

HAS J. G. MENDEL BEEN "TOO ACCURATE" IN HIS EXPERIMENTS?  
THE  $\chi^2$ -TEST AND ITS SIGNIFICANCE TO THE EVALUATION  
OF GENETIC SEGREGATION

by

Franz Weiling

A translation of the German version entitled  
HAT J. G. MENDEL BEI SEINEN VERSUCHEN „ZU GENAU" GEARBEITET?  
DER  $\chi^2$ -TEST UND SEINE BEDEUTUNG FÜR DIE  
BEURTEILUNG GENETISCHER SPALTUNGSVERHÄLTNISSE

in

Der Züchter, V. 36, 1966, pp. 359-65

by

Walter  
W. Piegorsch 19

with an abstract by the translator

BU-718-M

~~September 1980~~

Revised July 1982

HAS J. G. MENDEL BEEN "TOO ACCURATE" IN HIS EXPERIMENTS?  
THE  $\chi^2$ -TEST AND ITS SIGNIFICANCE TO THE EVALUATION  
OF GENETIC SEGREGATION

W. W. Piegorsch  
Biometrics Unit, Cornell University, Ithaca, New York

BU-718-M

September 1980

*Rev. July 1982*

Abstract

Ever since R. A. Fisher published his 1936 article, "Has Mendel's work been rediscovered?", the surprisingly high conformity between ~~his~~ Gregor Mendel's observed and expected ratios in his momentous experiments with peas has plagued ~~the~~ historians of biology and statistics alike. Fisher's calculated  $\chi^2$ -statistic of the experiments, ~~taken as a whole~~ suggested that results on a par with or better than those Mendel reported could only be expected to occur about ~~seven~~ <sup>three</sup> times in every 100,000 attempts, ~~and this~~

*The* arising controversy as to whether or not the good Father "sophisticated" his data continues, unanswered, to this very day.

In a series of lectures given at the 13<sup>th</sup> Biometric Colloquium of the International Biometric Society, German Region, and the Second International Berlin Convention of Mathematical Statistics and Their Applications (Mainz, March 1966 and Berlin, May 1966, respectively), Franz Weiling presented his views on the matter, challenging portions of Fisher's analysis and proposing that the  $\chi^2$  statistic Fisher calculated was in fact an underestimate. He reasoned that the underlying pollination effects in peas are not of a completely random fashion - i.e., binomial - but exhibit more of a semi-random nature, <sup>raising interesting questions for statisticians and experimenters interested in</sup> ~~and he attempted to estimate the factor by which Fisher's  $\chi^2$  is in~~

*genetic count data.*

~~error.~~ Weiling consolidated these results, in <sup>P.ii</sup> German, in a 1966 article, ~~which~~ published in Der Züchter (36, 359-65). ~~No English translation has, until now,~~

~~The only English translation of Weiling's theories has been published in~~

~~theories has been published in~~

(no 7)

In 1970 he published a short synopsis of these theories, in English, in ~~the~~ the annual selection of papers, ~~relating to Mendel's~~ ~~papers~~ (from the Acta Musei Moraviae), Folia Mendeliana (v. 5, 75-77). However, no complete translation has, until now, been made readily available, ~~even~~ though the controversy is no more settled

today than when Weiling first published this piece. This translation will hopefully stimulate further investigation into the nature of the  $\chi^2$  and the problem of Mendel's excessively good fit, by making Weiling's thoughts on the matter more accessible.

~~Doctor Weiling is a member of the Institute for Agricultural Botany at the University of Bonn, West Germany.~~

HAS J. G. MENDEL BEEN "TOO ACCURATE" IN HIS EXPERIMENTS?  
THE  $\chi^2$ -TEST AND ITS SIGNIFICANCE TO THE EVALUATION  
OF GENETIC SEGREGATION

by

Franz Weiling

Institute for Agricultural Botany, University of Bonn, West Germany

and translated by

*Walter*

~~W.~~ W. Piegorsch

Biometrics Unit, Cornell University, Ithaca, New York

BU-718-M\*

September 1980

*Rev. July 1982*

Introduction

In the year 1936, R. A. Fisher published a paper entitled "Has Mendel's work been rediscovered?" In it he investigates the statistical merit of Mendel's work.

Fisher maintains that it was not <sup>only that</sup> ~~just~~ Mendel's contemporaries ~~that~~ didn't understand him—since they perceived his work as a reiteration of previously-known general types of genetic crossing results—but also that his rediscoverers comprehended him only as far as the level of their research demanded.

A. The Statistical Objection to Mendel's Work

Even Fisher, in confronting Mendel's work, expresses reservations. He examines the question of whether the experiments could have been carried out in the manner with which Mendel described them, and comes to the conclusion

---

\* In the Biometrics Unit Mimeo Series, Cornell University, Ithaca, New York.

that Mendel's experimental design and execution must be taken verbatim. However, the examination (carried out with the help of the  $\chi^2$ -test) shows that the degree of accuracy of the experiments is too high; moreover that in two cases, namely the evaluation of the  $F_3$  segregation of the monohybrid and trihybrid crosses, Mendel concludes with incorrect expected values.

The unusually good statistical accuracy of all the experiments, together with the fact that the agreement of the observed values with the resulting comparison values (which were based on incorrect expected values) is of a highly accurate nature, allows Fisher to put forth the question of how genuine the given values were. Fisher certainly doesn't doubt Mendel's honesty; he simply suggests that Mendel may have had an assistant in his work who could possibly have fixed up the data to agree with what he believed to be the expected values. On the other hand, since Mendel was a strong believer in pedagogy, it ~~is in some cases~~ <sup>can be</sup> supposed, <sup>in some cases,</sup> that previous to the beginning of his work (sometime after the enumeration of the 3:1 ratio of segregating seed characteristics - about the year 1858) Mendel may have recognized the regularity of the underlying ~~heredity~~ <sup>genetic</sup> phenomenon, and therefore saw the experimental results as having more of a demonstrative than a true [experimental] character.

Zirkle's (1964) opinion is harsher; probably based on Fisher's analysis. He writes [p. 66]:

"Some modern statisticians, who are armed with the mathematical tools of modern statistics, have reported that Mendel's results were significant - in fact, a little too significant. They were too good, better than we would have a right to expect purely on the basis of chance. Could the good Father Mendel have fudged his results just a little? Could he have omitted a few unusual ratios?"

De Beer (1964) ~~also~~ shares a similar opinion.

B. Proof of the Objection

The expositions Fisher and Zirkle both put forth include reviews of not only the work, but also the personality of the man named Mendel. It is in these reviews that significant clarifications are revealed.

Anyone who takes more than a superficial look at both Mendel's works and his letters (the latter of which R.A. Fisher did not consider) would find it hard to conclude that Mendel sophisticated his data in any way. It also seems unlikely that Mendel had some sort of help at his disposal in his segregation counts. He specifically pointed out that the analysis of the tri-hybrid crosses "demanded the most time and effort of all the experiments", and it would seem imprudent for him not to have identified all <sup>the</sup> help he received in the counts, since he usually stated all of the pertinent procedures in his specifications [of the experiments].

I refer above to the fact that Fisher had not taken Mendel's letters to Nägeli into consideration, illustrating an error of a year in his reconstruction of the experiments, i.e., he begins the experiments in 1857 instead of 1856. There are also two further points of Fisher's exposition that need be carefully considered, as the following discussions will show.

*P* <sup>In all, the</sup> ~~the~~ decisive question that we have before us strikes me as follows: To what degree do the available  $\chi^2$ -based suppositions comply with the existing genetic segregation relationships?

I. Short Overview of Mendel's Experiments Before we study this question in detail, we need to take a quick look at the nature of Mendel's experiments:

Mendel's work consisted of analyses of mono-, di-, and trihybrid crosses. In the monohybrid crosses, the first [two] experiments dealt with the analysis of seed characteristics (1. seed shape: round-wrinkled, 2. seed color: yellow-green). The choice of these characters was especially well

accommodating, since the ripe seeds can be examined immediately after one cross, i.e., [the traits] appear in the same generation. The analysis of the  $F_2$  and  $F_3$  [generations] required the breeding of relatively fewer  $F_1$  and  $F_2$  plants, respectively, since under these conditions each plant yielded an average 30 seeds, thus a large selection was not as important.

The third through seventh experiments concerned themselves with plant form and flower related characters: 3. the color of the seed coat and flowers (grey-brown-white and violet-red-white, respectively); 4. the shape of the pods (somewhat arched-contracted); 5. the color of the unripe pods (green-yellow); 6. the position of the flowers (axial-terminal); 7. the length of the stem (long-short).

The analysis of the  $F_2$  progeny resulted in plants which exhibited offspring of both the recessive and dominant characters. Whereas the former remained constant, the latter exhibited segregations. Exceptional difficulties were encountered in the analysis of the plant characters (experiments 3-7); an  $F_3$  breeding, in which the largest of the individual progeny were limited in their plot space, was required. Mendel analyzed 100 progeny per experiment, the cultivation of which required 10 seeds per progeny to be sown. Therefore he had to account for the selection effects, i.e., the progeny which did not segregate (due to the small sample number) ~~even though they [appeared as dominant] hybrids.~~ *Appearing*

The expected value of this segregation ratio therefore amounted to, provided all 10 seeds germinated, 1.8874:1.1126, instead of 2:1 .

In the dihybrid cross, Mendel again chose seed characters (seed shape round-wrinkled, seed color yellow-green). The  $F_3$  analysis could therefore take place with the seeds of the  $F_2$  plants, without having to consider the particular selection process. The plants segregated in a 1:1:1:1:2:2:2:2:4 ratio.

The trihybrid cross involved the aforementioned seed characters and the color of the seed coats and flowers (grey-brown-white and violet-red-white, respectively). Mendel, who probably classified the soil with the aid of the flower colors, needed an  $F_3$ . He makes no explicit remarks on this, yet he notes that these experiments took "the most time and effort" of all the others. From this, R.A. Fisher's supposition that he also analyzed 10 seeds per offspring in this experiment seems plausible. Again, this is a selection effect that needs to be taken into account in the analysis.

Mendel carried out reciprocal and selfed backcrosses on <sup>forms</sup> ~~forms~~, which were distinguishable in regards to their seed shapes and colors. ✓

In order to demonstrate the variation of the segregation ratios, he presented the first 10 segregation values of <sup>the two</sup> ~~both~~ early experiments <sup>involving</sup> ~~of~~ the monohybrid crosses. ✓

The total  $\chi^2$  [statistic] obtained for the described experiments is presented by Fisher in the following table (Table I). The selection effect is not considered. Rather, the theoretical segregation ratios that Mendel reported are used for the expected ratios. The worst agreement between observed and expected results is reported in the  $F_3$  analysis of the monohybrid crosses, especially in the plant characters ( $\chi^2 = 4.575$ , P-value = 0.6 with 6 degrees of freedom). The total  $\chi^2$  over the entire experiment comes to 41.6056 with 84 degrees of freedom. This corresponds to a [tail] probability of ~~99.993%~~.

99.99[7]%, ✓

Table I. Test of the Conformity with Expectation of Mendel's Segregations  
(from R.A. Fisher, 1936)

	Degrees of Freedom	$\chi^2$	P
3:1 ratios	Seed Characters	2	0.2779
	Plant Characters	5	1.8610
		7	2.1389
2:1 ratios	Seed Characters	2	0.5983
	Plant Characters	6	4.5750
		8	5.1733
Bi-factorial Experiment		8	2.8110
Gametic ratios		15	3.6730
Tri-factorial Experiment		26	15.3224
	Total	64	29.1186
Example of Variation of Individual Plants		20	12.4870
	Total	84	41.6056

0.9999[7] ✓

II. Evaluation of the Mendelian Numerical Data In the  $F_3$  analysis of both the third through the seventh monohybrid experiments and the trihybrid crosses Mendel used false expected values as a basis, in that he did not take the selection effect into account. If we adopt Fisher's argument that he had 10 plants per offspring at his disposal, then [Mendel] should have reckoned that  $(0.75)^{10} = 5.63\%$  of the heterozygous  $F_2$  progeny would not have been recognized as such, since they would not have exhibited segregation. If we accept the resulting segregation ratio of 1.8874:1.1126, then the resulting total  $\chi^2$  for the entire experiment becomes 48.910 with 84 degrees of freedom<sup>1</sup>. This value is also highly significant ( $P \approx 0.999$ ).

The question thus arises: To what degree is the  $\chi^2$ -test [statistic] of the given genetic segregation ratios [actually] correct?

(a) The  $\chi^2$ -Test and its Significance to the Evaluation of Genetic Segregation

The  $\chi^2$ -distribution is based on the squares of normally distributed data. Mendelian segregation ratios are not however, as a rule, distributed normally, but binomially, and even then only in the most ideal of situations.

The basic model of the binomial distribution is the "urn model with replacement", <sup>involving</sup> ~~under~~ a fixed relationship dependent upon the [different] existing sorts of balls in the urn. This model is, however, barely acceptable for genetic segregation ratios. An intrinsic supposition of the urn model is a thorough mix of the balls, or rather, <sup>a</sup> ~~the~~ prerequisite that the choosing of the balls occurs totally at random.

The genetic segregation ratios occur in plants on the basis of the union of specific pollen cells with specific egg cells. For the urn model to apply,

---

<sup>1</sup> All calculations were carried out with the help of the Institute for Applied Mathematics' IBM 7090 computer (at Bonn University). I [wish to] thank the Institute's director, Professor Unger, and his colleagues for their most courteous, eager assistance.

this union must occur totally at random.

This condition is certainly fulfilled in the case of the egg cells, since the maturation division for which of the different alleles is passed on in the egg is decided without a question of doubt. On the other hand, the relationships lie otherwise in Androeceum. Here, normally, four tetrads and ultimately four pollen cells are produced; from the division of the heterozygous characters Aa we receive two A characterized pollen cells, while the other two are "a". Correspondingly, the contents of the pollen sack of a heterozygous plant exhibit half A-pollen, half a-pollen. These are, however, not as well mixed as the balls in the urn model; instead they lie as they were produced, in tetrad arrangements, i.e., they are, as a rule, ordered "half-randomly".

In some plant species, e.g., varieties from the families Ericaceen, Apocynaceen, Asclepiadaceen, and Juncaceen, as well as the varieties Drosera, Anona, Elodea, Typha, Drimys and others, the ripe pollen remains well bound together in tetrand arrangements or in large groups of eight to 64 cells (e.g., in mimosa varieties)(Maheshwari, 1950). Occasionally the contents of almost an entire sporangium is carried to the stigma of some flower—something like the case of certain known Orchidaceen. Even when the individual pollen cells aren't tightly bound together, [we find] cases where they fasten together by means of fine oil droplets; or perhaps they adhere together with a warty, thorny Exine, i.e., in some cases the pollinating bee carries whole balls or bundles of unmixed pollen from flower to flower.

The ratios are completely different with many wind pollinating plants. In this case the pollen is generally separated by wind transport.

If several ovules lie together in the ovary, then the genetic variance will be naturally lowered by the pollination of arriving pollen bound together

in tetrads. A lower genetic variance also occurs when pollen, forming a loose mass in the pollen sac, arrives much like a single chance cell transported by wind.

In the  $\chi^2$ -test, whose general form reads

$$\frac{\chi^2}{n} = \frac{s^2}{\sigma^2} , \quad (1)$$

the variance of a binomial distribution is equated to the character's, or rather, the genetic variance only through the estimation of the segregation ratios; i.e., in the segregation of two characters, the  $\chi^2$ -test reads

$$\chi^2 = \frac{(x-Np)^2}{Np(1-p)} , \quad (2)$$

where  $N$  is the number of individuals in the particular random sample, and  $p$  is the expected fraction of individuals with the characteristic in question.

However, this expression corresponds to the [random] biological relationship depicted above. In order to correct this, we need to enlarge our test [statistic] with a [constant],  $c$ , of currently unknown magnitude. We write —

$$\chi^2 = \frac{(x-Np)^2}{c \cdot Np(1-p)} , \quad (3)$$

where  $c \in (0,1)$  .

(b) Determination of the Magnitude of  $c$  The simplest method of ascertaining how significantly different from 1 the value of  $c$  is, and of obtaining an estimate for  $c$ , clearly seems to be execution of a correspondingly large number of reciprocal backcrosses [of the form]:

1.  $Aa \times aa$       and      2.  $aa \times Aa$  .

In the former case, the pollen used in the backcross is completely identical, while the distribution of characters  $A$  and  $a$  in the egg cells is

random, i.e., it can be viewed as following a binomial distribution. In this case the factor  $c$  is accordingly expected to equal 1. In the latter case, the eggs are all [of] equivalent [genotypes]; in this case the genetic variance is in certain cases lower than that from a binomial distribution, due to the pollination [effects] of the arriving male gametes.

With the use of the usual  $\chi^2$ -test (equation (2)), values of [the statistic] are expected to be increased by the value of  $c$ , in the latter case, from those of the former case. Comparison of the value of  $\chi^2$  in these backcrosses suggests an estimate for  $c$ .

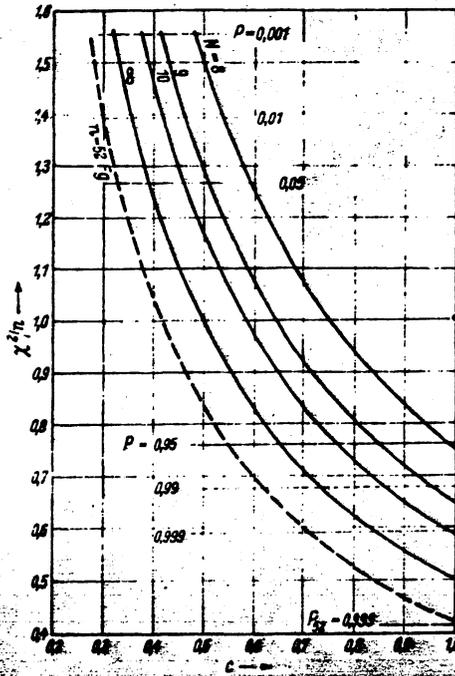
Unfortunately, it is not possible at this time ~~for me~~ to perform ~~such~~ <sup>such</sup> ~~an~~ <sup>[involved]</sup> experiment. In order to, however, preserve some impression of how the value of  $\chi^2$ , or rather  $\chi^2/n$ , changes with respect to  $c$ , we can perform the test on Mendel's data with shrinking values of  $c$ . Illustration 1 exemplifies this result (see the curve for  $N=10$ ). It is especially important to note that the value of  $\chi^2/n$  over the total segregation is not linearly dependent upon  $c$ , rather that it climbs only slightly at first—with values near  $c=1$ —and then grows significantly for values which are substantially less than  $c=1$ . An insignificant deviation of the genetic variance from that of a binomial distribution will not therefore [be easily discernable] from the size of the values  $\chi^2$  or  $\chi^2/n$ .

### III. The Influence of the Selection Process on the Evaluation of Mendel's Data

As mentioned, Mendel laid out 10 seed per progeny for the analysis of the  $F_3$  plant characters and (possibly) also for the trihybrid segregation.

From this, however, one can hardly conclude with Fisher that Mendel also had 10 plants per  $F_3$  progeny at his disposal. It is hardly imaginable

ILL. 1: Test of the segregation ratios observed by Mendel; using increasing values of  $c$  in equation (3).  $\chi^2/n$ -values are pictured as ordinates ( $n$  = degrees of freedom; in the given case,  $n = 84$ ) and  $c$ -values are abscissae. The curves ( $N = \infty, 10, \text{etc.}$ ) suggest that the  $F_3$  progeny are assumed infinitely large or consist of 10, 9, [or] 8 individuals, respectively. The solid horizontal lines indicate the different limiting probabilities. In the ideal case, i.e.,  $c = 1$ , the correct curve should near the point  $\{c, \chi^2/n\} = (1, 1)$ . The broken-line curve indicates the case for the 52 degrees of freedom (not including the  $F_3$  analyses). The corresponding limiting probability is simply indicated by the border  $P_{52} = 0.999$  (lower right).



that all [the] seeds germinated, and when they finally appeared [that they were all] utilized and evaluated (poor germination, feeding birds, etc.).

The expected ratios change correspondingly: if approximately nine plants per progeny appear, then the expected segregation ratio amounts to 1.8498:1.1502, whereas with eight plants per progeny [we get] 1.7998:1.2002 .

Mendel himself stated that in the dihybrid experiment 556 planted seeds yielded 529 (=95.14%) evaluable plants; in the trihybrid experiment, 687 planted seeds yielded 639 plants (=93.01%). In the first backcross 98 round yellow seeds produced 90 evaluable plants (=91.84%), and in the third experiment 94 round yellow seeds produced 87 (=92.55%) plants. Since the raising of a considerably large number of plants—as was necessary in the F<sub>3</sub> analysis—substantially increases the [care and] pains required [to produce them], the possibility always exists that a fraction of the evaluable plants in these experiments was, given certain circumstances, even smaller [than 90%]<sup>2</sup>.

<sup>2</sup> On the occasion of, while pending printing of this essay, a Brünn (Czechoslovakia) lecture delivered on this same topic, a Czech plant breeder remarked that the fraction of plants obtained from sowing pea seeds lies [somewhere] between 80 and 90% of the sown seeds. Thru the kind assistance of Dr. J. Rod, CSc., Brünn, I received thereupon the following statistics (from the Agricultural Center for Control and Research Affairs, Dept. of Varietal Testing—main office, Brünn) for which at this point I also [wish to] express [my] sincere thanks:

Variety	# of Tested Varieties	Period of Experimentation	# of Test Years	Average Fraction of (a) germinating pea seeds (b) evaluable plants	Range of Variation
a) Fraction of germinating, demonstrated pea seeds in laboratory experiments:					
Tall-growing	7	1962-66	4(2)	96.3%	82 -100%
Dwarf-growing	2	1962-66	4	94.5%	91 - 98%
"Pal" peas	3	1960-66	7(4)	95.8%	90 - 99%
"Mark" peas	5	1960-66	7(4)	93.1%	80 -100%
b) Fraction of evaluable plants in field experiments:					
Tall	7	1962-66	4(2)	80.5%	57.5-100%
Dwarf	2	1962-66	4	83.0%	67.9-100%

The values given in column 4 in parentheses are [for] the varieties which were tested in all the specified years. The data mentioned in (b) is based on the counts of plants in field lots, each obtained from drilled seeds (average number of machine drilled seeds = 106 pea seeds/m<sup>2</sup>).

We have not only derived, from these grounds, the ratios for this case - that of 10 available plants per  $F_3$  progeny - but also for the case of nine or eight seeds, respectively, germinating and evaluable plants.

The result is also depicted in Ill. 1 in the form of nearly parallel curves. It specifically points out that with the supposition of eight plants per  $F_3$  progeny and for  $c=1$ , the total data of Mendel['s experiments] results in a P-value of 95.6% (for the test of agreement between observed and expected results), a value which obviously lies near the 95% limit. If one otherwise collectively omitted the  $F_3$  analyses of plant characteristics (including the trihybrid analysis), then ratios are obtained as ~~they~~ are approximately given, even if one assumes 10 individuals for each  $F_3$  offspring. This curve is the lower, [broken] curve in Ill. 1. The position of the curve is, however, dependent on the [number of the] corresponding degrees of freedom, which in this case number only 52. To allow [the reader] to bear these ratios in mind, each curve's limit for the probability of exceeding the value, [the so-called P-value], are supplied.

In the case where the observed segregation ratios are completely determined and the genetic variance corresponds to the variance of a binomial distribution, we expect our curves to drawn near to the point  $(c, \chi^2/n) = (1, 1)$  - the specific position near the point determined by the corresponding degrees of freedom. Under these suppositions we assume an estimate of  $c$  for Mendel's data to lie [in the interval] between 0.4 (0.6 for when the number of estimable  $F_3$  progeny averages eight) and 1.0.

IV. Comparison with the Data of other Authors      A further possibility we [need to] consider in answering our question is how other authors' data on pea [experiments] compares. In conjunction with, and connected to, the rediscovery

of Mendel's work, ~~experiments~~ <sup>experiments</sup> from different ~~schools of heredity~~ <sup>schools of heredity</sup> were being performed. Table 2 gives an overview of the general results obtained in F<sub>2</sub> analyses of a series of such works. ✓

With the help of equation (2) for the test of agreement between observed and expected [results], I was able to find the probabilities above for all the accessible data (F<sub>2</sub> and F<sub>3</sub> analyses, and backcrosses), [presented] in Table 3.

Table 2. The Total Numerical Ratios of Monohybrid F<sub>2</sub> Segregations in Peas [Obtained] by Various Authors; and Their Agreement with Expectation (from Hertwig, 1964)

Author		Observed		Ratios		Standard Error from 3:1
		yellow	green	yellow	green	
Mendel	1865	6022	2001	3.0007	0.9993	± 0.0193
Correns	1900	1394	453	3.0197	0.8803	± 0.0403
Bateson	1900	11903	3903	3.0128	0.9872	± 0.0138
Darbishire	1909	109060	36186	3.0041	0.9959	± 0.0135

Table 3. X<sup>2</sup>-Values, Degrees of Freedom, and P-Values for the Agreement with Expectation of Different Authors' Experiments with Peas

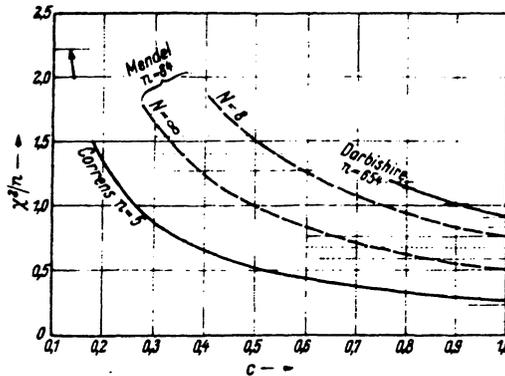
			X <sup>2</sup>	df	P-value
Mendel	1865	(N(F <sub>3</sub> ) = ∞)	41.606	84	0.99993
Mendel	1865	(N(F <sub>3</sub> ) = 10)	48.910	84	0.999
Mendel	1865	(N(F <sub>3</sub> ) = 9)	54.074	84	0.994
Mendel	1865	(N(F <sub>3</sub> ) = 8)	62.997	84	0.956
Mendel	1865	No F <sub>3</sub>	21.707	52	0.9998
Correns	1900		1.303	5	0.92
Tschermak	1900		110.528	132	0.911
Bateson and Killby	1905		411.101	408	0.451
Darbishire	1908/09		597.689	654	0.943
TOTAL	(without Mendel)		1120.621	1199	0.947

With the exception of the experiments by Bateson and Killby (1905), the obtained P-values lie above 90%, while the P-value for Bateson's and Killby's experiment was approximately 45%. The P-value for the summed experiments of the four papers (excluding Mendel's) comes to 94.7%. This agrees quite well with Mendel's result, when we assume that he averaged only eight, or even nine evaluable plants/progeny in his  $F_3$  analysis—an assumption which by no means seems [all that] unreasonable.

Yet, the data taken as a whole accept the hypothesis that the average probability of agreement between observation and expectation in peas exceeds 50%.

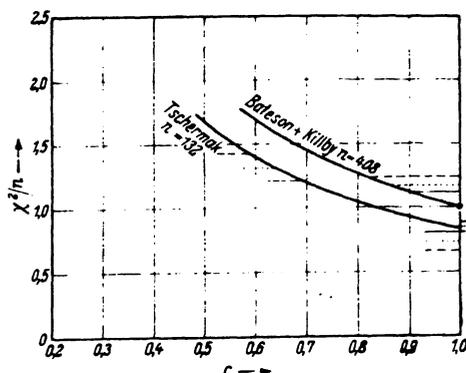
The dependence of the value of  $\chi^2/n$  on  $c$  is [thus] exemplified for these data points by [consulting] Ill. 2 and 3.

ILL. 2: Dependence of the value of  $\chi^2/n$  on  $c$  in the peer segregations [performed] by Correns (1900) and Darbishire (1908, 1909). The relationships for Mendel ( $N = \infty$  and  $N = 8$ ) are [included] for a comparison. Cf. the interpretation in Ill. 1.



When considering all portrayable conditions, therefore, it is truly hard to maintain that Mendel [in some way] falsified his experimental results. It is far more probable to speculate that the value of  $c$  is [significantly] diff-

ILL. 3. Dependence of the value of  $\chi^2/n$  on  $c$  in the pea segregations of von Tschermak (1900) as well as Bateson and Killby (1905). Cf. the interpretation in Ill. 1.



erent than 1. From this we can include the possibility that this value is different in certain circumstances for different strains of peas.

V. Conclusions for the Use of the  $\chi^2$ -Test for the Evaluation of Genetic Segregation

If  $c < 1$ , then the [relatively] unimportant question of what further use the  $\chi^2$ -test in its original form is to us (for the evaluation of segregation ratios) does not appear ~~[necessary]~~ [warranted].

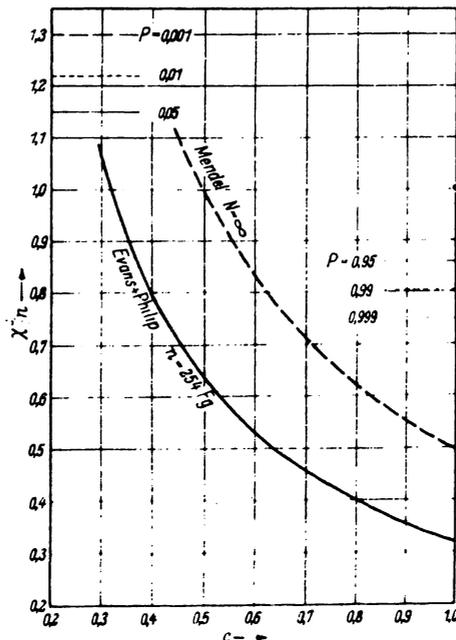
We have attempted to test this question, in that we've ascertained the frequency distributions of the resulting probabilities of the collective works for given authors. These need to be equally divided based on the expectations.

In order to better evaluate an eventual shift of the calculated distributions towards 100%, only the total sums of the given single values in the probability intervals (0, .5) and (.5, 1) are presented. For the pea we obtain little significant displacement towards 100% for either the individual authors or the overall sum of all the observed values. This need not, however, be the general case and is in fact certainly not such (cf. Coelopa frigida (Fab.)).

Table 4. Distribution of the Separate Derived Probabilities in the Pea Segregations of the Given Authors, [set] in Probability Intervals of (0,50) and (50,100%)

Author	df	0 < P < 0.5	0.5 < P < 1
Mendel 1865 ( $N(F_3) = \infty$ )	84	15	27
Mendel 1865 ( $N(F_3) = 10$ )	84	15	27
Mendel 1865 ( $N(F_3) = 9$ )	84	16	26
Mendel 1865 ( $N(F_3) = 8$ )	84	18	24
Correns 1900	5	1	4
Tschermak 1900	132	46	44
Bates and Killby 1905	408	120	116
Darbishire 1908/09	654	307	347
TOTAL		474	511

ILL. 4: Dependence of the value of  $\chi^2/n$  on  $c$  in the seaweed fly Coelopa frigida (Fab.) in results from Evans and Philip (1964).



Discussion

These inquiries have shown that R. A. Fisher, in his evaluation of the precise fit of Mendel's experiments, did not take the following aspects into account (and truly could not, since he approached the analysis ~~more~~ <sup>more</sup> mathematically):

1. The determination of the expected values for the  $F_3$  segregations should not be based on the number of individually sown seeds for progeny, but should instead consider the fact that the number of resulting plants at one's disposal is certainly always smaller [than the number planted].

2. Only in certain known cases, and in no way in general, can an "urn with replacement" model be used for the structure of the segregation relationships. Consequently, an essential assumption (of a binomial or random distribution of the segregation values), which allows the application of the  $\chi^2$ -test, is usually missing. This is for rigorously-taken, normally distributed data, whose squares follow a  $\chi^2$ -distribution. In some, perhaps even many cases, the distribution of segregation values is instead assumed as "semi-random".

If one exclusively considers the condition raised in the former point, ~~total~~ then the total  $\chi^2$  of all of Mendel's segregation data approaches a significance level of 5% (better put, an agreement with expectation of 95%), when the assumption of an average eight resulting evaluable plants from 10 sown seeds in the  $F_3$  analysis is made. This result agrees quite well with the total probability [i.e., P-value] of 94.7%, [attained] in experiments by Correns (1900), Tschermak (1900), Bateson and Killby (1905), and Darbishire (1908, 1909). Previously, Mendel's results could not be considered as unusual on these grounds. Moreover, on the basis of all the experiments, one can infer an estimate of the value of  $c$  between 0.6 and 1.0 .

In spite of the consequent results of our investigations into all the analyzed pea crosses (yielding an appreciable deviation of the value of  $c$  from 1), the  $\chi^2$ -test may, without hesitation, at least be applied to the evaluation of individual segregations. Then the test of frequency distribution of all other inquiries into the underlying individual probabilities shows no noticeable deviation in the direction of  $P=1.0$ , as Table 4 illustrates.

It does not however follow that this is the case for all organisms.

Evans and Philip (1964) inform us that for the zoological creature, the seaweed fly Coelopa frigida (Fab.), the variance of the distribution of the species departs considerably from that of a binomial (ILL. 4). In this case the investigation shows a significant deviation of the individual probabilities in the direction of  $P=1.0$ .

In Table 5 the total probabilities for the segregations observed in both the hen and various plants by Bateson and others are given. This analysis shows that the evaluated plants (Datura, Lychnis, Lathyrus and Salvia) yield overall P-values over 50%, with the exception of Salvia (the P-value for all the above plants amounts to 66.02%). In comparison, the observed P-values for the hen in all three cases were obviously lower than 50% (the total P-value in these segregations reached only 0.777%).

Using extensive genetic inquiries within certain species, we should in each case attempt to construct some estimate of the size of the factor  $c$  and the deviation of the segregation from the binomial relationships.

Gregor Mendel had not himself recognized the necessity of considering the selection effect in the evaluation of the segregations in his  $F_3$  analysis, inasmuch as plant, and not seed, characters bring other aspects [in these analyses] into account.

Table 5.  $\chi^2$ -Values, Degrees of Freedom and P-Values for the Agreement with Expectation for Various Plant Types and also for the Hen (Experiments from Bateson and others)

Author		Species	df	$\chi^2$	P
Bateson and Saunders	1902	<u>Datura</u>	105	99.230	0.60 < P < 0.70
Bateson and Saunders	1902	<u>Lychnis</u>	28	13.984	0.98 < P < 0.99
Bateson and Saunders	1902	Hen	60	74.951	0.05 < P < 0.10
Bateson and Punnett	1905	Hen	141	172.956	0.03 < P < 0.05
Hurst	1905	Hen	31	38.576	0.10 < P < 0.20
Bateson, Saunders and Punnett	1905	<u>Lathyrus</u>	29	26.953	0.50 < P < 0.60
Bateson, Saunders and Punnett	1905	<u>Salvia</u>	18	31.595	P $\approx$ 0.025

Summary

In a statistical analysis of Gregor Mendel's experiments, R. A. Fisher (1936) expresses the opinion that from a statistical point of view Mendel's experimental results are too exact. Supposing that Mendel recognized the regularities of segregation from even his first seed counts in 1858, Fisher believes that further experimentation by Mendel was only of demonstrative value. Several authors, e.g., C. Zirkle (1964), share in this opinion.

However, Fisher and the other authors have overlooked the fact that in judging  $F_3$  plant character—as opposed to the less necessary seed character—analyses of Mendel's experiments, the number of progeny available for classification could hardly be equal to the 10 seeds (in some cases presumably) planted, since one must count on losses through poor germination, birds, or other problems. It is shown that with a supposed average of eight evaluable plants per progeny, the probability of agreement with expectation in all of Mendel's [pea] experiments is effectively similar to the probabilities calcu-

lated by Correns (1900), von Tschermak (1900), Bateson and Killby (1905), and even Darbishire (totalled-1908, 1909) in their results with peas.

This excessive goodness of fit with the pea could be explained by considering the fact that the distribution of segregation data is obviously different for each genetic species, plant or animal; and will have in fact more of a "semi-random" rather than binomial nature. The resulting  $\chi^2$  [statistic] will thus be underestimated.

The factor, c, by which this  $\chi^2$  [statistic] is underestimated and the resulting consequences of the usefulness of the  $\chi^2$ -test are considered. A comparison of other authors' results in the literature is also presented, in which both animal and plant species were examined.

#### References

1. Bateson, W., and Miss Killby: Peas (*Pisum sativum*). Reports to the Evolution Committee of the Royal Society: Experimental Studies in the Physiology of Heredity, pp. 55-80 (1905).
- 1a. deBeer, G.: Mendel, Darwin, and Fisher (1865-1965) Notes and Records. Royal Society of London, v. 19, 192-226 (1964).
2. Correns, C.: G. Mendel's rules concerning the behavior of progeny of hybrid strains. German Botanical Society of Berlin, v. 18, 158-168 (1900).
3. Darbishire, A. D.: On the result of crossing round with wrinkled peas, with especial reference to their Starch-grains. Proceedings of the Royal Society of London, Series B, v. 80, 122-135 (1908).
4. Darbishire, A. D.: An experimental estimation of the theory of ancestral contributions in heredity. Proceedings of the Royal Society of London, Series B, v. 81, 61-79 (1909).
5. Evans, D. A., and U. Philip: On the distribution of Mendelian ratios. Biometrics, v. 20, 794-817 (1964).
6. Fisher, R. A.: Has Mendel's work been rediscovered? Annals of Science, v. 1, 115-137 (1936).

7. Hertwig, P.: Why are disturbances of Mendelian ratios of interest, and how can they help explain the case of tailless mice? Forschungen und Fortschritte, v. 38, 353-357 (1964).
8. Maheshwari, P.: An Introduction to the Embryology of Angiosperms. 453 S. New York: McGraw-Hill Book Company, 1950.
9. Mendel, G.: Experiments in Plant Hybridization. Proceedings of the Brünn Society for Scientific Research, v. 4, 3-47, 1865 (1866).
10. Mendel, G.: Letters to Carl Nägeli (1866-1873), edited by C. Correns: Essays in Mathematics and Physics of the Royal Class of the Saxon Society of Science, v. 29, 189-265, and C. Correns: Society's Essays, 1233-1297; Berlin: Springer-Verlag, 1924.
11. Tschermak-Seysenegg, E.: On artificial crossings of Pisum sativum. Journal of Agricultural Research Methods of Austria (Vienna), v. 3, 465-555 (1900).
12. Zirkle, C.: Some oddities in the delayed discovery of Mendelism. Journal of Heredity, v. 55, 65-72 (1964).

<sup>1</sup> All calculations were carried out with the help of the Institute for Applied Mathematic's IBM 7090 computer (at Bonn University). I [wish to] thank the Institute's director, Professor Unger, and his colleagues for their most courteous, eager assistance.

<sup>2</sup> On the occasion of, while pending printing of this essay, a Brünn (Czechoslovakia) lecture delivered on this same topic, a Czech plant breeder remarked that the fraction of plants obtained from sowing pea seeds lies [somewhere] between 80 and 90% of the sown seeds. Thru the kind assistance of Dr. J. Rod, CSc., Brünn, I received thereupon the following statistics (from the Agricultural Center for Control and Research Affairs, Dept. of Varietal Testing-main office, Brünn) for which at this point I also [wish to] express [my] sincere thanks:

Variety	# of Tested Varieties	Period of Experimentation	# of Test Years	Average Fraction of (a) germinating pea seeds (b) evaluable plants	Range of Variation
a) Fraction of germinating, demonstrated pea seeds in laboratory experiments:					
Tall-growing	7	1962-66	4(2)	96.3%	82 -100%
Dwarf-growing	2	1962-66	4	94.5%	91 - 98%
"Pal" peas	3	1960-66	7(4)	95.8%	90 - 99%
"Mark" peas	5	1960-66	7(4)	93.1%	80 -100%
b) Fraction of evaluable plants in field experiments:					
Tall	7	1962-66	4(2)	80.5%	57.5-100%
Dwarf	2	1962-66	4	83.0%	67.9-100%

The values given in column 4 in parentheses are [for] the varieties which were tested in all the specified years. The data mentioned in (b) is based on the counts of plants in field lots, each obtained from drilled seeds (average number of machine drilled seeds = 106 pea seeds/m<sup>2</sup>).