy between the observed facts and the requirements of the linear theory without interference was that the proportions of the chromosomes showing the different numbers of breaks would not be those actually found. The latter show a marked excess in the proportion having but one break, with a corresponding deficiency of those having two or more breaks (page 421). Interference, we have seen (p. 424), tends to produce just this effect; the question here is: does it produce this effect to just the extent demanded in order to bring the observations and the requirements into unison?

In table 13 are given the proportions for the different numbers of breaks observed in experiments directed upon this point, by MULLER (1916), PLOUGH (1917), WEINSTEIN (1918) and GOWEN (1919), together with the proportions required on the linear theory (a) without interference, (b) with 20 percent interference, and (c) with 25 percent interference. In all cases the observed proportion showing no breaks is considered the fixed point, since this proportion is not altered by interference. From this are computed the required proportions for 1, 2, 3 or 4 breaks. For no interference this computation is made by the formula (20 a), page 421; for the different degrees of interference the computation was made by the use of formulae 30, 31, 33 and 34, in connection with extensive tables made for producing the curves of figure 1.

Table 13 shows that with no interference the proportion with one break is uniformly much too low, while the proportions with 2, 3, 4 breaks are much too high, as compared with the observed proportions. As interference is made greater the proportion with one break gradually

TABLE 13

Comparison of the proportions showing 0, 1, 2, 3 or 4 breaks in chromosomes of *Drosophila*, with the proportions required by the linear theory, (a) when there is no interference; (b) when there is 20 percent interference; (c) when there is 25 percent interference. The data are all for *Drosophila melanogaster*, except the case last given, in which the data are for *Drosophila virilis*.

Number of breaks . .	0	1	2	3	4	AUTHORITY AND CHROMOSOMES	NUMBER OF CASES
Percentages:							
Observed	54.2	41.6	4.3	0	0	MULLER 1916	712
Required:						I. Yellow to bar	
Interference 0		33.1	10.2	2.1	0.3		
Interference 20%		40.0	5.8	0.1	0		
Interference 25%		41.6	4.3	0	0		
Observed	32.6	51.1	15.0	1.3	0	MULLER 1916	460
Required:						II. Star to balloon	
Interference 0		36.6	20.5	7.7	2.2		
Interference 20%		44.7	19.5	3.0	0.1		
Interference 25%		47.2	18.6	1.9	0.02		
Observed	77.3	22.5	0.2	0	0	PLOUGH 1917	5134
Required:						I. Vermilion to forked	
Interference 0		19.9	2.4	0.2	0		
Interference 20%		22.6	0.05	0	0		
Interference 25%		22.7	0	0	0		
Observed	75.2	23.5	1.3	0	0	PLOUGH 1917	17,225
Required:						II. Black to curved	
Interference 0		21.4	3.0	0.3	0		
Interference 20%		24.2	0.6	0	0		
Interference 25%		24.8	0	0	0		
Observed	78.1	21.8	0.1	0	0	WEINSTEIN 1918 (data from BRIDGES)	3394
Required:						I. Vermilion to forked	
Interference 0		19.3	2.4	0.2	0		
Interference 20%		21.7	0.1	0	0		
Interference 25%		21.9	0	0	0		
Observed	47.3	45.9	6.6	0.2	0	WEINSTEIN 1918	2572
Required:						I. Eosin to cleft	
Interference 0		35.4	13.2	3.3	0.6		
Interference 20%		42.6	9.7	0.4	0		
Interference 25%		44.8	7.9	0.1	0		
Observed	54.6	38.9	6.3	0.3	0.003	GOWEN 1919	31,456
Required:						III. Sepia to rough	
Interference 0		33.1	10.0	2.0	0.3		
Interference 20%		39.6	5.6	0.1	0		
Interference 25%		41.3	4.2	0	0		
Observed	51.2	43.4	5.1	0.3	0	GOWEN 1919	2381
Required:						III. Dichaete to rough	
Interference 0		34.3	11.5	2.6	0.4		
Interference 20%		41.4	7.3	0.2	0		
Interference 25%		43.1	5.8	0.01	0		
Observed	32.6	43.8	20.2	3.2	0.2	WEINSTEIN (unpublished)	4540
Required:						I. <i>Drosophila virilis</i>	
Interference 0		36.6	20.5	7.7	2.2		
Interference 20%		44.7	19.5	3.0	0.1		
Interference 25%		47.2	18.6	1.9	0		

increases, while the proportions with more than one correspondingly decrease. At 20 or 25 percent interference the proportions for the different numbers of breaks become extremely close to those observed.

Thus, if there is interference our second numerical discrepancy, like our first one, disappears; and the degree of interference required to make them disappear is the same for both.

12. The third numerical discrepancy between the results of the linear theory without interference and the observed relations was that the "map distances" determined for the chromosomes, are much too small (see table 6, p. 422). What effect will interference have on these relations?

These map distances are measured in units, each of which is the length required to give 1 percent crossing over. Table 14 shows the map distances

TABLE 14

Comparison of the map distances (in units of 1 percent crossing over), given by BRIDGES (1921), with those required by the linear theory: (a) without interference; (b) with 20 percent interference; (c) with 25 percent interference.

GENE	CHROMOSOME I DISTANCES FROM GENE WHITE				GENE	CHROMOSOME II DISTANCES FROM GENE STAR			
	Given by BRIDGES (1921)	(a) Inter- fer- ence 0	(b) Inter- fer- ence 20 per- cent	(c) Inter- fer- ence 25 per- cent		Given by BRIDGES (1921)	(a) Inter- fer- ence 0	(b) Inter- fer- ence 20 per- cent	(c) Inter- fer- ence 25 per- cent
White	0.0	0.0	0.0	0.0	Star	0.0	0.0	0.0	0.0
Bifid	5.8	5.5	5.3	5.3	Streak	15.4	18.2	15.4	15.4
Club	15.2	16.7	14.3	14.3	Dachs	29.1	42.5	31.4	29.7
Vermilion	31.5	46.6	34.0	32.0	Black	46.5	70.2	49.0	44.5
Miniature	34.6	54.0	38.8	36.0	Purple	52.7	102.8	69.2	60.2
Sable	41.5	86.0	58.8	52.3	Curved	73.5	123.8	83.1	62.7
Rudimentary	53.0	93.2	63.7	55.8	Speck	105.3	167.4	101.6	89.9
Bar	55.5	101.8	68.8	59.8					

given by BRIDGES (1921) for the best-known genes of chromosomes I and II of *Drosophila melanogaster*, in comparison with those required: (a) if there is no interference; (b) if there is 20 percent interference; (c) if there is 25 percent interference. These distances are computed from the most reliable data as to the crossover ratios for the two end genes. For chromosome I, the gene white is taken as the terminus, since its ratios with the others are much better determined than are the ratios with yellow,

commonly taken as the terminus of reference. The ratios employed are: for I, those given by MORGAN and BRIDGES (1916); for II, those given by BRIDGES and MORGAN (1919). The numbers of units required for different ratios, in the cases of 20 and 25 percent interference, were obtained in connection with the extensive tables computed for making the curves of figure 1.

As table 14 shows, while the distances required by the linear theory in case there were no interference are entirely incompatible with those set forth by BRIDGES, with 20 to 25 percent interference the required distances approach those given by BRIDGES. In chromosome I the distances for 25 percent interference are very close to those of BRIDGES (save in the case of the gene *sable*; the suspicion is raised that the crossover ratio of this with *white* (41.2 percent according to MORGAN and BRIDGES 1916) may have been inaccurately determined.) In chromosome II, the better agreement is given, on the whole, by 20 percent interference, although here there are considerable irregularities. This agrees with the facts set forth by BRIDGES (1921), namely, (a) that in the second chromosome different parts show different extents of interference, and (b) except near the ends, this chromosome shows less interference than chromosome I.

In view of the considerable irregularities to be expected from the many sources of error in determining the crossover ratios, and from the fact that different parts of a chromosome may have different degrees of interference, it is clear that the effect of interference is to bring the distances required by the linear theory into essential agreement with those set forth by the workers on *Drosophila*.

CONCLUSION AND DISCUSSION

Thus, if there is interference averaging about 20 to 25 percent, the theory that the genes are arranged in a linear series and that crossing over takes place at conjugation of the chromosomes by breaks at certain points in the series, with interchange of sections of the two chromosomes, yields and requires the entire system of complex and peculiar relations found in nature (set forth in the 17 propositions of pages 397–399); and at the same time it requires and accounts for the specific numerical values,—the relative crossover ratios of different genes, the “map distances,” and the relative proportions of chromosomes showing 0, 1, 2, 3 or more breaks between given genes.

No other theory of crossing over that has been proposed makes so much as a beginning in accounting for this system of relations. The theory of GOLDSCHMIDT (1917), as shown in a previous paper (JENNINGS 1918), re-

quires a definite system of relations between the crossover ratios, but it is not this system; it is a system quite inconsistent with this one. The theory set forth by CASTLE (1919 a, b) likewise does not yield this system; it appears now to have been abandoned by its author. The reduplication theory supported by some English students gives not the slightest ground for the existence of any such system of relations. In view of the complexity and extraordinary character of the system, and in view of the precise numerical proportions in which it is embodied, it is almost impossible to conceive that more than one set of conditions could produce it.

It is to be noted that the conditions required for producing the observed system are simply: (1) the arrangement of the genes in linear series, with breaking and exchange at the time of pairing of the chromosomes; and (2) interference extending over an average distance of 20 to 25 units. It is not required that breaks should be as frequent between any two genes as between any other two, nor that equal absolute distances should in all cases give the same crossover ratio. The possible case that this is true was first investigated, as the simplest condition, and this yielded the fundamental formula (1). In this formula the units of length are the single joints between the genes, and the crossover ratio varies, in the way shown by the formula, with the distance apart of the genes in these units. But as soon as we transform our formulae so as to express the relations in the "standard" units, each of which is the distance necessary to give one percent of crossing over (formula (6) and those following it for no interference; formulae (30), (31), (33) and (34) for interference), they no longer require that breaks should be equally frequent at all joints, nor that equal absolute distances should give the same ratios. The only difference made if breaks are not equally frequent in all parts of the chromosome is that the absolute length of the "standard units" varies in different regions: the formulae based on these units retain their validity, and the system of relations (propositions 1 to 17, page 397) remains unchanged. BRIDGES (1915), STURTEVANT (1917, 1919), PLOUGH (1917) and others have shown that various conditions, environmental or germinal, may alter the frequency of the breaks and exchanges in crossing over. The frequency may be reduced in one part of the chromosome, remaining unchanged in other parts (STURTEVANT 1919); then obviously the "standard units" are no longer of equal absolute length. GOWEN (1919) shows that the frequency of breaks is very variable, and DETLEFSEN and ROBERTS (1921) show that the variations are heritable and that by selection the frequency in a given stock can be greatly changed. But all this does not affect the validity of our formulae that are based on the

standard units; does not affect the general system of relations (propositions 1 to 17); these depend only on the serial arrangement of the genes. If it could be shown that in a set of genes having (according to the linear theory) the order *A-B-C*, the crossover ratios between *A* and *B* and between *B* and *C* can be changed without correspondingly altering that between *A* and *C*, then indeed there would be difficulty for the linear theory, but no indication of this has been shown.

DETLEFSEN (1920), in view of the change in the crossover ratios through selection, entitles a paper "Is crossing-over a function of distance?" and concludes that "In view of these considerations it would perhaps be simpler to conclude that linkage is not a function of distance; i.e., crossing over is not necessarily proportional to distance" (p. 670). Obviously in view of the considerations above set forth, crossing over does not depend on distance alone and is therefore "not necessarily proportional to distance"; many other things affect it. On this there is no dispute; "actual length of the section between loci is only one of the factors determining the amount of crossing over between the loci" (MORGAN 1922, p. 188). The standard unit yielding one percent of crossing over need not be the same in different cases: "It has often been pointed out (e.g., STURTEVANT 1913, p. 49; MORGAN, STURTEVANT, MULLER and BRIDGES 1915, pp. 67-68) that 1 percent of crossing over must not be supposed to represent the same actual morphological distance in different chromosomes or in different regions of the same chromosome" (STURTEVANT 1919, p. 238). But distance remains nevertheless an important factor, in that, of two genes differing in distance, in the same direction, from a given gene *A*, the more distant one will have the higher ratio (provided of course that interference does not go beyond 30 units).

If this relation of the ratios to distance be abandoned, the entire complex system of relations observed between the crossover ratios (page 397) becomes unintelligible and should not exist.

All the evidence is that the genes are in a linear series; the complex system of extraordinary relations shown by the crossover ratios is not explicable on any other ground. If the order of the genes is thus serial, this entire system results, whether the frequency of breaks between genes is or is not uniform in all parts of the chromosomes.

SUMMARY

1. In the only organism adequately^{er} studied (*Drosophila*), the diverse ratios of crossing over between different^r genes show among themselves a complex system of remarkable relations. These are illustrated in table 2,

page 396, and are formulated in seventeen propositions, pages 397-399. The reader should refer to this formulation as an essential part of the present summary.

2. Any set of conditions which are assumed to account for the diverse ratios of crossing over must, if valid, yield and account for this system of relations, and for the numerical proportions in which it is expressed.

The "linear theory,"—the theory that the genes are arranged in series in the chromosomes and that crossing over occurs by breaking and exchange at certain points in the series, at the time of conjugation,—is examined as to whether it fulfils this test.

3. *Crossing over without interference*: From the conditions in the linear theory, formulae are developed for the crossover ratio, first in case the occurrence of a break at one point does not affect the occurrence of breaks elsewhere ("no interference"). Formula (1), p. 405, gives the value of the ratio in terms of the number of genes or joints and of the frequency of a break at any joint; formula (8) the value for any number of "standard units" (each of such a length as to yield one percent of crossing over); formulae (10) and (11) the ratio produced by adding together lengths, for each of which when taken separately the ratio is known. Formulae (16) and (20) give for any value of the crossover ratio the proportion of cases showing any particular number of breaks, as 0, 1, 2, 3, 4, etc.

4. By the aid of these formulae, a series of crossover ratios as required by the linear theory without interference are computed and their relations developed (table 5, p. 418); these are compared with the conditions found in nature. These required ratios are found to show the entire system of complex relations found in nature and set forth in the 17 propositions of pp. 397-399, except the last one (that for which the hypothesis of interference has been proposed). But besides this exception, three important numerical discrepancies are found to exist between the conditions required by the linear theory without interference, and those observed in nature. These are:

(a) The observed numerical values of the ratios could not occur: the observed ratios for distant genes are regularly greater than they could be on the linear theory without interference.

(b) The proportions of the chromosomes showing the different numbers of breaks (0, 1, 2, 3, etc.) are not what are required; there is a marked excess in the proportion showing only one break, and a corresponding deficiency in the proportion showing 2, 3 or more breaks.

(c) The "map distances" of the genes, that have been determined for the chromosomes of *Drosophila* by MORGAN and associates are much less

than are required by the linear theory without interference (table 6, page 422).

Thus the linear theory without interference, in spite of the fact that it yields all but one of the *general* relations found in nature, can not account for the observed ratios and their precise relations.

5. *The linear theory with interference:* To account for the fact set forth in proposition 17, page 399, the hypothesis of interference has been proposed. This assumes that the occurrence of a break at a certain joint in the linear series of genes interferes in some way with the occurrence of another break at any joint within a certain distance.

If valid, the existence of interference must preserve intact the validity of propositions 1 to 16, pp. 397-399; and it must not only account for proposition 17, for which it was devised, but it must also do away with the three discrepancies just mentioned, for which it was not devised. To determine whether it can meet this severe test, mathematical formulae are developed for crossing over with any extent of interference. First are developed formulae for the case, possible but apparently not realized, of complete interference,—a break at any point in the chromosome preventing any other break in that chromosome (formulae (21) to (25)). For interference extending over only a part of the chromosome, for any given distance in units, the fundamental formula is (33), page 437; important special cases are covered in formulae (30), (31) and (34).

6. By the aid of the formulae for the crossover ratios required by the linear theory, with interference, a series of ratios are computed (table 8, page 439, and figure 1); and their interrelations are investigated and compared with those found in nature. The following are found to be the facts:

(a) Crossing over as required by the linear theory with interference extending over a distance of not more than about 30 percent (30 units) yields ratios not rising above 50 percent, and the greater the distance between two genes, the greater the crossover ratio between them. These are the same conditions occurring in crossing over without interference.

(b) If interference extends for a distance greater than 30 percent, then as the distance between the two genes becomes greater, the crossover ratio rises above 50 percent. If interference is sufficiently extensive, it may rise to any value less than 100 percent (see table 10, page 442).

As the distance between the genes is still farther increased, the crossover ratio begins to decrease, falling again below 50 percent. It thus undergoes oscillatory changes, rising and falling successively above and below 50 percent, but the oscillations becoming less and less marked, until at a sufficiently great distance the ratio settles down to about 50 percent

These oscillations are illustrated, for different extents of interference, in the curves of figure 1, page 443.

(c) Thus, for interference extending over long distances (above 30 percent), the serial order of genes would not be given by the increase of the crossover ratio measured from the end gene. Genes farther from the end gene would often have a lower crossover ratio with it than genes nearer to it. Thus, if the magnitude of the ratios were taken as giving the order of the genes, they would appear to be in different orders, when the ratios were measured from different basic genes.

(d) But crossing over as required by the linear theory, with interference extending on the average over a distance of not more than 30 units shows and requires all the general relations set forth in propositions 1 to 17, pages 397-399, as characteristic of the relations found in nature.

(e) Interference extending on the average to a distance equivalent to about 20 to 25 percent crossing over requires and accounts for the conditions giving rise to the three numerical discrepancies (a), (b) and (c) of paragraph 4 of this summary.

7. Thus, in sum, the linear theory of crossing over, with interference of 20 to 25 percent, gives the entire system of complex and peculiar relations found in nature (set forth in the 17 propositions of pages 397-399) and at the same time requires and accounts for the specific numerical data,—the relative ratios of different genes, the “map distances,” and the relative proportions of chromosomes showing 0, 1, 2, 3 or 4 breaks.

No other theory that has been suggested yields any such system. In view of the complexity and extraordinary character of this system, and in view of the precise numerical proportions it involves, it is difficult to conceive that any other set of conditions than that set forth in the “linear theory” could produce it.

LITERATURE CITED

- BRIDGES, C. B., 1915 A linkage variation in *Drosophila*. Jour. Exp. Zoöl. **19**: 1-21.
 1921 Current maps of the location of the mutant genes of *Drosophila melanogaster*. Proc. Nation. Acad. Sci. **7**: 127-132.
- BRIDGES, C. B., and MORGAN, T. H., 1919 The second chromosome group of mutant characters. Carnegie Inst. Washington Publ. 278, pp. 123-304.
- CASTLE, W. E., 1919 a Is the arrangement of the genes in the chromosome linear? The linkage system of eight sex-linked characters of *Drosophila virilis* (data of METZ). Proc. Nation. Acad. Sci. **5**: 25-36.
 1919 b Are genes linear or non-linear in arrangement. Proc. Nation. Acad. Sci. **5**: 500-506.
- DETLEFSEN, J. A., 1920 Is crossing over a function of distance? Proc. Nation. Acad. Sci. **6**: 663-670.
- DETLEFSEN, J. A., and ROBERTS, E., 1921 Studies on crossing over. 1. The effect of selection on crossover values. Jour. Exp. Zoöl. **32**: 333-354.

- GOLDSCHMIDT, R., 1917 Crossing over ohne Chiasmotypie? *Genetics* 2: 82-95.
- GOWEN, J. W., 1919 A biometrical study of crossing over. On the mechanism of crossing over in the third chromosome of *Drosophila melanogaster*. *Genetics* 4: 205-250.
- HALDANE, J. B. S., 1919 The combination of linkage values, and the calculation of distances between the loci of linked factors. *Jour. Genetics* 8: 299-309.
- IBSEN, H. L., 1922 A cross in guinea-pigs best explained by assuming 75 percent crossing over. *Anat. Rec.* 23: 96.
- JENNINGS, H. S., 1918 Disproof of a certain type of theories of crossing over between chromosomes. *Amer. Nat.* 52: 247-261.
- MORGAN, T. H., 1922 Croonian lecture on the mechanism of heredity. *Proc. Roy. Soc. B* 94: 162-197.
- MORGAN, T. H., and BRIDGES, C. B., 1916 Sex-linked inheritance in *Drosophila*. *Carnegie Inst. Washington Publ.* 237, 87 pp.
- MORGAN, T. H., STURTEVANT, A. H., MULLER, H. J., and BRIDGES, C. B., 1915 The mechanism of Mendelian heredity. xiii+262 pp. New York: Henry Holt and Co.
- MULLER, H. J., 1916 The mechanism of crossing over. *Amer. Nat.* 50: 350-366.
- PLOUGH, H. H., 1917 The effect of temperature on crossing over in *Drosophila*. *Jour. Exp. Zööl.* 24: 147-209.
- STURTEVANT, A. H., 1913 The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. *Jour. Exp. Zööl.* 14: 43-59.
- 1917 Genetic factors affecting the strength of linkage in *Drosophila*. *Proc. Nation. Acad. Sci.* 3: 555-558.
- 1919 Inherited linkage variations in the second chromosome. *Carnegie Inst. Washington Publ.* 278, pp. 305-387.
- WEINSTEIN, A., 1918 Coincidence of crossing over in *Drosophila melanogaster (ampelophila)*. *Genetics* 3: 135-172.