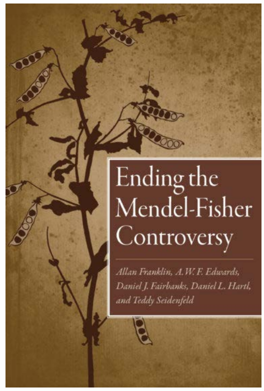
Ending the Mendel-Fisher Controversy
By Allan Franklin, A.W.F. Edwards,
Daniel J. Fairbanks, Daniel L. Hartl,
Teddy Seidenfeld, 2008



CHAPTER 1

The Mendel-Fisher Controversy

An Overview

ALLAN FRANKLIN

Gregor Mendel (1822–1884) is regarded as the founder of modern genetics. His experiments on pea plants reported in 1865 established the principles of segregation and of independent assortment. The former states that variation for contrasting traits is associated with a pair of factors that segregate to individual reproductive cells. The latter states that two or more of these factor-pairs assort independently to individual reproductive cells. It is well known that Mendel's work was neglected until its "rediscovery" in 1900 by Hugo de Vries, Carl Erich Correns, and Erich von Tschermak. It is less well known, however, that in 1936, the great British statistician and biologist R. A. Fisher analyzed Mendel's data and found that the fit to Mendel's theoretical expectations was too good (Fisher 1936). Using χ^2 analysis, Fisher found that the probability of obtaining a fit as good as Mendel's was only 7 in 100,000. Fisher also argued that because Mendel used only a limited sample of 10 plants in his experiment to determine the ratio of heterozygous plants (*Aa*) to homozygous plants (*AA*) in the F_2 generation produced by the self-pollination of hybrids, there was a 5.63% chance of misidentifying heterozygous plants as being homozygous. Thus, the ratio should be approximately 1.7 to 1, rather than Mendel's expectation of 2 to 1, although Mendel's data agreed more closely with the 2 to 1 ratio. Fisher concluded: "This possibility is supported by independent evidence that the data of most, if not all, of the experiments have been falsified so as to agree closely with Mendel's expectations" (*134*).[1] Fisher did not believe that Mendel was responsible for the falsification, but attributed it to an unknown assistant.

Fisher's work was overlooked. The first published comments on it appeared in 1964, about the time of the centenary of Mendel's paper, and since then at least 50 papers, letters, and discussions have been published on the controversy as to whether Fisher adequately showed that Mendel's data were falsified. These publications include explanations of Mendel's results and both criticisms and defenses of Fisher.

This chapter will provide an overview of that controversy, including summaries of both Mendel's and Fisher's papers, along with a discussion of most of the papers on the debate. It is not, however, a substitute for reading the original works. Therefore, this book contains the work of both Mendel and Fisher as well as four of the most significant discussions of the controversy, and updates by those four authors. I believe that taken together, these voices argue for an end to the controversy.

Mendel's Experimental Results

Mendel began his experiments on garden peas (*Pisum sativum* L.) in 1856 and continued them until 1863, a period of approximately eight years. His stated purpose was to investigate whether there was a general law for the formation and development of hybrids, something he noted had not yet been formulated:

> Those who survey the work done in this department will arrive at the conviction that among all the numerous experiments made, not one has been carried out to such an extent and in such a way as to make it possible to determine the number of different forms under which the offspring of hybrids appear, or to arrange these forms with certainty according to their separate generations, or definitely to ascertain their statistical relations.
>
> It requires indeed some courage to undertake a labour of such far-reaching extent; this appears however, to be the only right way by which we can finally reach the solution of a question the importance of which cannot be overestimated in connection with the history of the evolution of organic forms. (*79*)[2]

Mendel proposed to remedy the situation and did so. As Fisher remarked, "Mendel's paper is, as has been frequently noted, a model in respect of the order and lucidity with which the successive relevant facts are presented" (Fisher 1936, *121*). I will follow Mendel's plan in describing his experiments and will allow Mendel to speak for himself as much as possible.

In order to carry out such experiments successfully, Mendel required: "The experimental plants must necessarily—1. Possess constant differentiating characters. 2. The hybrids of such plants must, during the flowering period, be protected from the influence of all foreign pollen, or be easily capable of such protection. The hybrids and their offspring should suffer no marked disturbance in their fertility in the successive generations" (79). He further noted: "In order to discover the relations in which the hybrid forms stand towards each other and also towards their progenitors it appears to be necessary that all members of the series developed in each successive generation should be, *without exception*, subjected to observation" (79–80).

[F_2] The First Generation [Bred] from the Hybrids

Mendel began with 34 varieties of peas, from which he selected 22 varieties for further experiments. He had confirmed, in two years of experimentation, that these varieties bred true. He reported experiments on seven characters that had two easily distinguishable characteristics. I have listed these below, with the dominant form first:[3]

1. Seed shape: round or wrinkled
2. Cotyledon color: yellow or green
3. Seed-coat color: colored (gray, gray-brown, or leather-brown) or white. Colored seed coats were always associated with violet flower color and reddish markings at the leaf axils. White seed coats were associated with white flowers.
4. Pod shape: inflated or constricted
5. Pod color: green or yellow
6. Flower position: axial (along the stem) or terminal (at the end of the stem)
7. Stem length: long (six to seven feet) or short (three-quarters of a foot to one and a half feet)

The first two are seed characters because they are observed in seed cotyledons, which consist of embryonic tissue. Each seed is thus a genetically different individual and such characters may differ among the seeds produced on a heterozygous plant. Both yellow and green seeds may be observed on a single heterozygous plant. One may, in fact, observe these characters for the next generation, without the necessity of planting the seeds. The latter five are plant characters. As William Bateson remarked, "It will be observed that the [last] five are *plant-characters*. In order to see the result of crossing, the seeds must be sown and allowed to grow into plants. The [first] two characters belong to the *seeds* themselves. The seeds of course are members of a generation later than that of the plant which bears them" (Bateson 1909, 12). Because of this, Mendel would have had a reasonable expectation of what the results of his plant character experiments would be from his observations of the seed characters, before the plants of the next generation were grown.

Mendel's first experiment was to breed a generation of hybrids from his true breeding plants for each of the seven characters. His results for this generation (F_1) clearly showed dominance. He remarked: "In the case of each of the seven crosses the hybrid-character resembles that of one of the parental forms so closely that the other either escapes observation completely or cannot be detected with certainty" (*84*).

He then allowed these monohybrids to self-fertilize. He found a 3:1 ratio for plants that showed the dominant character to those that possessed the recessive character in this generation (F_2).[4] He found that this ratio held for all the characters observed in the experiments and that "*Transitional forms were not observed in any experiment*" (*85*). His results are shown in table 1.1. He concluded, "If now the results of the whole of the experiments be brought together, there is found, as between the number of forms with the dominant character and recessive characters, an average ratio of 2.98 to 1, or 3 to 1" (*87*).

Mendel also noted that the distribution of characters varied in both individual plants and in individual pods. He illustrated this with data from the first ten plants in the seed character experiments (see table 1.2). The variation in both the ratios of the characters and in the number of seeds per plant is considerable. Mendel also presented the extreme variations.

TABLE 1.1 Mendel's results for the F_2 generation of monohybrid experiments (from data in Mendel 1865, *85–87*)

	Expected ratio 3:1				
Trait	*Dominant*	*Number*	*Recessive*	*Number*	*Ratio*
1. Seed shape	Round	5474	Angular	1850	2.96
2. Cotyledon color	Yellow	6022	Green	2001	3.01
3. Seed coat color	Colored	705	White	224	3.15
4. Pod shape	Inflated	882	Constricted	299	2.95
5. Pod color	Green	428	Yellow	152	2.82
6. Flower position	Axial	651	Terminal	207	3.14
7. Stem length	Long	787	Short	277	2.89
Total	Dominant	14,949	Recessive	5010	2.98

TABLE 1.2 Mendel's results for the first 10 plants in the experiments on seed shape and seed color (from data in Mendel 1865, 86)

Plant	*Experiment 1: Shape of seeds*		*Experiment 2: Coloration of albumen*	
	Round	*Wrinkled*	*Yellow*	*Green*
1	45	12	25	11
2	27	8	32	7
3	24	7	14	5
4	19	10	70	27
5	32	11	24	13
6	26	6	20	6
7	88	24	32	13
8	22	10	44	9
9	28	6	50	14
10	25	7	44	18
Ratio	3.33 : 1		3.08 : 1	

Note: The fact that the number of seeds in each plant differs for each numbered plant shows clearly that the plants for Experiments 1 and 2 are different plants. Thus, plant 1 in Experiment 1 has 57 seeds, whereas plant 1 in Experiment 2 has 36 seeds.

"As extremes in the distribution of the two seed characters in one plant, there were observed in Expt. 1 an instance of 43 round and only 2 angular, and another of 14 round and 15 angular seeds. In Expt. 2 there was a case of 32 yellow and only 1 green seed, but also one of 20 yellow and 19 green" (*86*). Mendel was clearly willing to present data that deviated considerably from his expectations.[5]

Mendel also noted: "In well-developed pods which contained on the average six to nine seeds, it often happened that all the seeds were round (Expt. 1) or all yellow (Expt. 2); on the other hand, there were never observed more than 5 wrinkled or five green ones in one pod" (*86*).[6]

[F_3] The Second Generation [Bred] from the Hybrids

At the end of the section describing the first-generation experiments, Mendel remarked that the dominant character could have a "*double signification.*" It could be either a pure parental (dominant) character or a hybrid character. "In which of the two significations it appears in each separate case can only be determined by the following generation. As a parental character it must pass over unchanged to the whole of the offspring; as a hybrid-character, on the other hand, it must maintain the same behaviour as in the first generation" (*87*).[7] He further noted that those plants that show the recessive character in the first generation (F_2) do not vary in the second generation (F_3).[8] They breed true. That was not the case for those plants showing the dominant character: "Of these *two*-thirds yield offspring which display the dominant and the recessive characters in the proportion of 3 to 1, and thereby show exactly the same ratio as the hybrid forms, while only *one*-third remains with the dominant character constant"(*88*). In other words, of those F_2 generation plants showing the dominant character, two-thirds were heterozygous (*Aa*), or hybrid, and one third homozygous (*AA*). For the seed characters Mendel reported the following results: (1) from 565 plants raised from round seeds, 372 produced both round and wrinkled seeds in the proportion of 3 to 1 whereas 193 yielded only round seeds, a ratio of 1.93 to 1; (2) for plants raised from yellow seeds, 353 yielded both yellow and green seeds in the proportion 3 to 1, whereas 166 yielded only yellow seeds, a ratio of 2.13 to 1.

The experiments on plant characters required more effort: "For each separate trial in the following experiments [on plant characters] 100 plants were selected which displayed the dominant character in the first generation [F_2], and in order to ascertain the significance of this, ten [F_3] seeds of each were cultivated" (*88*).[9] A plant was classified as homozygous if all of the 10 offspring had the dominant character and classified as heterozygous otherwise.[10] Mendel's results for the plant characteristics are shown in table 1.3. Mendel noted that the first two experiments on seed characters were of special importance because of the large number of plants that could be compared. Those experiments yielded a total of 725 hybrid plants and 359 dominant plants that "gave together almost exactly the average ratio of 2 to 1" (*89*). Experiment 6 also yielded almost the exact ratio expected, whereas for the other experiments, as Mendel noted, "the ratio varies more or less, as was only to be expected in view of the smaller number of 100 trial plants" (*89*). Mendel was, however concerned about Experiment 5 (the color of unripe pods), in which the result was 60 to 40. He regarded these numbers as deviating too much from the expected 2 to 1 ratio.[11] Mendel repeated the experiment and obtained a ratio of 65 to 35, and was satisfied: "*The average ratio of 2 to 1 appears, therefore, as fixed with certainty*" (*89*). It is clear that Mendel did not attempt to hide any of his results, especially those that deviated from his expectations, because he presented the results for both the original Experiment 5 as well as its repetition. The sum totals for the six plant characteristic experiments, in-

TABLE 1.3 Mendel's results for the heterozygous-homozygous experiment (the 2 to 1 experiment) (from data in Mendel 1865, 88)

Experiment	Dominant	Hybrid
3. Seed coat color (grey-brown or white)	36	64
4. Pod shape (smooth or constricted)	29	71
5. Pod color (green or yellow)	40	60
6. Flower location (axillary or terminal)	33	67
7. Stem length (long or short)	28	72
8. Repetition of Experiment 5	35	65
Total	201	399
	Ratio (hybrid to dominant) 1.99	

cluding the repetition of Experiment 5, were 399 (hybrid) to 201 (dominant), or 1.99 to 1.

Mendel's conclusion was quite clear:

> The ratio of 3 to 1, in accordance with which the distribution of the dominant and recessive characters results in the first generation, resolves itself into a ratio of 2:1:1 if the dominant character be differentiated according to its significance as a hybrid-character or as a parental one. Since the members of the first generation [F_2] spring directly from the seed of the hybrids [F_1], *it is now clear that the hybrids form seeds having one or the other of the two differentiating characters, and of these one-half develop again the hybrid form, while the other half yield plants which remain constant and receive the dominant or the recessive characters, [respectively], in equal numbers.* (89)

The Subsequent Generations Bred from the Hybrids

Mendel suspected that the results he had obtained from the first and second generations produced from monohybrids were probably valid for all of the subsequent progeny. He continued the experiments on the two seed characters, shape and color, for six generations; the experiments on seed-coat color and stem length for five generations; and the remaining three experiments on pod shape, color of pods, and position of flowers for four generations, "and no departure from the rule has been perceptible. The offspring of the hybrids separated in each generation in the ratio of 2:1:1 into hybrids and constant forms [pure dominant and pure recessive]" (89). He did not, however, present his data for the experiments on the subsequent generations.[12] He went on to state, "If A be taken as denoting one of the two constant characters, for instance the dominant, a, the recessive, and Aa the hybrid form in which both are conjoined, the expression $A + 2Aa + a$ shows the terms in the series for the progeny of the hybrids of two differentiating characters" (*89*).[13]

The Offspring of Hybrids in which Several Differentiating Characters Are Associated

Mendel's next task, as he put it, was to investigate whether the laws he had found for monohybrid plants also "applied to each pair of differentiating characters when several diverse characters are united in the hybrid by crossing" (*90*).

He went on to describe the experiments. "Two experiments were made with a considerable number of plants. In the first experiment the parental plants differed in the form of the seed and in the colour of the albumen; in the second in the form of the seed, in the colour of the albumen, and in the colour of the seed-coats. Experiments with seed characters give the result in the simplest and most certain way" (*91*). He was no doubt referring to the greater number of seeds than plants, which provides data with greater statistical significance, and also to the fact that the shape of the seeds and the color of albumen (cotyledons) could be seen in the second generation, without the need to plant a third generation. Daniel Fairbanks and Bryce Rytting (2001) later remarked with reference to seed-coat color, which, as noted above, was correlated with the presence or absence of axillary pigmentation, could be scored in seedlings, and was also used as the third factor in the trifactorial experiment: "Because this trait can be scored in seedlings, it is an excellent choice for the third trait in the trihybrid experiment because it creates at most a three-week delay between data collection for the first two traits and the third. Garden space is not as critical because many seedlings can be grown in the space occupied by a single mature plant" (*276*).

In these experiments Mendel distinguished between the differing characters in the seed plant and the pollen plant. A, B, and C represented the dominant characters of the seed plant and a, b, and c the recessive characters of the pollen plant, with hybrids represented as Aa, Bb, and Cc.[14]

First Experiment (Bifactorial)

Mendel's first experiment used two seed characters in which the seed plant (AB) was A (round shape) and B (yellow cotyledon), and the pollen plant (ab) was a (wrinkled shape) and b (green albumen). The fertilized

TABLE 1.4 Mendel's results for the bifactorial experiment (from Mendel 1865, 91–92)

	A (round)	*Aa* (hybrid)	*a* (angular)
B	*AB* (round, yellow): 38	*AaB* (round yellow and angular yellow): 60	*aB* (angular, yellow): 28
Bb	*ABb* (round yellow and green): 65	*AaBb* (round yellow and green and angular yellow and green): 138	*aBb* (angular yellow and green: green): 68
b	*Ab* (round green): 35	*Aab* (round and angular green): 67	*ab* (angular green): 30

seeds were all round and yellow, as expected. He then raised plants from these seeds and obtained 15 plants with 556 seeds distributed as follows:

315 round and yellow
101 wrinkled and yellow
108 round and green
32 wrinkled and green[15]

All of these seeds were planted in the following year and Mendel's results are shown in table 1.4.

Mendel separately recorded the results for each set of the 556 seeds (i.e., round and yellow, round and green, wrinkled and yellow, wrinkled and green).[16] He noted that there were nine different forms (we would say genotypes) and classified them this way:

> The whole of the forms may be classed into three essentially different groups. The first includes those with the signs *AB, Ab, aB, ab*: they possess only constant characters and do not vary again in the next generation. Each of these forms is represented on the average thirty-three times. The second group includes the signs *ABb, aBb, AaB, Aab*: these are constant in one character and hybrid in another, and vary in the next generation only as regards the hybrid-character. Each of these appears on an average sixty-five times. The form *AaBb* occurs 138 times: it is hybrid in both characters, and behaves exactly as do the hybrids from which it is derived.
>
> If the numbers in which the forms belonging to these classes appear to be compared, the ratios of 1, 2, 4 are unmistakably evident. The numbers 32, 65, 138 present very fair approximations to the ratio numbers of 33, 66, 132. (92)

Mendel had a very good feel for his data and an ability to see the underlying patterns in his results despite statistical fluctuations. Mendel concluded that these results "indisputably" showed that the results could be explained by the combination of $A + 2Aa + a$ and $B + 2Bb + b$ (i.e., $AB + 2AaB + aB + 2ABb + 4AaBb + 2aBb + Ab + 2Aab + ab$).

Second Experiment (Trifactorial)

In this experiment, Mendel investigated whether the results he had obtained in both the monohybrid and bifactorial experiments held for an experiment in which three different characters were examined, the trifactorial experiment. He remarked, "Among all the experiments it demanded the most time and trouble" (93). The characters investigated for the seed plant (*ABC*) were: *A* (round shape), *B* (yellow albumen), and *C* (gray-brown seed coat); and for pollen plant (*abc*): *a* (wrinkled seed), *b* (green albumen), and *c* (white seed coat). The first two were seed characters and could be observed immediately, whereas seed-coat color, a plant character, required plants from the next generation.[17] Mendel obtained 687 seeds from 24 hybrid plants, from which he successfully grew 639 plants and "as further investigations showed,"[18] he obtained the results depicted in table 1.5. He summarized his data as follows:

> The whole expression contains 27 terms. Of these 8 are constant in all characters, and each appears on the average 10 times; 12 are constant in two characters, and hybrid in the third; each appears on the average 19 times; 6 are constant in one character and hybrid in the other two; each appears on the average 43 times. One form appears 78 times and is hybrid in all of the characters. The ratios 10, 19, 43, 78 agree so closely with the ratios 10, 20, 40, 80, or 1, 2, 4, 8, that this last undoubtedly represents the true value. (94)[19]

TABLE 1.5 Mendel's results for the trifactorial experiment (Mendel 1865, 93)

8 plants	*ABC*	22 plants	*ABCc*	45 plants	*ABbCc*
14 "	*ABc*	17 "	*AbCc*	36 "	*aBbCc*
9 "	*AbC*	25 "	*aBCc*	38 "	*AaBCc*
11 "	*Abc*	20 "	*abCc*	40 "	*AabCc*
8 "	*aBC*	15 "	*ABbC*	49 "	*AaBbC*
10 "	*aBc*	18 "	*ABbc*	48 "	*AaBbc*
10 "	*abC*	19 "	*aBbC*		
7 "	*abc*	24 "	*aBbc*		
		14 "	*AaBC*	78 "	*AaBbCc*
		18 "	*AaBc*		
		20 "	*AabC*		
		16 "	*Aabc*		

Mendel went on to say that this series resulted from combining $A + 2Aa + a$, $B + 2Bb + b$, and $C + 2Cc + c$. He had a strong feeling about the expected results and was willing to accept conclusions despite limited statistics. As Fisher remarked, "He evidently felt no anxiety lest his counts should be regarded as insufficient to prove his theory" (121).

Mendel remarked that he had conducted several other experiments in which the remaining characters were combined in twos and threes and that these gave approximately equal results, but he presented none of his data for these experiments. He concluded:

> There is therefore no doubt that for the whole of the characters involved in the experiments the principle applies that *the offspring of the hybrids in which several essentially different characters are combined exhibit the terms of a series of combinations, in which the developmental series for each pair of differentiating characters are united.* It is demonstrated at the same time that *the relation of each pair of different characters in hybrid union is independent of the other differences in the two original parental stocks.* (94)

In Mendel's opinion, his results justified belief that the same behavior applied to characters that could not be so easily distinguished. He noted, however, the difficulty of such experiments: "An experiment with peduncles of different lengths gave on the whole a fairly satisfactory result, although the differentiation and serial arrangement of the forms could not be effected with that certainty which is indispensable for correct experiment" (95).

The Reproductive Cells of Hybrids

In his bifactorial and trifactorial experiments, Mendel used seed plants with the dominant characters and pollen plants with the recessive characters. The question remained whether his results would remain the same if those parental types were reversed. He stated that in hybrid plants, it was reasonable to assume that there were as many kinds of egg and pollen cells as there were possibilities for constant combination forms. He further noted that this assumption, combined with the idea that the different kinds of egg and pollen cells are produced on average in equal numbers, would explain all of his previous results.

Mendel proposed to investigate these issues explicitly in a series of experiments. He chose true breeding plants as follows: seed plant (AB); where A and B were round shape and yellow albumen, respectively; pollen plant ab, where a and b were wrinkled shape and green albumen, respectively. These were artificially fertilized and the hybrid $AaBb$ obtained. Both the artificially fertilized seeds, together with several seeds from the two parental plants, were sown. He then performed the following fertilizations:

1. The hybrids with the pollen from *AB*
2. The hybrids with the pollen from *ab*
3. *AB* with pollen of the hybrid
4. *ab* with pollen of the hybrid

For each of these experiments, all of the flowers on three plants were fertilized. Mendel stated that if his assumptions were correct, then the hybrids would contain egg and pollen cells of the form *AB*, *Ab*, *aB*, and *ab*. When combined with the egg and pollen cells from the parental plants *AB* and *ab*, the following patterns emerge.

1. *AB*, *ABb*, *AaB*, *AaBb*
2. *AaBb*, *Aab*, *aBb*, *ab*
3. *AB*, *ABb*, *AaB*, *AaBb*
4. *AaBb*, *Aab*, *aBb*, *ab*

These genotypes should occur with equal frequency in each experiment. Experiments 1 and 3, as well as experiments 2 and 4, would demonstrate that the results are independent of which parent is used for pollen and which is used for seed. Mendel also noted that there would be statistical fluctuations in his data.

> If, furthermore, the several forms of the egg and pollen cells of the hybrids were produced on an average in equal numbers, then in each experiment the said four combinations should stand in the same ratio to each other. A perfect agreement in the numerical relations was, however, not to be expected, since in each fertilisation, even in normal cases, some egg cells remain undeveloped or subsequently die, and many even of the well-formed seeds fail to germinate when sown. The above assumption is also limited in so far that, while it demands the formation of an equal number of the various sorts of egg and pollen cells, it does not require that this should apply to each separate hybrid with mathematical exactness. (97)

Mendel predicted that in Experiments 1 and 3 all of the seeds produced would be round and yellow, the result of dominance. For Experiments 2 and 4, his expectations were that round yellow seeds, round green seeds, wrinkled yellow seeds, and wrinkled green seeds would be produced in equal proportions. He reported: "The crop fulfilled these expectations perfectly" (98). Experiments 1 and 3 produced 98 and 94 exclusively round and yellow seeds, respectively. Experiment 2 produced 31 round yellow seeds, 26 round green seeds, 27 wrinkled yellow seeds, and 26 wrinkled green seeds. Experiment 4 produced 24 round yellow seeds, 25 round green seeds, 22 wrinkled yellow seeds, and 27 wrinkled green seeds. Mendel noted: "There could scarcely be now any doubt of the success of the experiment; the next generation must afford the final proof" (98).

Mendel sowed all of the seeds obtained in the first experiment, and 90 plants from 98 seeds bore fruit. In the third experiment, 87 plants from 94 seeds bore fruit.[20] Mendel reported on his other results:

> In the second and fourth experiments the round and yellow seeds yielded plants with round and wrinkled yellow and green seeds, *AaBb*.
>
> From the round green seeds plants resulted with round and wrinkled green seeds, *Aab*.
>
> The wrinkled yellow seeds gave plants with wrinkled yellow and green seeds, *aBb*.
>
> From the wrinkled green seeds plants were raised which yielded again only wrinkled green seeds, *ab*. (98)

Mendel's results are also shown in tables 1.6 and 1.7. He concluded, "In all the experiments, therefore, there appeared all the forms which the proposed theory demands, and they came in nearly equal numbers" (99).

TABLE 1.6 Mendel's results from the gametic experiments 1 and 3 (Mendel 1865, 98)

1st Exp.	*3rd Exp.*		
20	25	round yellow seeds	*AB*
23	19	round yellow and green seeds	*ABb*
25	22	round and wrinkled yellow seeds	*AaB*
22	21	round and wrinkled yellow and green seeds	*AaBb*

TABLE 1.7 Mendel's results from the gametic experiments 2 and 4 (Mendel 1865, 99)

2nd Exp.	*4th Exp.*			
31	24	plants of the form		*AaBb*
26	25	"	"	*AaB*
27	22	"	"	*aBb*
26	27	"	"	*ab*

TABLE 1.8 Mendel's results for the flower color-stem length experiments (Mendel 1865, *100*)

Class	*Color of flower*	*Stem*	
1 [*AaBb*]	violet-red	long	47 times
2 [*aBb*]	white	long	40 "
3 [*Aab*]	violet-red	short	38 "
4 [*Ab*]	white	short	41 "

TABLE 1.9 Mendel's subsequent results for the flower color-stem length experiments (from Mendel 1865, *100*)

Trait	*Number*
violet-red flower color (*Aa*)	85 plants
white flower color (*a*)	81 plants
long stem (*Bb*)	87 plants
short stem (*b*)	79 plants

Mendel conducted a second set of experiments to test his assumptions. For these trials, he made selections so that each character should occur in half the plants if his assumptions were correct. In these experiments, *A* conferred violet-red flowers, *a* conferred white flowers, *B* long stems, and *b* short stems. He fertilized *Ab* (violet-red flowers, short stem) with *ab* (white flowers, short stem) producing hybrid *Aab*. In addition, *aB* (white flowers, long stem) was also fertilized with *ab*, yielding hybrid *aBb*. In the second year, the hybrid *Aab* was used as the seed plant and hybrid *aBb* as pollen plant. This should produce the combinations *AaBb*, *aBb*, *Aab*, and *ab*. In the third year, half the plants would have *Aa* (violet-red flowers), half *a* (white flowers), half *Bb* (long stems), and half *b* (short stems). The results are shown in tables 1.8 and 1.9. Mendel modestly concluded, "The theory adduced is therefore satisfactorily confirmed in this experiment also" (*100*). Mendel also performed other experiments, with fewer plants, on pod shape, pod color, and flower position, and "results obtained in perfect agreement" (*100*). No numerical data were presented.

As a result of this research, Mendel deduced, "Experimentally, therefore, the theory is confirmed that *the pea hybrids form egg and pollen cells which, in their constitution, represent in equal numbers all constant forms which result from the combination of characters united in fertilisation*" (*100*). He also stated, "It was furthermore shown by the whole of the experiments that it is perfectly immaterial whether the dominant character belong to the seed-bearer or to the pollen-parent; the form of the hybrid remains identical in both cases" (*84*).[21]

In discussing his results, Mendel demonstrated that he understood, at least qualitatively, the statistical nature of his data. He stated:

> This represents the average results of the self-fertilisation of the hybrids when two differentiating characters are united in them. In individual flowers and in individual plants, however, the ratios in which the forms of the series are produced

may suffer not inconsiderable fluctuations. Apart from the fact that the numbers in which both sorts of egg cells occur in the seed vessels can only be regarded as equal on the average, it remains purely a matter of chance which of the two sorts of pollen may fertilise each separate egg cell. For this reason the separate values must necessarily be subject to fluctuations, and there are even extreme cases possible, as were described earlier in connection with the experiments on the form of the seed and the colour of the albumen. The true ratios of the numbers can only be ascertained by an average deduced from the sum of as many single values as possible; the greater the number the more are merely chance effects eliminated. (*102*)

All of Mendel's numerical data from his pea experiments have now been presented, and these are the data on which Fisher based his analysis.

Mendel's Experiments on Other Species

Mendel also reported several experiments on *Phaseolus* (beans). The experiments on *Phaseolus vulgaris* and *Phaseolus nanus* "gave results in perfect agreement" (*103*). Those with *Phaseolus nanus*, L., as the seed plant, and *Phaseolus multiflorus*, W., as the pollen plant, did not. The former had white flowers and small white seeds, whereas the latter had purple-red flowers and seeds with black flecks or splashes on a peach-blood-red background. Mendel reported that the hybrids more closely resembled the pollen plant. He obtained only a few plants but, within limited statistics, he found that for recessive plant characters such as axis length and the form of the pod were the ratio of recessive to dominant was 1 : 3.

Mendel summarized his work as follows.

Despite the many disturbing factors with which the observations had to contend, it is nevertheless seen by this experiment that the development of the hybrids, with regard to those characters which concern the form of the plants, follows the same laws as in *Pisum*. With regard to the colour characters, it certainly appears difficult to perceive a substantial agreement. Apart from the fact that from the union of a white and a purple-red colouring a whole series of colours results [in F_2], from purple to pale violet and white, the circumstance is a striking one that among thirty-one flowering plants only one received the recessive character of the white colour, while in *Pisum* this occurs on the average in every fourth plant. (*105*)

Thus, Mendel not only reported blending inheritance, but also results that disagreed with his previous experiments.

Mendel also conducted experiments on *Hieracium* (hawkweed) (Mendel 1870). Again, the results did not always agree with those he had obtained previously. He remarked on the difficulty of the experiments and that he had obtained very few hybrids.

If finally we compare the described results, still very uncertain, with those obtained by crosses made between forms of *Pisum*, which I had the honor of communicating in the year 1865, we find a very real distinction. In *Pisum* the hybrids, obtained from the immediate crossing of two forms, all have the same type, but their posterity, on the contrary, are variable and follow a definite law in their variations. In *Hieracium* according to the present experiment the exactly opposite phenomenon seems to be exhibited. (qtd. in Stern and Sherwood 1966, 55)[22]

Summary

There are several points worth noting about Mendel's paper that will be important in the discussion of the Mendel-Fisher controversy. The first is that, as he remarks on several occasions, Mendel did not publish all of his data. The published data, however, also include results that differ considerably from Mendel's expectations. Mendel also knew what results he expected, either from theory or from his early observations. It also seems clear that Mendel had a good understanding of the principles of segregation and of independent assortment that form the basis of modern genetics.

Fisher's Analysis of Mendel's Data

Fisher's Early Thoughts

Although it was not until 1936 that R. A. Fisher published the paper on Mendel that would engender the longstanding controversy, that paper was not his first comment on Mendel's results. In a 1911 talk given to the Cambridge University Eugenics Society, Fisher commented, "It is interesting that Mendel's original results all fall within the limits of probable error;[23] if his experiments were repeated the odds against getting such good results is about 16 to one. It may just have been luck; or it may be that the worthy German abbot, in his ignorance of probable error, unconsciously placed doubtful plants on the side which favoured his hypothesis" (qtd. in Norton and Pearson 1976, 160). Fisher later changed his mind and attributed these results to the work of an assistant.

Fisher, in all probability, based these early comments on the analysis of Mendel's results provided by W. F. R. Weldon (1902). Weldon thought Mendel's work quite interesting and, in a letter to Karl Pearson, wrote, "About pleasanter things I have heard of and read a paper by one, Mendel, on the results of crossing peas, which I think you would like to read" (qtd. in Froggatt and Nevin 1971, 13). In his comments on Mendel, Weldon discussed Mendel's results on the 3:1 ratio in the first generation bred from hybrids. He presented Mendel's data along with the deviation of obser-

TABLE 1.10 Individuals with dominant characters in the second hybrid generation (Weldon 1902, 233)

Characters crossed	*Individuals of second hybrid generation*	*Number of dominant individuals*	*Dominant individuals on Mendel's theory*	*Probable error of theory*	*Deviation of observation from theory*
1. (Shape of seeds)	7324	5474	5493	±24.995	−19
2. (Color of cotyledons)	8023	6022	6017.25	±26.160	+4.75
3. (Color of seed coats)	929	705	696.75	±8.902	+8.25
4. (Shape of pod)	1181	882	885.75	±10.037	−3.75
5. Color of pod)	580	428	435	±7.034	−7
6. (Distribution of flowers)	858	651	643.5	±8.555	+7.5
7. (Height of plant)	1064	787	798	±9.527	−11

vation from theory along with a calculation of the probable error (table 1.10). He remarked:

Here are seven determinations of a frequency which is said to obey the law of Chance. Only one determination has a deviation from the hypothetical frequency greater than the probable error of the determination, and one has a deviation sensibly equal to the probable error; so that a discrepancy between the hypothesis and the observations which is greater to or equal to the probable error occurs twice out of seven times, and deviations much greater than the probable error do not occur at all. These results then accord so remarkably with Mendel's summary that if they were repeated a second time, under similar conditions and on a similar scale, the chance that the agreement between observation and hypothesis would be worse than that actually obtained is about 16 to 1. (Weldon 1902, 233)

Weldon also commented on Mendel's experiments on the 2:1 ratio and noted, "Mendel's statement is admirably in accord with his experiment" (Weldon 1902, 234). He then went on to discuss the results of the trifactorial experiment and commented, "Applying the method of Pearson (No. 25)[24] [χ^2 analysis] the chance that a system will exhibit deviations as great or greater than these from the result indicated by Mendel's hypothesis is about 0.95, or if the experiment were repeated a hundred times, we should expect to get a worse result about 95 times, or odds against a result as good as this or better are 20 to 1" (235). This was one of the early uses, perhaps even the first use, of the χ^2 test.

Weldon did not comment further in his paper on the goodness of fit of Mendel's data to his expectations, nor did he give even the slightest hint that he believed that Mendel's results were fraudulent in any way.[25] In a letter to Karl Pearson of November 1901, however, Weldon wrote: "Remembering his shaven crown [an allusion to Mendel's status as a monk] I cannot help wondering if they [Mendel's results] were not too good" (qtd. in Magnello 2004, 23). This line was crossed out and followed by the statement, "I do not see that the results are so good as to be suspicious." This was, in all probability, the first suggestion that Mendel's data were "too good." When Weldon wrote again to Pearson on 28 November 1901, he stated that he was certain that Mendel "cooked his figures, but that he was *substantially* right" (qtd. in Magnello 2004, 23).

In his 1902 paper, Weldon did comment further on both the value of Mendel's work and on some difficulties with Mendel's conclusions:

> Mendel's experiments are based upon work extending over eight years. The remarkable results obtained are well worth even the great amount of labour they must have cost, and the question at once arises, how far the laws deduced from them are of general application. It is almost a matter of common knowledge that they do not hold for all characters, even in Peas, and Mendel does not suggest that they do. At the same time I see no escape from the conclusion that they do not hold universally for the characters of Peas which Mendel so carefully describes. In trying to summarise the evidence on which my opinion rests, I have no wish to belittle the importance of Mendel's achievement. I wish simply to call attention to a series of facts which seem to suggest fruitful lines of inquiry. (Weldon 1902, 235)

The rest of Weldon's paper is devoted to a discussion of some of the evidence for his reservations about Mendel's work.[26]

Fisher's Seminal Paper

In 1936, R. A. Fisher published a paper entitled "Has Mendel's Work Been Rediscovered?" (*117*). This is the paper that engendered, albeit after a considerable delay, the so-called Mendel-Fisher controversy. Fisher did not question whether people knew of Mendel's work, but rather whether they really understood what Mendel had written. He noted that the story of Mendel's work and its rediscovery had become traditional in the teaching of biology: "A careful scrutiny can but strengthen the truth in such a tradition, and may serve to free it from such accretions as prejudice or hasty judgment may have woven into the story" (*117*). Fisher proposed to provide such a careful scrutiny and remarked, "When the History of Science is taken seriously the number of enquiries which such a story suggests is somewhat formidable. We want to know first: What did Mendel discover? How did he discover it? And what did he think he discovered? Next, what was the relevance of his discoveries to the science of his time, and what was its reaction to them?" (*118*).

Fisher was concerned that misconceptions about Mendel's work had been propagated by Bateson, particularly claims that Darwinism was responsible for the neglect of Mendel's work and that Mendel was hostile to Darwinism. Fisher presented persuasive arguments against both these views. He was also concerned about Bateson's assertion that Mendel's description of his experiments should not be taken literally. Bateson, in commenting on the monohybrid experiments, stated: "This statement of Mendel's in the light of present knowledge is open to some misconception. Though his work makes evident that such varieties may exist, it is very unlikely that Mendel could have had seven pairs of varieties such that the members of each pair differed from each other in *only* one considerable character (*wesentliches Merkmal*). The point is probably of little theoretical or *practical consequence, but a rather heavy stress is thrown on 'wesentlich'*" (Bateson 1909, 332).[27] Fisher proposed two possible solutions to this problem. Mendel might have arbitrarily chosen one factor for which the particular cross was designated as an experiment and ignored other factors; or he might have scored each plant in all factors and assembled the data for that factor from all of the crosses in which it had been involved and reported the result as a single experiment on a single factor. Fisher noted that the first solution seemed incredibly wasteful of data, but added, "This objection is not so strong as it might seem, since it can be shown that Mendel left uncounted, or at least unpublished, far more material than appears in his paper" (*121*). Fisher believed that the second option was what most modern geneticists would do, but thought it unlikely that Mendel had done so: "[T]he style throughout suggests that he [Mendel] expects to be taken literally; if his facts have suffered much manipulation the style of his report must be judged disingenuous. Consequently, unless real contradictions are encountered in reconstructing his experiments from his paper, regarded as a literal account, this view must be preferred to all alternatives, even though it implies that Mendel had a good understanding of the factorial system, and the frequency ratios which constitute his laws of inheritance, before he carried out the experiments reported in his first and chief paper" (*122*).

Fisher's Reconstruction of Mendel's Data

As far as the subsequent controversy is concerned, the most important section of Fisher's paper is the one entitled "An Attempted Reconstruction." Fisher constructed a chronology of the eight years of Mendel's experiments, including which experiments were done and how many plants were grown in a given year, what Mendel's results were, and in what order those results were obtained.

Fisher inferred that the experiments on seed characters (yellow or green and round or wrinkled) were completed in 1859 and that "Mendel does not test the significance of the deviation, but states the ratios as 2.96 : 1 and 3.01 : 1, without giving any probable error" (*123*). He went on to remark, "The discovery, or demonstration, whichever it may have been, of the 3 : 1 ratio was evidently the critical point in Mendel's researches" (*124*). Fisher believed that Mendel's satisfaction with these approximate ratios was intelligible if "he had convinced himself as to their explanation, and framed the entire Mendelian theory of genetic factors and gametic segregation" (*124*). He further noted:

> In 1930,[28] as a result of a study of the development of Darwin's ideas, I pointed out that the modern genetical system, apart from such special features as dominance and linkages, could have been inferred by any abstract thinker in the middle of the nineteenth century if he were led to postulate that inheritance was particulate, that the germinal material was structural, and that the contributions of the two parents were equivalent. I had at that time no suspicion that Mendel had arrived at his discovery in this way. From an examination of Mendel's work it now appears not improbable that he did so and that his ready assumption of the equivalence of the gametes was a potent factor in leading him to his theory. *In this way his experimental programme becomes intelligible as a carefully planned demonstration of his conclusions.* (*125*, emphasis added)

In other words, Fisher believed that Mendel was, in fact, a Mendelian.[29]

Fisher went on to discuss Mendel's experiments of 1860 in which the 3 : 1 ratio was shown to be 1 : 2 : 1, where 1 is the homozygous dominants or recessives and 2 is the heterozygous hybrid. On several occasions, Fisher commented on the comparison of the observed deviations from the expected results to the standard deviation expected. Thus, in discussing the experiments on plants raised from yellow seeds (which yielded 166 plants with only yellow seeds and 353 plants with both yellow and green seeds) and that on plants grown from round seeds (which yielded 193 plants with only round seeds and 372 plants with both round and wrinkled seeds), Fisher stated: "The ratios in both cases show deviations from the expected 2 : 1 ratio less than their standard errors" (*126*). For the 1861 experiments on plants bred from colored flowers and from tall plants (see table 1.3), Fisher commented, "In neither case does the ratio depart significantly from the 2 : 1 ratio expected, although in the second case the deviation does exceed the standard deviation of random sampling" (*126*). For the experiment on yellow pods (which yielded a 60 : 40 ratio), Fisher remarked on "a relatively large, but not a significant, deviation" (*127*). He further noted, "It is remarkable as the only case in the record in which Mendel was moved to verify a ratio by repeating the trial" (*127*).

TABLE 1.11 Classification of plants grown in the trifactorial experiment (Fisher 1936, table II)

	CC				*Cc*				*cc*				*Total*			
	AA	*Aa*	*aa*	Total	*AA*	*Aa*	*aa*	Total	*AA*	*Aa*	*aa*	Total	*AA*	*Aa*	*aa*	Total
BB	8	14	8	30	22	38	25	85	14	18	10	42	44	70	43	157
Bb	15	49	19	83	45	78	36	159	18	48	24	90	78	175	79	332
bb	9	20	10	39	17	40	20	77	11	16	7	34	37	76	37	150
Total	32	83	37	152	84	156	81	321	43	82	41	166	159	321	159	639

Fisher, obviously concerned, went on to critically examine the experiments in which such deviations occurred. It was at this point that he first announced the problem of the 2:1 ratio:

> In connection with these tests of homozygosity by examining ten offspring formed by self-fertilization, it is disconcerting to find that the proportion of plants misclassified by this test is not inappreciable. If each offspring has an independent probability, .75, of displaying the dominant character, the probability that all ten will do so is $(.75)^{10}$ or .0563. Consequently, between 5 and 6 per cent of the heterozygous parents will be classified as homozygotes, and the expected ratio of segregating to non-segregating families is not 2:1 but 1.8874:1.1126 or approximately 377.5:222.5 out of 600. Now among the 600 plants tested by Mendel 201 were classified as homozygous and 399 as heterozygous [see table 1.3]. Although these numbers agree extremely closely with his expectations of 200:400, yet, when allowance is made for the limited size of the test progenies, the deviation is one to be taken seriously. It seems extremely improbable that Mendel made any such allowance, or that the numbers he recorded are "corrected" values, rounded off to the nearest integer, obtained by dividing the numbers observed to segregate by .9437. We might suppose that sampling errors in this case caused a deviation in the right direction, and of almost exactly the right magnitude, to compensate for the error in theory. A deviation as fortunate as Mendel's is to be expected once in twenty-nine trials. Unfortunately the same thing occurs again with the trifactorial data [table 1.11] (*127*).

Fisher's further examination of those trifactorial data yielded detailed comments that are also worth examining.

> In the case of the 600 plants tested for homozygosity in the first group of experiments Mendel states his practice to have been to sow ten seeds from each self-fertilized [F_2] plant. In the case of the 473 plants with coloured flowers from the trifactorial cross he does not restate his procedure. It was presumably the same as before. As before, however, it leads to the difficulty that between 5 and 6 per cent of heterozygous plants so tested would give only coloured progeny, so that the expected ratio of those showing segregation to those not showing it is really

TABLE 1.12 Comparison of numbers reported with uncorrected and corrected expectations (Fisher 1936, table III)

	Number of plants tested	Number of non-segregating progenies observed	Number expected		Deviation	
			Without correction	Corrected	Without correction	Corrected
1st group of experiments	600	201	200.0	222.5	+1.0	−21.5
Trifactorial experiment	473	152	157.7	175.4	−5.7	−23.4
Total	1073	353	357.7	397.9	−4.7	−44.9

lower than 2:1, while Mendel's reported observations agree with the uncorrected theory.

The comparisons are shown in Table III [table 1.12]. A total deviation of the magnitude observed, and in the right direction, is only to be expected once in 444 trials; there is therefore here a serious discrepancy. (*130*)

The reliability of Mendel's results had been called into question.

Fisher then offered several possible solutions to the 2:1 ratio problem. He pointed out that if Mendel had backcrossed the 473 trifactorial plants, the probability of misclassification of heterozygotes would be reduced by a factor of 50. (This would have involved a considerable amount of labor.) If, for example, the plants were backcrossed with a recessive plant, then the probability of observing the recessive character in a single plant would be 0.5 for a heterozygote. For 10 plants, the probability of misclassification is then $(0.5)^{10}$, or 0.00098.

A second possibility was that Mendel had used a larger number of progeny in his test, say 15 instead of 10. The probability of misclassification, in this case, is reduced to 0.013, which gives a ratio of 1.974:1.026 = 1.924, much closer to 2. Fisher noted, however, that this would have required a larger number of plants grown in a single year than Mendel had, in fact, ever planted. In addition, it would not apply to the earlier experiments, in which Mendel had explicitly stated that he used 10 progeny.

The third possibility was that the selection of plants for testing favored the heterozygotes. Fisher remarked that in some crosses, it was possible that the heterozygote plants were larger and that "the larger plants might have been unconsciously preferred" (*131*). Fisher presented three arguments against this possible solution: (1) in the trifactorial experiment all plants were counted; (2) it was improbable that the compensating selection would work equally well for all five plant characters; and (3) the total compensation for all plants was unlikely to have given the exact number needed.

Fisher stated, however, that the question of whether the trifactorial data had been manipulated could be tested. He proceeded to use the χ^2 test (discussed further in this book's appendix), and commented:

> The possibility that the data for the trifactorial experiment do not represent objective counts, but are the product of some process of sophistication, is not incapable of being tested. Fictitious data can seldom survive a careful scrutiny, and, since most men underestimate the frequency of large deviations arising by chance, such data may be expected generally to agree more closely with expectation than genuine data would. The twenty-seven classes in the trifactorial experiment supply twenty-six degrees of freedom for the calculation of χ^2. The value obtained is 15.3224, decidedly less than its average value for genuine data, 26, though this value by itself might occur once in twenty genuine trials.[30] (*131*)

Fisher then applied the test to various subdivisions of the trifactorial data, with similar results, and subsequently applied the analysis to all of the experiments performed in 1863. These included the trifactorial experiment, the bifactorial experiment, the experiment on gametic ratios (those involving the question of whether the results depended on which plant produced the pollen and which the egg), and the repetition of the yellow-pod experiment. His results, shown in table 1.13, gave a χ^2 of 15.5464 for 41 degrees of freedom, and prompted him to write: "The discrepancy is strongly significant, and so low a value could scarcely occur by chance once in 2000 trials. There can be no doubt that the data from the later years of the experiment have been biased strongly in the direction of agreement with expectation" (*132*).

Fisher explained that in tests where seeds were deformed or discolored, bias rather than theory might help to explain Mendel's results, but also noted that this would not apply to the tests of gametic ratios or to other experiments based on classification of whole plants.

TABLE 1.13 Measure of deviation expected and observed in 1863 (Fisher 1936, table IV)

	Expectation	*χ^2 observed*
Trifactorial experiment	17	8.9374
Bifactorial experiment	8	2.8110
Gametic ratios	15	3.6730
Repeated 2 : 1 test	1	0.1250
Total	41	15.5464

TABLE 1.14 Deviations expected and observed in all experiments (Fisher 1936, table V)

		Expectation	χ^2		Probability of exceeding deviations observed
3 : 1 ratios	Seed characters	2	0.2779		
	Plant characters	5	1.8610		
		— 7	—	2.1389	.95
2 : 1 ratios	Seed characters	2	0.5983		
	Plant characters	6	4.5750		
		— 8	—	5.1733	.74
Bifactorial experiment		8		2.8110	.94
Gametic ratios		15		3.6730	.9987
Trifactorial experiment		26		15.3224	.95
Total		64		29.1186	.99987
Illustrations of plant variation		20		12.4870	.90
Total		84	41.6056		.99993

While Fisher did collect the χ^2 values for all the experiments in his Table V (here, table 1.14), the analysis of all of Mendel's results seems to be almost an afterthought. It was the agreement of Mendel's data with what Fisher regarded as the incorrect 2:1 ratio that was most important for Fisher.[31] Fisher's result for all of Mendel's experiments, which included 84 degrees of freedom, was a total χ^2 of 41.6056. The probability of exceeding this χ^2 value was 0.99993, or the probability of getting such a good fit to the expectation was 7 in 100,000. Fisher makes no comment on this extraordinary result except to note that "the bias seems to pervade the whole of the data" (*133*).

Fisher's Conclusions

Fisher's first conclusion (and the one undoubtedly most important to him) was that Mendel's account of his experiments was "to be taken entirely literally." Fisher fully believed that Mendel's experiments "were carried out in just the way and much in the order that they are recounted. The detailed reconstruction of his programme on this assumption leads to no discrepancy whatever" (*134*). Bateson and others who had suggested otherwise were, in Fisher's view, conclusively refuted. However, Fisher went on to state explicitly that he believed that most of Mendel's data has been falsified:

> A serious and almost inexplicable discrepancy has, however, appeared, in that in one series of results the numbers observed agree excellently with the two to one ratio, which Mendel himself expected, but differ significantly from what should have been expected had his theory been corrected to allow for the small size of his test progenies. To suppose that Mendel recognized this theoretical complication, and adjusted the frequencies supposedly observed to allow for it, would be to contravene the weight of the evidence supplied in detail by his paper as a whole. Although no explanation can be expected to be satisfactory, it remains a possibility among others that Mendel was deceived by some assistant who knew too well what was expected. This possibility is supported by independent evidence that the data of most, if not all, of the experiments have been falsified so as to agree closely with Mendel's expectations. (*134*)

Fisher concluded that Mendel regarded the numerical ratios as demonstrating the truth of his factorial system "and that he was never much concerned to demonstrate either their exactitude or their consistency" (*135*). Perhaps as a way of rationalizing *why* Mendel might falsify his data, Fisher wrote, "it is clear, from the form his experiments took, that he knew very surely what to expect, and designed them as a demonstration for others rather than for his own enlightenment" (*135*). Yet Fisher clearly attributes the falsification to someone else, such as an assistant. The last section of Fisher's paper is devoted to examining how Mendel's contemporaries reacted to Mendel's work and why they largely overlooked it. Fisher remarked that the journal in which Mendel published his results was widely distributed and reasonably well known. Moreover, Fisher noted, Mendel's paper was not inaccessible; in fact, "the new ideas are explained most simply, and amply illustrated by the experimental results" (*136*). Yet Karl Wilhelm von Nägeli, with whom Mendel corresponded, was either unimpressed by Mendel's results or anxious to warn students against paying attention to them. Fisher also cited W. O. Focke, "who, in his *Pflanzenmischlinge* [1881], makes no less than fifteen references to Mendel" (*137*). As Fisher made clear, however, Focke did not understand Mendel's work and seemed to prefer the more comprehensive contributions of Joseph Gottlieb Kolreuter, Carl Friedrich von Gartner, and others. Fisher remarked that Focke had "overlooked, in his chosen field, experimental researches conclusive in their results, faultlessly lucid in presentation, and vital to the understanding not of one problem of current interest, but of many" (*139*). It is hard to imagine a more positive opinion of Mendel's work. There is no mention here of any falsification or fraud.

Fisher ended his paper with an exhortation to, and criticism of, his colleagues for failing to carefully examine and understand Mendel's work:

> The peculiar incident in the history of biological thought, which it has been the purpose of this study to elucidate, is not without at least one moral—namely, that there is no substitute for a careful, or even meticulous, examination of all original papers purporting to establish new facts. Mendel's contemporaries may be blamed for failing to recognize his discovery, perhaps through resting too great a confidence on comprehensive compilations. It is equally clear, however, that since 1900, in spite of the immense publicity it has received, his work has not often been examined with sufficient care to prevent its many extraordinary features being overlooked, and the opinions of its author being misrepresented. Each generation, perhaps, found in Mendel's paper only what it expected to find . . . (139)

It is clear that Fisher admired both Mendel and his work.

Fisher's Later Thoughts

Even before Fisher's paper was published, he discussed its contents with E. B. Ford, then departmental demonstrator in the Department of Zoology at Oxford University. In a letter of 2 January 1936, Fisher wrote: "I have had the shocking experience lately of coming to the conclusion that the data given in Mendel's paper must be practically all faked"; he referred to this information as his "abominable discovery" (R. A. Fisher Digital Archive, 2, 4). Just a few days later, on 8 January 1936, Fisher wrote to Douglas McKie, the editor of *Annals of Science*, to submit his paper for publication. In this letter, Fisher stated: "I had not expected to find the strong evidence which has appeared that the data had been cooked. This makes my paper far more sensational than ever I had intended" (R. A. Fisher Digital Archive, 1).

In the letter to Ford, Fisher emphasized, however, that he had concluded that Mendel's experiments were planned and performed exactly as Mendel had recorded and "this is what I was really studying his paper for" (2). He further remarked, "I don't believe that this touches Mendel's own *bona-fides*, or the reality of the experiments he carried out" (3).

Ford was horrified. In his response to Fisher of 5 January 1936, he exclaimed: "I am appalled by your discovery. Your analysis is a remarkable piece of work, but what it reveals is really very shocking. Clearly, as you say, Mendel himself is not to blame" (R. A. Fisher Digital Archive, 3). Ford, like Fisher, refused to believe that Mendel was guilty of fraud: "[I]t is simply incredible that a man of his intelligence could want to fake, *after* he had found out what to look for by honest work" (4). Ford regarded Fisher's paper as extremely important and encouraged its publication. "Too much has been hung on Mendel's results to suppress the matter though one does not want to wash dirty linen in public unnecessarily" (5). In a subsequent letter of 11 January, Ford further noted that Fisher had dealt with "the difficult matter of the faked data in a most tactful way" (R. A. Fisher Digital Archive, 1). He suggested that Fisher might include a summary of the paper because "You see you have naturally had to be so careful in your statements on the faking, that the point seems rather immersed in the paper as a whole. A summary would obviate this" (1–2). Fisher did not act on this suggestion.

It is clear both from Fisher's paper itself and from Ford's comment that the statements concerning Mendel's alleged falsification were not emphasized in the paper. I might suggest that this was more than tact—that it represented Fisher's view that the numerical falsification, if it was that, was relatively less important than Mendel's conclusions. This view seems to have been shared by Ford, who wrote in his letter of 5 January 1936: "When you write, do stress that Mendel's greatness lies not so much in his discoveries as in his deductions—and in planning the work" (4). Fisher clearly agreed. There doesn't seem to be any further Fisher correspondence on Mendel's data. J. Henry Bennett's book (1983), which contains selections from Fisher's letters, does not include any such correspondence.

Fisher did include a discussion of the problems with Mendel's data in his university lectures. Alan Cock reports that in Fisher's lectures in 1942–1943, he presented the following anecdote: "A lay brother assigned the task of weeding Mendel's pea beds, is grumbling to himself about his bad back. 'Brother Gregor and his experiments are all very well, but my back is killing me.' Knowing from experience that if the result were 'bad,' Brother Gregor might insist in repeating the experiment, he 'accidentally' lets his hoe slip to demolish a few chosen plants and ensure a 'good' result" (A. Cock, letter to A. W. F. Edwards, 25 March 1988).

In 1955, at the request of an editor of a proposed series on source papers in science, Fisher wrote both an introduction to, and a commentary with marginal notes on, Mendel's paper. The series never came about, but Fisher's introduction and commentary (Fisher 1965a, 1965b) were included in a book edited by Bennett (1965), containing a translation of Mendel's paper. Fisher's introduction again emphasized the quality and importance of Mendel's work as well as Mendel's contribution to the methodology of research on hybrids. Commenting on the "rediscovery" of Mendel's work in 1900, Fisher wrote, "The facts available in 1900 were at least sufficient to establish Mendel's contribution as one of the greatest experimental advances in the history of biology" (Fisher 1965a, 2).

Fisher then discussed Mendel's method:

If we read his introduction literally we do not find him expressing the purpose of solving a great problem or reporting a resounding discovery. He represents his work rather as a contribution to the *methodology* of research into plant inheritance. He had studied the earlier writers and tells us just in what three respects he thinks their work should be improved upon. If proper care were given, he suggests, to the distinction between generations, to the identification of genotypes, and, to this end, to the frequency ratios exhibited by their progeny, when based on an adequate statistical enumeration, studies in the inheritance of other organisms would yield an understanding of the hereditary process as clear as that which he here exhibits for the varieties of garden pea. There is no hint of a tendency to premature generalization, but an unmistakable emphasis on the question of method. (Fisher 1965a, 3)

Here, using the very same language that he had used in his 1936 paper, Fisher emphasized Mendel's reasoning:

The fact that Mendel was principally concerned to justify a method of investigation, and not primarily to exhibit particular results, is at least a partial explanation of another group of peculiarities of his paper, which flow from the fact that he is reporting a *carefully planned demonstration, rather than the protocol of the first observations which led to the formation of his ideas.* The simplicity of his plan, and the adequacy of the numbers of the first crosses reported, are indications that he knew in advance very much what he intended to do, and what he ought to expect. He constantly omits reference to the confirmation of his first conclusions, which the later generations and other experiments reported must have supplied in abundance. Only once is he led to repeat a test. He seems never to be unsure of the sufficiency of the first evidence reported, even when it is not really so strong as might be wished . . . (Fisher 1965a, 4, emphasis added)

In his marginal notes, Fisher again stated that all of Mendel's data were not reported: "It is remarkable how much of the material, which Mendel must have bred, has not been reported, either in confirmation, or for comparison with what he has given" (Fisher 1965b, 54).

Fisher's introduction made no mention at all either of the problem of the 2:1 ratio or of the overall goodness of fit of Mendel's results. In the marginal notes, only the 2:1 ratio was emphasized. Fisher once again pointed out that if Mendel's method was as reported, the 2:1 ratio should have been 1.8874:1.1126. He noted that the problem appeared in both the monohybrid experiments and in the trifactorial experiment. He did, however, remark that an examination of the general level of agreement between Mendel's expectations and his reported results showed that it is closer than would be expected in the best of several thousand repetitions. "The data have evidently been sophisticated systematically" (Fisher 1965b, 53). Again Fisher placed the blame on a deceiving assistant.

Thus, it is evident that Fisher admired Mendel's achievement and believed that Mendel's most important contribution was his method of experimentation and his deductions from his results and not the numerical results themselves. Fisher also concluded that much of Mendel's data had been falsified. He considered the most important falsification to have been on the 2:1 ratio experiments and believed that the overall "goodness of fit" was less important. It also seems clear that Fisher believed that Mendel's achievements outweighed any possible falsification.

The Mendel-Fisher Controversy

Just as Mendel's work was neglected, so was Fisher's analysis of Mendel's work. This latter neglect lasted for more than 25 years. As discussed earlier, there seems to be no correspondence between Fisher and others on the subject after 1936. Similarly, Vítězslav Orel (1996), a biographer of Mendel, mentions no papers on the controversy published before 1964. Nor do any papers on the controversy published after 1964 make any reference to papers published before that date except, of course, to Fisher's paper.[32] One may speculate that the reason for this neglect was the lack of emphasis on Mendel's possible fraud in Fisher's paper.

The controversy divides itself reasonably into three time periods: (1) the 1960s, about the time of the Mendel centenary; (2) the late 1970s to the mid-1980s; and (3) from 1990 to the present. Many of the papers on the controversy include discussions of other aspects of Mendel's work. I will restrict my discussion to only those sections that deal with Mendel and Fisher. I will also attempt to keep to a chronological account. If there are errors in a paper, I will delay discussion of them until I come to the time in which those criticisms appeared in the published literature. In this way, the reader will get a better feel for the actual history of the controversy.

The 1960s

The year 1965 marked the centenary of Mendel's discovery[33] and, as with other significant scientific discoveries, the occasion was commemorated with conferences, papers, and books. All of these celebrated Mendel's achievement and the evaluations were unanimously and enthusiastically favorable. This attention to Mendel's work also seems to have resulted in attention to Fisher's 1936 analysis paper. This provided the initial impetus for the Mendel-Fisher controversy. In a talk given to the New Jersey Academy of Science on 20 April 1963, published in the *Journal of Heredity*, Conway Zirkle observed that Mendel's results seemed to fit his expectations too well: "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 a little *too good*, better than we would have a right to expect on the basis of chance. Could the good Father Mendel have fudged his results a little?" (Zirkle 1964, 66, emphasis added). Interestingly, Zirkle made no mention of Fisher's 1936 paper, although he cited several other works by Fisher. This seems to have been the first published mention that Mendel's results were "too good." Zirkle also remarked that if Mendel's results had not been so good "he might never have discovered Mendelism" (66).[34] He illustrated this with a story concerning Darwin's experiments on corn. In examining two types of grain on an ear of corn, Darwin found a ratio of 88:37, which differed only slightly from the 94:31 ratio expected for a 3:1 ratio.[35] Darwin missed its significance. Zirkle concluded, "Mendel, perhaps, was lucky, but was he?" (66). With such statements, Zirkle seemed to acknowledge the possibility that Mendel falsified his data.

In 1964, Gavin de Beer published a paper on Mendel, Darwin, and Fisher in which he also constructed a chronology of Mendel's work, slightly correcting Fisher's chronology. De Beer, like Fisher, was an admirer of Mendel. In discussing Mendel's paper, he stated: "It is not too much to claim that this communication, which was the foundation of the science of genetics and introduced mathematics and the theory of probability into the study of inheritance, represented an advance of knowledge in biological science of an order of magnitude comparable with that of evolution by natural selection" (DeBeer 1964, 192). He noted that it was now possible to appraise the true worth of Mendel's work: "This is largely due to the experiments, demonstrations, and conclusions of Sir Ronald Fisher, F.R.S., who, in 1930, brought out the full significance of Mendel's work" (192). De Beer clearly did not regard Fisher as a detractor of Mendel. He did, however, present a summary of Fisher's 1936 paper including Fisher's conclusion that Mendel's data had been falsified. For De Beer, however, as for Fisher, this conclusion had no effect on his admiration of Mendel: "None of this, of course, detracts in the slightest degree from the genius of Mendel in planning and carrying out his experiments, nor from the validity of the principles of heredity, his Laws, that he discovered" (200). De Beer also suggested that whoever had been responsible for Mendel's results was, in fact, a benefactor to science. He presented an analysis of Darwin's ear of corn experiment similar to Zirkle's analysis, noting that Darwin's observed deviation was "well within the acceptable limits for agreement with a 3:1 ratio" (201).

Another paper commemorating the Mendel centenary was by Leslie Clarence Dunn (1965). Dunn stated that there was a strong indication that Mendel knew what numerical results to expect, and that the agreement of Mendel's results with his expectations could not be accounted for by luck alone. He noted that this had first been pointed out by Fisher and that Fisher had concluded that fraud was involved. Dunn offered several possible explanations of Mendel's results, including the possibility, already mentioned by Fisher, that Mendel had tested more than 10 plants in the 2:1 ratio experiments:

> Dr. Sewall Wright has pointed out to me his view that Mendel, who clearly knew how to compute probabilities, could hardly have been unaware of the likelihood of no recessives would appear in some groups of ten progeny and could have estimated this to be about one in eighteen (0.056). Perhaps he chose the inadequate number ten because of lack of space for growing plants; but perhaps he in fact tested more than ten plants in order to have at least ten left after the inevitable losses. If the average of "at least ten" should be twelve the probability of misclassifying falls from 0.056 to 0.031 and the discrepancy from Mendel's 2:1 expectation is not a serious one.
>
> Those who have experience in tallying such outcomes become aware of the danger that unconscious bias in favor of the expected result will creep in and that the count may be stopped at a point which is favorable to the theory. (Dunn 1965, 194)

Although Dunn acknowledged the correctness of Fisher's analysis, he concluded: "There is no evidence of conscious fraud and he [Mendel] was careful to report wide deviations in some parts of some experiments which he would not have done if bent on fraud" (194). Dunn also believed that Mendel had his theory in mind "when the data *as reported* were tallied" (194). In his final assessment of Mendel, Dunn wrote: "Mendel however stands as a clear example and guide to a new way of studying a biological problem with a sharp, clear experimental design applied to a single question stated with simplicity because it had been reduced to its essentials" (198).

Bennett's aforementioned *Experiments in Plant Hybridisation by Gregor Mendel* (1965) included a translation of Mendel's paper along with Fisher's paper, introduction, and marginal notes. Bennett emphasized Fisher's "remarkable findings" on both the 2:1 ratio experiments and on the overall goodness of fit. This was stated on the first page of the editor's preface along with Fisher's conclusion that fraud had occurred and that this was due to Mendel's assistant.

Fisher and Mendel were also discussed in *A History of Genetics* by Alfred Sturtevant (1965). Sturtevant commented that Fisher had shown that if one examined all of Mendel's experiments, the probability of getting as good a fit as Mendel had was 1 in 14,000. Sturtevant did not seem to be overly impressed by this observation and offered the first technical criticism of Fisher's analysis.

> If this were all, one might not be too disturbed, for it is possible to question the logic of the argument that a fit is too close to expectation. If I report that I tossed 1000 coins and got exactly 500 heads and 500 tails, a statistician will raise his eyebrows, though this is the most probable exactly specified result. If I report 480 heads and 520 tails, the statistician will say that is about what one would expect—though this result is less probable than the 500:500 one. He will arrive at this by adding the probabilities for all results between 480:520 and 520:480, whereas for the exact agreement he will consider only the probability of 500:500 itself. If I now report that I tossed 1000 coins ten times, and got 500:500 every time, our statistician will surely conclude that I am lying, though this is the most probable result thus exactly specified. (Sturtevant 1965, 13)

Sturtevant was much more impressed with Fisher's analysis of the 2:1 ratio experiments because "In the present case, however, it appears that in one series of experiments Mendel got an equally close fit to a *wrong* expectation" (13–14). He remarked that Fisher's analysis was correct, but only if exactly 10 seeds were planted. For more than 10 seeds, as we have seen, the correction to the expectation is less and "Fisher's most telling point will be weakened" (14). He went on to note: "The statement by Mendel seems unequivocal, but the possibility remains that he may have used more than 10 seeds in some or many cases" (15).[36]

Sturtevant also examined eight experiments on cotyledon color in peas, including Mendel's experiment and seven others performed between 1900 and 1924, to see if they also reported unexpectedly close agreement with expectation (table 1.15). He noted that half the experiments showed a ratio of the observed deviation to the probable error greater than one, which is what one expects. He concluded, "The over-all impression is that the agreement with expectation is neither too good nor too poor" (16).

Sturtevant also offered a possible botanical explanation for Mendel's results. He stated that in self-pollination an anther will usually break at one point, leading not to a random sample of pollen grains but to one in which all or most of the pollen grains come from one or a few pollen-mother cells. He admitted that this was unlikely to be important in Mendel's experiments, and, "Calculations based on this improbable limiting assumption indicate that Fisher's conclusions would still hold good; but the point remains that in any such analysis one needs to examine the assumptions very carefully, to make sure there may not be some alternative explanation" (15).

TABLE 1.15 F_2 results, pea crosses (from Sturtevant 1965, 15)

Source	Yellow	Green	Total	Deviation from 3 in 4	Prob. error	Dev./P.E.
Mendel 1866	6,022	2,001	8,023	+.0024	±.0130	.18
Correns 1900	1,394	453	1,847	+.0189	±.0272	.70
Tschermak 1900	3,580	1,190	4,770	+.0021	±.0169	.12
Hurst 1904	1,310	445	1,775	−.0142	±.0279	.51
Bateson 1905	11,902	3,903	15,806	+.0123	±.0093	1.23
Lock 1905	1,438	514	1,952	−.0533	±.0264	2.04
Darbishire 1909	109,060	36,186	145,246	+.0035	±.0030	1.16
Winge 1924	19,195	6,553	25,748	−.0180	±.0125	1.44

Despite his criticisms of Fisher, Sturtevant concluded, "In summary, then, Fisher's analysis of Mendel's data must stand essentially as he stated it" (16). He offered three possible explanations for Mendel's data: (1) unconscious bias, (2) omission of aberrant families, or (3) biased action by an assistant. He admitted, "None of these alternatives is wholly satisfactory, since they seem out of character, as judged by the whole tone of the paper" (16). Sturtevant ended his analysis by stating: "Perhaps the best answer—with which I think Fisher would have agreed—is that, after all, Mendel was right!" (16).

Another translation of Mendel's paper appeared in *The Origin of Genetics: A Mendel Source Book*, edited by Curt Stern and Eva Sherwood (1966). The title gives the editors' judgment about the importance of Mendel's work. They stated: "Gregor Mendel's short treatise 'Experiments on Plant Hybrids' is one of the triumphs of the human mind. It does not simply announce the discovery of important facts by new methods of observation and experiment. Rather, in an act of the highest creativity, it presents these facts in a conceptual scheme which gives them general meaning. Mendel's paper is not solely a historical document. It remains alive as a supreme example of scientific experimentation and profound penetration of data" (v).

The book also included Fisher's 1936 analysis paper and a commentary on why Mendel's data were "too good" by Sewall Wright, a leading population geneticist. Wright remarked: "the excessive goodness of fit of Mendel's data is certainly one of the most disconcerting items that a historian of genetics has to deal with" (Wright 1966, 173). He reported that he had repeated Fisher's calculations and obtained substantially the same results and agreed: "There is no question that the data fit the ratios much more closely than can be expected from accidents of sampling" (173). Wright cited Raymond Pearl (1940) on the difficulty of making repeatable counts as evidence for such possible bias. Pearl reported an experiment in which 15 trained observers (two plant pathologists, two professors of agronomy, one professor of philosophy [originally trained as a biologist], four biologists, one computer, one practical corn breeder, and one professor and three assistants in plant physiology) were asked to count the same 532 kernels of corn from a single ear. The traits examined were color (yellow or white) and form of the kernel (starchy or sweet). The Mendelian predictions were nine yellow starchy; three yellow sweet; three white starchy; and one white sweet. Presumably each of these observers knew the results expected. Pearl's results are shown in table 1.16. He remarked, "It must be remembered that each individual handled, sorted, and counted *the same identical kernels of corn.* They were required to discriminate only with reference to the color and the form of each kernel. *Yet no two of the fifteen highly trained and competent observers agreed* as to the distribution of these 532 kernels" (Pearl 1940, 87).

Although Pearl's data show the difficulty of obtaining repeatable results, they do not reveal any subconscious bias on the part of the observers. Even though the observers presumably knew what to expect, the results show fluctuations well beyond what one might expect if these were different sets of kernels that were randomly distributed, and larger fluctuations than he expected for good observers observing the same kernels. These results would seem to argue that Mendel's excessively good fit to expectations *was* rather the result of some sort of "sophistication" or bias, not an unbiased observation.

Wright also considered the agreement in the case of the 2:1 ratio experiments as the most serious evidence of fraud but noted that Mendel had presented wide variations in his data that "would hardly have been reported by one bent on fraud" (174). He also suggested that Mendel's claim that he used only ten plants in such experiments should not be taken literally. He further suggested that it would take only very small changes in Mendel's data, caused by bias, to allow his results to agree so well with expectations. He concluded, "Taking everything into account, I am confident, however, that there was no deliberate effort at falsification" (175).

J. M. Thoday (1966) continued the discussion. He was an admirer of both Mendel and Fisher and remarked, "Fisher in a classic essay, which

TABLE 1.16 Showing the classification of the kernels of ear no. 8 by the different observers (from Pearl 1940, 87)

	Classes of kernels					
Observer	*Yellow starchy*	*Yellow sweet*	*White starchy*	*White sweet*	*Total starchy*	*Total sweet*
Mendelian expectation	299.25	99.75	99.75	33.25	399.00	133.00
I	352	102	52	26	404	128
II	322	49	82	79	404	128
III	298	75	108	51	406	126
IV	332	101	71	28	403	129
V	305	101	86	40	391	141
VI	313	100	90	29	403	129
VII	308	86	95	43	403	129
VIII	311	101	92	28	403	129
IX	327	101	78	26	405	127
X	308	92	95	37	403	129
XI	311	97	92	32	403	129
XII	313	99	91	29	404	128
XIII	308	97	95	32	403	129
XIV	312	104	91	25	403	129
XV	333	97	73	29	406	126
Totals	4753	1402	1291	534	6044	1936
Means	316.87	93.47	86.07	35.60	402.93	129.07

is a model that every would-be historian of science should be required to digest, made a thorough logical and statistical analysis of Mendel's paper" (122). Thoday noted the problems presented by both the overall fit to the data and by the 2:1 ratio experiments. He went on to offer a biological explanation for Mendel's results: "Diffident as I am in taking issue with one of Fisher's standing," Thoday remarked, he believed that there was still a question concerning the explanation of Mendel's data (122). He remarked that the data were too good only if one assumed that the distributions were in accord with binomial expansions.

In fact, a little reflection shows that they [the distributions] should not, for though egg cells no doubt come at random, pollen grains do not, they come in tetrads, and tetrads give exact ratios. Unless therefore the number of pollen grains is vastly in excess of the number of ovules and the many tetrads are thoroughly randomized, we expect ratios that are better than Fisherian. How much departure from Fisherian ratios we expect will depend on many factors in the biology of the particular organism. No doubt it also will vary with the particular conditions of growth and so on, and we cannot judge these for Mendel's peas in Mendel's garden a century ago. A recent paper by Dr. Ursula Philip has come up with exactly the same hypothesis to explain excessively good data with the sea weed fly (Coleopa frigida). (Thoday 1966, 123)[37]

If there is a reduced expected variance in peas, then Fisher's argument concerning the overall goodness of fit will no longer hold. A reduced expected variance results in a larger χ^2 (see this volume's appendix), thus making Mendel's results more probable on the basis of chance. This explanation does not, however, account for the very good agreement in the gametic ratio experiments in which the plants were artificially fertilized.

George Beadle (1967) made the same suggestion:

Pollen populations in a given anther or flower are definitely finite and produced in *exactly* a one-to-one ratio for a given gene pair, because each mother cell gives rise to four daughters, the two kinds always distributed in a two-to-two ratio. Furthermore, since pea flowers are self-fertilized before they fully open, the pollen grains that fertilize the flower that produces a given pea pod are very few in number and likely to come from an even smaller number of two-to-two quartets. Thus if one pea seed in a pod carries one allele of a pair from the pollen, the next one is more likely than not to carry the alternate allele. (Beadle 1967, 337)[38]

He further remarked that he and Sturtevant had "explored this possibility to see if it was sufficient to account for the apparent bias. It works in the right direction but is not sufficient, even if alternate forms of pollen grains are assumed to have functioned always in an exact one-to-one ratio" (337). No details of the calculation were presented. Beadle decided that the best explanation for Mendel's results was unconscious bias that resulted from Mendel stopping his count when the results looked good, remarking: "I therefore conclude that Mendel was very human, but not dishonest" (338).

In 1966, Franz Weiling began what would be a twenty-five-year defense of Mendel (Weiling 1965; 1966; 1971). Weiling did not believe that Mendel engaged in fraud and rejected Fisher's views that the falsification was done by an assistant and that Mendel's paper was a demonstration. In fact, Weiling denied that Mendel had an assistant. De Beer, on the other hand, named three such assistants, Alipius Winkelmayer, Josef Lindenthal, and Josef Maresch (De Beer 1964, 200).[39]

TABLE 1.17 Mendel's segregations and their χ values (excerpted from Edwards 1986a, table 2)

Character	Expected	Observed		Total	χ
		F_3, Plant characters			
Color of seed coats	0.63 : 0.37	64	36	100	+0.2252
Form of pods	0.63 : 0.37	71	29	100	+1.6743
Color of unripe pods	0.63 : 0.37	60	40	100	−0.6029
Position of flowers	0.63 : 0.37	67	33	100	+0.8462
Length of stem	0.63 : 0.37	72	28	100	+1/8813
Color of unripe pods	0.63 : 0.37	65	35	100	+0.4322
		F_3, Trifactorial experiment, plant character			
Color of seed coats, among *AaBb*	0.63 : 0.37	78	49	127	−0.3488
Color of seed coats, among *AaBB*	0.63 : 0.37	38	14	52	+1.5174
Color of seed coats, among *AABb*	0.63 : 0.37	45	15	60	+1.9384
Color of seed coats, among *AABB*	0.63 : 0.37	22	8	30	+1.1816
Color of seed coats, among *Aabb*	0.63 : 0.37	40	20	60	+0.6020
Color of seed coats, among *AAbb*	0.63 : 0.37	17	9	26	+0.2610
Color of seed coats, among *aaBb*	0.63 : 0.37	36	19	55	+0.3903
Color of seed coats, among *aaBB*	0.63 : 0.37	25	8	33	+1.5276
Color of seed coats, among *aabb*	0.63 : 0.37	20	10	30	+0.4257
Total	0.63 : 0.37	720	353	1073	+2.8408

Weiling considered other experiments on peas, including several of those discussed by Sturtevant (table 1.17), and concluded that there was no great difference between Mendel's results and those of the later experiments. Weiling further stated that if one thought that Mendel had corrected his data, then the same comment should be made about Correns, von Tschermak, and geneticist Arthur Darbishire. This is in contrast to Sturtevant's conclusion that the data were "neither too good nor too poor."

Weiling also suggested that in the 2:1 ratio experiments, fewer than 10 plants survived. He stated, in contrast to Fisher, Dunn, Sturtevant, and Wright, who suggested that a number larger than 10 was required to improve the fit between Mendel's observed results and his expectations, that this smaller number would achieve that result. Despite Weiling's arguments, the fact is that this smaller number actually worsens the fit.

Weiling also questioned whether χ^2 analysis was appropriate for analyzing Mendel's data because it assumes a binomial distribution. He emphasized that Mendel's pea plants did not, in fact, obey the expected binomial distribution, citing the tetrad model of pollen. He concluded that this would result in a variance that was reduced by a factor *c*, which was a function of the number of surviving plants, resulting in a higher χ^2. Weiling found that if he assumed that 8 out of 10 plants survived, that *c* was between 0.7 and 0.8, whereas for 9 out of 10 plants surviving, *c* was between 0.6 and 0.7. This resulted in a good fit to randomness and a much less improbable value of χ^2. Weiling also raised the issue of whether Mendel's results on the 2:1 ratio experiments should be compared to the 2:1 ratio Mendel expected or to the correct value of 1.8874:1.1126. He calculated the χ^2 using the corrected value and found that the total c^2 increased from 41.6506 to 48.910, which was also highly significant.[40] The probability of a worse fit decreased from 0.99993 to 0.9992. Unfortunately, Weiling seems to have estimated *c* by asking what value of that parameter would give a reasonable value of χ^2. As Edwards (1986a) subsequently pointed out, this really begs the question.

Theodosius Dobzhansky (1967), in a review of six then-recent books on Mendel and the history of genetics, offered a somewhat different explanation for Mendel's results.

> Few experimenters are lucky enough to have no mistakes or accidents happen in any of their experiments, and it is only common sense to have such failures discarded. The evident danger is ascribing to mistakes and expunging from the record perfectly authentic results which do not fit one's expectations. Not having been familiar with chi-squares and other statistical tests, Mendel may have, in perfect conscience, thrown out some crosses which he suspected to involve contaminations with foreign pollen or other accidents. (Dobzhansky 1967, 1588)

Dobzhansky also cited Mendel's publication of his *Hieracium* (hawkweed) data (Mendel 1870), discussed earlier, which disagreed with both his expectations and his previous results, as evidence that Mendel accepted discordant data and had been unlikely to engage in falsification.

A novel suggestion concerning hypothetical experiments was put forth by Bartel Leendert van der Waerden (1968). Van der Waerden examined the effect on the χ^2 if a hypothetical experimenter continued performing an experiment until there was sufficient agreement with his expectations. For calculational purposes he assumed that such an experimenter continued with a sequence of experiments until the χ^2 for that particular experiment was ≤ 1.69.[41] Van der Waerden then showed that such a procedure might account for overall goodness of fit in Mendel's results. There is, however, no evidence that Mendel did this. Van der Waerden went on to discuss the gametic cell experiments and stated, "Yet, I cannot but agree with Fisher's conclusion that the data have probably been biased. By the time the last series of five experiments was performed, Mendel and his gardening assistants knew too well what they had to expect" (van der Waerden 1968, 287). Van der Waerden also discussed the trifactorial experiment, remarking that Mendel might have planted 12 or 15 seeds or used back-crossing as an additional check: "One or two back-crosses of each of his 473 doubtful plants with cc-plants would reduce the systematic error of his counts by a factor of ½ or ¼" (285). He admitted that although the data for the later years were probably biased, the evidence was not quite as strong as Fisher believed, and he thought that the results of the original 3:1 experiments were probably not biased.[42]

A review of the status of the controversy was written by Vítězslav Orel (1968). It included virtually all of the papers that we have discussed in this section. He added both Jaroslav Kříženecký and Robert Olby (1966) to the group of people who believed that Mendel had stopped counting when his results looked good. He also cited the work of Herbert Lamprecht (1968), who thought that either selective pollination or mutant genes that affected pollination might account for the reduced variance in Mendel's results. Orel, however, thought that Lamprecht's omission of the arguments given by others greatly weakened his case. Orel concluded that given all the evidence, "there are no proofs either of the possibility that Mendel was deceived by somebody, or that he had unconsciously or deliberately falsified the data from his experiments" (Orel 1968, 778). He offered the hope that "the story of 'too good' results in Mendel's experiments may be closed by quoting Dobzhansky: 'Far from this, he was a most careful experimenter and a most penetrating analytical mind'" (778).

Contrary to Orel's hope, the controversy continued. His summary was too optimistic. Although a possible biological explanation of Mendel's "too good" results had been offered in the tetrad-pollen model, there was no evidence that it applied to pea plants, or if it did, that it was sufficient to account for the results. None of the plausible explanations of Mendel's results—deception by an assistant, conscious or unconscious bias, or stopping the count when the results agreed with his expectations—had very much supporting evidence, and none of them were universally accepted.

The Middle Period

There were few papers published on the Mendel-Fisher controversy during the 1970s. One interesting mention occurs in A. W. F. Edwards's book, *Likelihood* (1972). In his discussion of the χ^2 test, Edwards presciently remarked, "It would be interesting to rework Fisher's analysis" (190). Edwards also referred the reader to discussion of χ^2 contained in Fisher's *The Design of Experiments* (1935), written only a year before Fisher's analysis of Mendel's data.

Margaret Campbell (1976) summarized the situation regarding the possible explanations of Mendel's results. These included suggestions, previously mentioned in this discussion, that Mendel may have: (1) been deceived by an assistant, (2) faked his results, (3) stopped counting when the results agreed with his expectations, (4) subconsciously favored his expected results, (5) been able to distinguish genotypes from observation of phenotypes, or (6) been lucky. Campbell rejected the first suggestion because she thought that, given the care with which Mendel performed his experiments, it was unlikely that he would have delegated such an important task to an assistant. The second was deemed unlikely because had Mendel wanted to fake his results, he could have withheld results or used selection procedures that would have given better results. She also noted that Mendel had presented data that disagreed with his expectations as an argument against his possible falsification of data. Campbell offered an interesting and novel argument against the view that Mendel had planted seeds from more than 100 plants in the 2:1 ratio experiments. Had he done so, she said, "he could have selected from which 100 lots of 10 he made his observations and there would seem no reason to repeat an experiment [Experiment 5] had this been his procedure" (162). She noted that Experiment 7 on stem length, which gave a large observed deviation from the expected result, was not repeated. If Mendel had selected for hybrid vigor, which this trait showed, then Campbell thought that the results should have been closer to expectation. Campbell also regarded Mendel's comment "that all members of the series developed in each successive generation should be, *without exception*, subjected to observation" (80), as an argument against the view that Mendel stopped counting when the results agreed with his expectations. She believed that Mendel may very well have been able to determine genotypes from phenotypes, in at least some experiments, and conceded that luck might have been a factor in Mendel's results.

Campbell also asked whether an appropriate statistical model had been used. She noted both Beadle's suggestion of the tetrad-pollen model and his failure to account for Mendel's results using this model. She concluded that subconscious bias combined with an ability to detect heterozygosity in the absence of recessives, along with Beadle's tetrad-pollen model, might be sufficient to explain Mendel's results in the 2:1 ratio experiments. No calculations were presented.

Two statisticians proposed alternatives to Fisher's χ^2 analysis. Tom Leonard (1977) suggested a Bayesian approach to multinomial estimates. He applied his technique to Mendel's data and stated, "We have no wish to enter into the general controversy but simply wish to make the point that, in situations like this, with χ^2 smaller than the degrees of freedom, the raw proportions could still give better estimates" (873). He argued that his method gave a lower probability of a worse fit to Mendel's results than did Fisher's calculation. Tim Robertson (1978) proposed an order restriction on multinomial parameters.[43] When he applied these restrictions to an analysis of Mendel's data he, too, found a reduced probability.

In a later review of the controversy, Walter Piegorsch (1983) commented, "Both of these approaches provide somewhat more reasonable interpretations of the fit for this particular subdivision of the data. Their overall impact, however, helps advance the debate only slightly. Still, the various statistical and experimental alternatives proposed do tend to give one the impression that Fisher's original conclusions may have been a bit extreme" (2300). Piegorsch's review also included a detailed summary of Weiling's views, including his argument for a reduced variance because of the tetrad-pollen model.[44] Piegorsch commented that Weiling's view "has not received a great deal of further attention" (2298).[45]

Another review of the controversy was presented by Robert Root-Bernstein (1983). One criticism he raised was that in analyzing the 2:1 ratio experiments, Fisher should have calculated the χ^2 using the corrected values 1.8874:1.1126 rather than 2:1. In that case, Mendel's fits are about what one would expect on the basis of probability. He commented that Fisher used the same "ploy" in analyzing the trifactorial experiment.

Root-Bernstein also added a new explanation of Mendel's "too good" results. He argued that Mendel had attempted to segregate continuously varying quantities into discrete categories. He presented evidence from Pearl's experiments, discussed earlier, and from experiments of his own, on the difficulty of obtaining repeatable results. In one of his experiments, he asked undergraduate students to classify maize kernels into two groups, yellow and purple. He separated the students into two groups. One group was asked to use only the categories "yellow" and "purple," whereas the second group was allowed to add a third category, "indeterminant." The first group obtained a poor fit to the expected results. For the second group, Root-Bernstein first excluded all the "indeterminant" kernels and obtained a good fit to the expectations, "although these were not better than one would expect on statistical grounds. . . . Finally, students were asked to reassign the 'indeterminant' kernels to the 'ideal' categories so as to achieve the closest possible approximations to the expected Mendelian ratio. This group of students produced results that were as statistically unlikely as those reported by Mendel" (284).[46]

Root-Bernstein argued that this was, in fact, Mendel's method. He also concluded, on the basis of these experiments, that there were a sufficient number of "difficult to classify" (284) individuals to account for Mendel's results. He stated that in Mendel's work on seed-coat color, Mendel originally began with six categories and, when he could not obtain good fits to his expectations, he reduced the number to the three expected. "One could hardly ask for better evidence that Mendel recognized the arbitrary nature of his imposition of discrete categories on nature and was willing to revise them to fit theory" (288).[47]

Root-Bernstein recognized that he had, at least at first glance, offered an argument in favor of fraud on Mendel's part. He concluded, however, "If Mendel was conscious of manipulating his data to achieve the verification of his hypothesis, does this not make him guilty of the charge of fraudulence? Did he not 'doctor' his data as has been alleged? I believe not. One can allege fraud only if one can demonstrate that an objective truth presented itself to Mendel and he ignored it in favor of some preconceptions. In fact, the point of this essay has been to argue that Mendel's peas did not represent an 'objective truth' that Mendel could unambiguously interpret" (289).

I believe that there are several problems with Root-Bernstein's explanation, not the least of which is that it is inconsistent with what Mendel himself stated. Mendel said quite emphatically that he had used traits that could be easily distinguished and that there were others, for which the categorization was not as easy, which were not used. Mendel further commented, "In counting the seeds, also, especially in Expt. 2, some care is requisite, since in some of the seeds of many plants the green colour of the albumen is less developed, and at first may be easily overlooked. . . . In luxuriant plants this appearance was frequently noted. Seeds which are damaged by insects during their development often vary in colour and form, but, with a little practice in sorting, errors are easily avoided" (*86–87*). In order to absolve Mendel from fraud, Root-Bernstein would make him a liar. I also question Root-Bernstein's view that an objective truth that was ignored is a requirement for fraud. Mendel clearly did not realize at the time he performed his experiment that he could not accurately count peas, whether this is true or not. In all probability, Mendel also had a good idea of the results he expected. Thus, contrary to Root-Bernstein, if Mendel had manipulated his data to agree with an incorrect, or even impossible, expectation, this would be evidence of fraud.[48]

The defense of Mendel was continued by Ira Pilgrim (1984). Pilgrim stated that his purpose was "to demonstrate that Fisher's reasoning was faulty and to clear the name of an honest man" (501). He faulted Fisher's analysis of Mendel's work this way: "Here is the paradox: the closer his results are to his expectations the less credible they become and the farther they are from his expectations the more credible they become. In other words, if his results are excellent, he is accused of dishonesty, and if his results are poor, they do not support his theory" (501). Pilgrim also maintained that Fisher obtained such a low probability for Mendel's results by successively multiplying the probability of each of Mendel's results. Thus, because the probability of any result is less than one, the probability will decrease with an increasing number of experiments. He further argued that Fisher assumed that getting a royal flush in a poker hand was evidence of cheating because of its low probability. Pilgrim concluded that fictitious data can never be discerned by statistical means, but "only the time honored method of the critical repetition of the work can do that [corroborate results], as it did for Mendel's work" (502). In sum, Pilgrim stated, "There is no evidence that Mendel did anything but report his data with impeccable fidelity. It is to the discredit of science that it did not recognize him during his lifetime. It is a disgrace to slander him now" (502).

Pilgrim was answered by A. W. F. Edwards (1986b). Edwards agreed with Pilgrim that an extremely good fit to one's expectations in a single experiment is not evidence of fraud, but that in Mendel's case it applied to 84 different experiments. "One can applaud the lucky gambler; but when he is lucky again tomorrow, and the next day, and the following day, one is entitled to become a little suspicious" (138). Edwards also corrected Pilgrim's error by asserting that Fisher did not, in fact, perform multiplication of successive probabilities. Fisher added independent χ^2 values, a justified mathematical procedure. Edwards concluded, "[Fisher's] grounds for concluding that Mendel's data were falsified were not that it was exceedingly improbable that they would recur exactly on a repeat of the experiment (which it is) but that it was very improbable than *any* results so close to expectation would recur on a repeat" (138).

Pilgrim (1986) responded by admitting his error on the multiplication of probabilities. He claimed, however, "If data are honest and 'good' adding the chi-squares will increase the P value, making good data seem excessively good" (138).[49] Pilgrim agreed with Edwards about the "lucky gambler" but believed that the evidence was insufficient to support an accusation of cheating. "However, one had better have a good deal more evidence (such as a set of loaded dice or perhaps the information that the man is a known cheat) before accusing someone of cheating, which is what Fisher did to Mendel" (138).

Floyd Monaghan and Alain Corcos (1985) joined the debate on Mendel's behalf. They maintained that there was, in fact, no evidence of bias in Mendel's results. If there were bias, they argued, and Mendel knew what results to expect, then one should expect the later χ^2 values to be smaller than the earlier ones. Monaghan and Corcos found no evidence that this was so. When they examined the gametic ratio experiments they admitted, "Obviously, these chi-square values are small. Since these experiments were done to test his hypothesis of gametic formation, one could think that Mendel in this case could have consciously or unconsciously biased his results" (308). They also calculated and summarized the χ^2 values for all of Mendel's experiments and found a probability of between 0.95 and 0.99 of obtaining a worse fit, which they did not seem to regard as remarkable. Monaghan and Corcos examined the seven early-twentieth-century experiments and noted that the results of Bateson and von Tschermak deviated widely from those of other experimenters. They concluded that although Fisher's statistical procedure "is undoubtedly correct, the conclusion seems to us to be illogical. We have a series of independent experiments, none of which shows evidence of bias and whose chi-square values show no systematic trend. Yet the sum of these individually unbiased experiments is judged as showing bias" (309). Recall, however, Edwards's comment about a gambler being lucky on successive days. Monaghan and Corcos also suggested that the solution to the problem might lie in the biology of egg and pollen formation as suggested by Beadle and Thoday.

Olby (1985), in a revised version of his 1966 book, noted the arguments of Weiling concerning semi-random pollination and van der Waerden's suggestion of sequential experiments. He presented further evidence that Mendel had stopped his counts when the results were close to his expectations by arguing that the number of seeds counted was smaller than one would expect given the number of plants Mendel had examined.

Weiling (1986) continued his defense of Mendel. He added to his argument for reduced variance by citing the work of Evans and Philip (1964), Thoday (1966), and his own previous work. He also examined seven other early-twentieth-century experiments on seed color and, although four of the eight experiments (which included Mendel's experiment) showed χ^2 values greater than that for a probability greater than 0.5 and four less, Weiling believed they provided evidence in support of reduced variance. (The overall χ^2 for the eight experiments was 6.368, whereas the χ^2 for probability equal to 0.5 was 7.34.) Weiling further argued that for biological reasons, the variances for the different experiments performed by Mendel were, in fact, different, unlike the homogeneous variance assumed by Fisher. He also argued that in the 2:1 ratio experiments, one should use a hypergeometric, rather than a binomial, distribution because Mendel's 10 plants were chosen from a finite, not an infinite, number of seeds. Weiling also commented on the fact that Fisher's statement of "too good to be true" data had led popular authors to consider Mendel a "betrayer of the truth" (see Broad and Wade 1982).[50] He expressed the hope that because, as he believed, he had shown that Fisher's two decisive statements were incorrect, that the "defamatory questioning of Mendel's accuracy henceforth will stand corrected" (283).

The next article in the same journal was a correction offered by Corcos and Monaghan (1986) of their 1985 paper. Weiling had pointed out that they had incorrectly stated that the number of degrees of freedom for Mendel's testcross experiments was 9 when it was actually 15. This reduced the χ^2 per degree of freedom considerably and gave a probability of a worse fit of greater than 0.995, which they admitted was evidence of bias. They suggested that because Mendel knew what to expect, he might not have been as careful as he should have been in scoring phenotypes. Weiling had also pointed out that they had made an error in reporting the number of plants observed by von Tschermak. Corcos and Monaghan had found an error in their report of Bateson's results. Both of these errors reduced the χ^2 of these experiments, and they claimed that this supported their view that Mendel and these other experimenters were equally biased. Thus, they stated, "The above errors . . . with the possible exception of the testcross data, in no way alter our conclusion" (283).

The middle period closed with A. W. F. Edwards's paper, "Are Mendel's Results Really Too Close?" (1986a; chapter 4 of this volume). In this paper, Edwards provided a survey of previous explanations of Mendel's results, a critical review of the issues raised about Fisher's statistical analysis, a discussion of the relevance and appropriateness of using χ^2 analysis, and a new analysis of Mendel's data. Edwards noted, "Fisher's 1936 conclusion slowly became the received wisdom, but his painstaking analysis and his defence of Mendel's integrity have sometimes been incorrectly reported as having exposed a scientific fraud of major proportions" (*142*).

Edwards began by surveying the explanations offered for Mendel's results. He remarked that while a large number of commentators had accepted the suggestion that Mendel had stopped counting peas when his results agreed with his expectations, "The difficulty is that there is no evidence whatsoever that Mendel did this, and in many of the experiments his description clearly excludes the possibility, so that to accept it as a means of exonerating Mendel from having reported data which had been adjusted in some way is to saddle him with the charge of not having reported his experimental method accurately" (*144*). Edwards also considered the suggestion that peas exhibit a non-binomial variability, the tetrad-pollen model of Beadle and Thoday, and cited Beadle's comment that it was not sufficient to explain Mendel's data. Edwards also presented evidence that peas did exhibit normal variability, a point that will be discussed further later in this introduction. In a detailed consideration of the arguments offered by Weiling, Edwards remarked that Weiling had suggested that one might explain the results of the 2:1 ratio experiment by assuming that only 8 of the 10 seeds sown produced plants. Edwards noted that this would make Mendel's fit to the correct results even worse and cited Sturtevant, Wright, Dunn, and Fisher, all of whom suggested that one could explain the results only if Mendel had examined *more* than 10 plants. "Nothing could illustrate better than these two opposing theories the ingenuity that has been expended on accounting for Mendel's surprisingly good data. Here are two explanations, one postulating that *fewer* than ten plants were scored, and the other that *more* than ten were scored, both with a view to accounting for results judged too close to a 2:1 ratio" (*145*). Edwards further remarked that there was, in fact, no evidence that Mendel had done either.

Edwards devoted considerable attention to previous statistical analyses. He raised the issue of whether Fisher should have computed his χ^2 by comparing Mendel's data to Mendel's expected 2:1 ratio or to the corrected value of 1.8874:1.1126. Edwards noted that this issue was relatively unimportant because Weiling (1966) had adjusted the total χ^2 using the corrected value and found that the value increased from 41.6506 to 48.910 and that this value was also "highly significant" (*147*). (The probability of a worse fit to the data decreased from 0.99993 to 0.9992.) Edwards also considered Weiling's suggestion of a binomial variability reduced by a factor of *c*. "Of course, once one has estimated *c* (for which Weiling found the broad limits 0.6–1.0) there is nothing left to test" (*148*). Edwards presented the evidence from the experiments of Bateson and Kilby (1905) and of Darbishire (1908; 1909) for which Weiling had calculated a total χ^2 of 1008.79 for 1062 degrees of freedom. This gave an estimate of *c* of 1008.79/1062 = 0.95. "In other words these massive data exhibit a standard deviation of about 97.5% of that expected on a binomial model, and far from this lending support to the biological hypothesis that the pea does not segregate randomly, it fills one with admiration for the perfection of the randomizing mechanism!" (*148*). Of Weiling's calculation that had found that 463 one degree of freedom experiments had χ^2 values less than the theoretical median of 0.4529 and 427 above that median, Edwards said the data "lend no real support to the infra-binomial-variance hypothesis" (*148*).

While Edwards offered, "Weiling is to be congratulated for his determined attempts to rescue the Mendelian experiments from the Fisherian conclusion" (*149*), he concluded, "But in the end the attempt fails: there is too much to explain. The diminished-variance hypothesis is not supported by other more extensive data, and, as I show in my own analysis below, simply reducing the variance is in any case not enough to explain the peculiarities of Mendel's data" (*149*).

Edwards found the defense of Mendel offered by Monaghan and Corcos less than admirable. He noted that many of the χ^2 values given were incorrect to the accuracy claimed, that their addition of the χ^2 values was wrong, and that their report of the results of Bateson and of von Tschermak was incorrect. Referring to the data contained in table III of Monaghan and Corcos (1985), Edwards wrote that it "contains the astonishing assertion that $0.95 < P < 0.99$ for $\chi^2_{67} = 32.57$, on which obvious error the authors hang their conclusions" (*149*). (The probability is, in fact, 0.9998.)

Edwards concluded, "My overall impression from reading all of the commentaries since Fisher (1936) is that a good deal of special pleading, not to mention downright advocacy, has failed to make any substantial impact on Fisher's conclusion" (*150*).

The issue of the appropriateness of using χ^2 analysis was also discussed by Edwards. He noted that the total χ^2 was greatly influenced by large individual χ^2 values, for which the probability was low. "The criticism of Mendel's results is that they are too close to expectation in the normal space as judged by being too far from expectation in the χ^2 space" (*151*). He then offered the following amusing illustration. "Suppose the χ^2 test had antedated Mendel, and that in his paper he had reported a value of 84.0000 on 84 d.f. [degrees of freedom] The reaction of a latter-day Fisher might well have been to conclude that Mendel's assistant had known that what Mendel really needed for his paper was not good Mendelian ratios but a good value of χ^2" (*151*). He ended his discussion of χ^2 with this cautionary statement: "To sum up, we must not allow our judgment to be dominated by tests of significance and other calculations of probability which are at best pointers for further thought and at worst misleading" (*152*).

The most significant aspect of Edwards's paper was his new analysis of Mendel's data. Edwards chose to work with χ rather than with χ^2, because it preserved the direction of the departure from expectation. He also chose to compare Mendel's results on the 2:1 ratio experiments to the corrected value of 0.6291:3709 rather than the 2:1 ratio Mendel expected. The results for those 15 experiments are shown in table 1.16. (All of the data are shown in table 2 in chapter 4 of this volume.) He found that of the 15 χ values, 13 were positive and 2 negative, "suggesting something of a bias towards the larger class" (*154*). The total numbers of plants that Mendel observed in these experiments was 720:353 (heterozygote:homozygote), or 0.6710:0.3290, "with an associated χ value of +2.8408, indicating a very poor fit [to 0.6291:0.3709] indeed" (*154*). (The probability that such a deviation is due to chance is 0.0045.) Edwards believed this result substantiated "what Sturtevant (1965) called 'Fisher's most telling point' in his argument that the data have been biased in the direction of agreement with what Mendel expected" (*154–55*).

Edwards then investigated the remaining 69 results obtained by Mendel. He found that the mean values were unexceptionable, but that the sum of χ^2 values for these 69 results was 30.8138, which, he said, is "remarkable on any interpretation of tests of significance" (*155*). (The probability of a worse fit is 0.9999.) He then went "further than any previous analysis of Mendel's data" (*159*) and examined the distribution of the 69 χ values, omitting the 15 results from the 2:1 ratio experiments. Edwards's results are shown in figure 1.1. He contrasted these with the first 69 random normal deviates from the tables provided by Lindley and Scott (1984) (figure 1.2). If the data were randomly distributed, these graphs should fit the straight line shown in the figures. This is true for the random deviates (see figure 1.2), but not for Mendel's data (see figure 1.1). "It is immediately obvious that the reduced variance is not characteristic of the whole data, as Weiling's theory would require, but is confined to the tails of the distribution, where the extreme variates are not extreme enough to conform to expectation" (*159*). The graph of Edwards's results shows no significant deviations in the region between the 25th and 75th percentiles, but does show an excess of observations between the 5th and 25th and the 75th and 95th percentiles, and a lack of observations between 0 and 5% and between 95 and 100%. "The inescapable conclusion," Edwards wrote, "is that some segregations beyond the outer 5-percentiles (approximately) have been systematically biased towards their expectations so as to fall between the 5-percentiles and the 25-percentiles. Further analysis shows that the effect is not confined to particular sample sizes or segregation ratios, but is quite general" (*159*). Edwards did not speculate on how the data had been adjusted.[51]

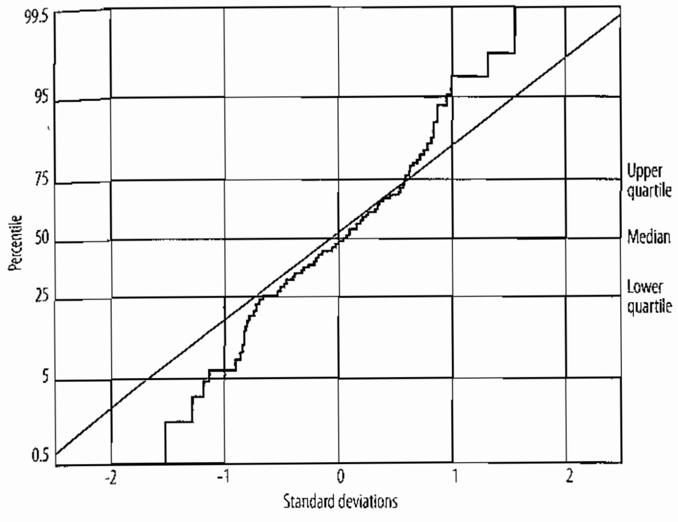


FIG. 1.1 The values of χ for the 69 segregations with undisputed expectations plotted on normal probability paper

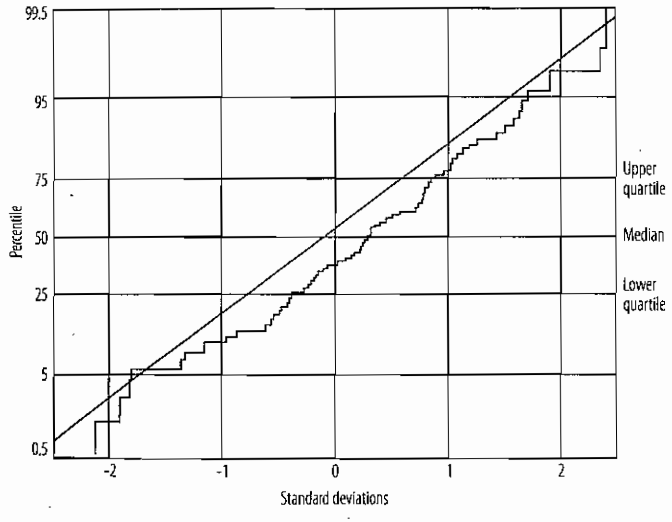


FIG. 1.2 The values of 69 simulated standard normal deviates plotted for comparison with fig. 1.1

Edwards prefaced his concluding remarks by noting that he had originally thought Fisher's analysis might be faulted because of Edwards's own doubts concerning χ^2 analysis, "but a complete review of the whole problem has now persuaded me that his 'abominable discovery' must stand" (*161*). He thought, however, that "any criticism of Mendel himself is quite unwarranted" (*161*). He also suggested, in agreement with Dobzhansky, that Mendel should not be judged by modern standards of data recording; he quoted Dobzhansky's view: "not having been familiar with chi-squares and other statistical tests, Mendel may have, in perfect conscience, thrown out some crosses which he suspected to involve contamination with foreign pollen or other accident" (Dobzhansky 1967, 1589). Edwards's final comment was, "In the words of my title, Mendel's results really are too close" (*162*).

Thus, by the middle of the 1980s, the Mendel-Fisher controversy was no nearer to a solution than it had been in 1970. Fisher's analysis was still seen as correct by most scholars, there was no generally accepted explanation of why Mendel's results were "too good," and the additional mystery of the distribution of χ values pointed out by Edwards had no apparent explanation.

The Contemporary Period

Weiling (1989) continued his defense of Mendel and argued that the model used by Fisher in calculating the χ^2 values was incorrect. He again argued that, for botanical reasons, the hypergeometric distribution, which gives a smaller variance, and thus a larger χ^2 value and a "less good" fit to Mendel's data, should be preferred to the binomial distribution used by Fisher. Weiling further suggested that the samples contained in Mendel's data were too small to allow the appropriate use of χ^2 analysis. He also claimed that his analysis showed that all of the data were moved to smaller χ^2 values, not just the larger deviations moved inwards as shown by Edwards.[52] With regard to the 2:1 ratio experiments, Weiling criticized Fisher for using Mendel's expected 2:1 ratio rather than the corrected value of 1.8874:1.1126 in calculating the χ^2 values for these experiments. He did not, however, mention his own calculation, discussed earlier, which found that the total χ^2 increased from 41.6506 to 48.910 and that this value was also "highly significant." (The probability of a worse fit to the data decreased from 0.99993 to 0.9992.) Weiling further argued that in considering the 10 seeds planted in these experiments, one should calculate the probability using a "balls taken from an urn without replacement" model. For a plant with an average of 30 seeds, this results in a probability of $23/30 \times 22/29 \times 21/28 \times 20/27 \times 19/26 \times 18/25 \times 17/24 \times 16/23 \times 15/22 \times 14/21 = 0.0381$. In that case the probability of misclassification would be 0.0381 rather than the 0.0563 calculated by Fisher.[53]

Fisher's contributions to genetics were reviewed by Piegorsch (1990), who included a discussion of the Mendel-Fisher controversy. He noted that Mendel's data exhibited unusually close agreement with his [Mendel's] expectations and: "The controversy over the data's nature and origin remains unresolved" (921). Piegorsch emphasized that Fisher, despite his concern over the possible fraudulent data, was a great admirer of Mendel: "In spite of the concern over the 'cooked' data, Fisher argued that Mendel's experimental methodology was an advance well ahead of its time" (922). Remarking on Fisher's comment that each generation found in Mendel's work only what it expected to find, he noted:

Even Fisher's work has fallen prey to some modern writers' inability to read past their own (pre)conceptions. Fisher is often identified among popular writers as Mendel's intellectual assassin and scourge for his discoveries regarding the 'too good' fit of Mendel's data. As noted above, however, Fisher *never* argued that the (possible) falsification of the data was Mendel's doing, and valiantly suggested another alternative: the 'overzealous' assistant. Although this particular solution to the data falsification problem remains in doubt, the fact remains that one of Fisher's basic conclusions espoused support and admiration for Mendel's scientific work. It is inappropriate and misleading to paint Fisher in any other light. (Piegorsch 1990, 922)

Jan Sapp (1990) also supported the view that Fisher was not attempting to discredit Mendel, but rather "meant to celebrate his [Mendel's] power of abstract reasoning" (157). Sapp also discussed several methodological issues raised by the controversy including observer bias, the theory-ladenness of observation, and whether the validity of experimental results could be tested by statistical methods. He discussed a point similar to one made previously by Pilgrim:

data that looks good to the experimentalist looks bad for the statistician and *vice versa*. There seems to be a methodological incommensurability concerning the nature of statistical and experimental modes of reasoning. This might be called the 'experimentalist-statistician paradox': From a statistical point of view geneticists should not provide 'too much data' and have their results come too close to theoretical expectations, for the closer they come to the 'truth' the less true they will appear to be. This strange paradox is based on faulty reasoning. The reason why data are considered less true the closer they reach theoretical expectations is based on the idea that geneticists *should be* studying a random sample. It assumes that experiments should be carried out independently of the law or theory the observer is using for explanation. . . . The theory informs the experimenter what kind of experiment to perform, what kind of phenomena to examine how results are to be understood; it also tells the experimenter when the experiment is over. (Sapp 1990, 159)

Sapp regarded this last point as underlying the suggestion that Mendel stopped counting when the results agreed with his expectations.

In 1991, Federico Di Trocchio (1991) proposed a novel way of explaining Mendel's "too good" results. Mendel did not report any linkage between the characters he studied when, in Di Trocchio's view, he should have observed such linkage. Thus, Di Trocchio concluded that Mendel actually performed 22 × 22 = 484 crosses for the 22 varieties of peas he had selected for further experimentation.[54] Di Trocchio agreed with Bateson that Mendel's experiments "are fictitious in the sense that they were made just on paper by disaggregating the data of many polyhybrid experiments" (511). In Di Trocchio's view, Mendel then selected that data that best agreed with his 3:1 expectation, and this explains the excessively good fit: "This way of proceeding cannot be considered a form of manipulation or falsification of the data: Mendel simply chose to work with and report the data that showed the closest approximation to the ratio 3:1. Since in my opinion Mendel had in his possession a much wider sample than that reported in *Versuche*, he had available to him various tables containing columns of data on the characteristics yellow/green, wrinkled/smooth, tall/dwarf, and so on. There was nothing to prevent him from choosing the table most suitable to his purposes" (Di Trocchio 1991, 514).

As Daniel Hartl and Vítězslav Orel (1992) pointed out, it seems hard to believe that Mendel would have performed such an elaborate procedure "without saying so, since he roundly criticized Gärtner for not describing his experiments in sufficient detail to allow Mendel to repeat them" (246). Hartl also remarked in a personal communication to this author, "Mendel had carried out a trihybrid cross, and complained that of all his experiments it took the most time and effort. Because he knew how much effort went into multifactorial crosses, why would he make these his method of choice, and then conceal it by disaggregation."

Moreover, Fairbanks and Rytting (2001) pointed out on the basis of botanical evidence that "the nature of variation in pea varieties (both old and modern) facilitates, rather than prevents, the construction of monohybrid experiments" and that Di Trocchio's "claim that Mendel hybridized his 22 varieties in all possible combinations runs counter to the experimental design that Mendel described and the logic on which it is based" (283–84). Their analysis fully supports Fisher's conclusion that Mendel's "report is to be taken entirely literally, and that his experiments were carried out in just the way and in much the order that they are recounted" (Fisher 1936, 134).

Ironically, after having accused Mendel of, at the very least, deception in reporting both his methods and his results and of selectivity and bias in his choice of data, Di Trocchio concluded, "His [Mendel's] research is in no way the fruit of methodological mistakes or forgery, and it remains a landmark in the history of science. . . . We must still consider him the father and founder of genetics" (519).

When Hartl and Orel (1992) discussed and criticized Di Trocchio's view, they did not discuss other aspects of the "goodness of fit" question. They did, however, discuss in detail the questions raised by Fisher in his article on Mendel: What was Mendel trying to discover? What did he discover? What did he think he had discovered? Hartl and Orel also addressed some revisionist views of Mendel's work, including the view that Mendel was not interested in heredity, that he was interested only in hybrids, that he did not perform all the experiments attributed to him, as well as other questions such as the reasons for the neglect of Mendel's work.

In their 1993 text *Gregor Mendel's Experiments on Plant Hybrids: A Guided Study,* Corcos and Monaghan restated their conclusion, citing both their own previous work and that of Weiling, that "Mendel's reputation was indeed restored" (196). They presented a discussion of χ^2 testing and noted that if an experiment had a χ^2 of 0.00015 for one degree of freedom, the probability that that result occurred by chance is less than 1%. They further stated that if a second experiment had the same χ^2, the probability that both experiments would have such a low χ^2 is $0.01 \times 0.01 = 0.0001$. They offered this as part of an argument that there was, in fact, no bias in Mendel's results. Their calculation of probability is correct, but it is not, in fact, what Fisher claimed. His claim was that if all 84 experiments performed by Mendel were repeated, the probability of a worse fit to Mendel's expectations was 0.99993, or the probability of a better fit of 0.00007. This was not calculated by multiplying the probabilities of each of Mendel's results, but by summing their χ^2 values and then computing the probability.[55]

Two new possible explanations of Mendel's results were offered by Moti Nissani (1994). In support of his view that Mendel had not falsified his data, he noted that Mendel had published data that disagreed with his expectations, particularly his results on *Hieracium*, and that Mendel had sent 140 packets of *Pisum* seeds to Nägeli, in the hope that Nägeli would repeat his experiments. After reviewing the history, Nissani stated that he thought that no one had as yet cast any serious doubt on Fisher's analysis, and that the previous explanations of Mendel's results were both out of character and not wholly satisfactory. He then proposed two possible explanations that "involve the contention that Mendel *consciously* presented biased data, but that in doing so he acted honorably and in the best interests of science" (188). Nissani asked the following questions: "What is the proper conduct when a conflict arises between the norm of communicating one's findings and the norm of communicating them as faithfully as possible? What is one to do when the only way to communicate a larger truth is to tamper with some inconsequential details?" (188). Nissani suggested that Mendel, in order to make his contributions, which he believed were of great significance, more acceptable and believable to his intended audience, may have omitted or adjusted his data. Nissani's second explanation posited that Mendel may have omitted data because of length limitations imposed by the editor of the journal.[56]

Nissani quoted Broad and Wade (1982), who argued that "some who commit fraud do so to persuade their refractory colleagues of a theory they know is right. . . . If history has been kind to scientists such as these, it is because the theories turned out to be correct. But for the moralist, no distinction can be made between an Isaac Newton who lied for the truth and was right, and a Cyril Burt who lied for truth and was wrong" (212–13). Nissani concluded, "Sometimes, however, the demand to faithfully report one's data must be sacrificed for the higher value of advancing knowledge. It is, of course, impossible to know whether Mendel (and perhaps also others) faced such a procrustean dilemma. But the information presented in this paper raises the possibility that he may have. In that case, Mendel's choice merits our compassion and thanks, not our disapprobation" (195).

In the year after Nissani's paper was published, Orel and Hartl (1994; chapter 5 of this volume) presented a longer and more detailed discussion of the issues they had raised in their 1992 paper. In particular, they discussed the question of Mendel's 2 : 1 ratio experiments, which had triggered Fisher's suspicions. They wrote:

> Among 600 plants tested, therefore, the true expected ratio is 377 : 223. However, Mendel reports 399 : 201, a ratio in much better agreement with 0.67 : 0.33 (that is, 2 : 1) than with 0.63 : 0.37. However, a χ^2 test of the reported ratio against the expected 377 : 223 yields $\chi^2 = 3.3$, for which $P > 0.05$. The observed result is, therefore, not significantly more deviant from the true expectations than could be expected by chance alone. In other words, this series of progeny tests yields no evidence that the data had been adjusted. (Orel and Hartl 1994, *196*)[57]

The second set of results on the 2 : 1 ratio, from the trifactorial experiments, was, they noted, more problematic. Assuming 10 seeds were sown from each plant, the probability that the results were due to chance alone was less than 0.01. As Orel and Hartl remarked, Mendel had stated that these experiments had been conducted in a manner quite similar to the preceding one. They also noted that there had been sufficient space in Mendel's garden for more than 5000 plants and that this would have allowed Mendel to have planted more than 10 seeds per plant.

Perhaps what Mendel meant by saying that the method in the second series was "quite similar" to that in the first is that he allocated a certain plot of his garden for the purpose of progeny testing and cultivated as many plants per parent as he could to fill this space. If the space allocation was adequate for 6,000 plants then, in the second series of progeny tests, Mendel could have cultivated an average of 12.7 seeds per parent. He may well have cultivated more than 6,000 plants in the second series because he commented on the amount of work it required, saying that "of all experiments it required the most time and effort." Hence, Fisher's dismissal of the explanation based on more than 10 progeny is too facile, especially in light of Mendel's vague specification of how similar the two experiments actually were and the plausible alternative interpretation of Mendel's text. Fisher's "abominable discovery" is therefore much less damaging than first appears. In short, although Mendel's expectations are certainly wrong, Fisher's expectations may be wrong as well. Thus, the uncertainties in the experiment and the ambiguities in the analysis discredit any inference of deliberate manipulation or falsification of data. (Orel and Hartl 1994, *197*)

Fisher's analysis received support, however, from the work of Charles Novitski (1995). Novitski performed a Monte Carlo simulation of the experiments on gametic ratios, which had enhanced the "too good nature" of Mendel's results (see table 1.6). He generated random events corresponding to Mendel's five experiments on gametic ratios. He considered each set of five experiments as a trial and found that in 100,002 trials only 138, or about 1/725, gave χ^2 values less than the 3.673 calculated by Fisher for these experiments. This agreed quite well with the value 1/728 calculated from the χ^2. Novitski found that Mendel's results were, in fact, even more unusual. He examined the χ^2 values for each of Mendel's experiments and found that in only 47 of the 100,002 trials was the largest χ^2 value as small as the largest value, 1.0843, calculated from Mendel's data. He also computed the variance of Mendel's χ^2 values, which was 0.2183, and found from his simulation that in only 21 of 608 trials with χ^2 between 3.6 and 3.7, the Mendel value, was the variance as low as that of Mendel's data. He estimated that the probability of getting Mendel's results for these experiments was closer to 1/20,000 rather than 1/700. He also concluded that none of the explanations previously offered was sufficient to explain Mendel's results.

An extensive discussion of both Mendel's data and of Fisher's analysis was provided by Teddy Seidenfeld (1998; chapter 6 of this volume). He noted, "Fisher did not write 'Has Mendel's Work Been Rediscovered?' either to question Mendel's integrity or to challenge his rightful place among those at the center of modern genetics" (*216*). He also remarked that Mendel had a good idea of the results he expected. "It hardly needs saying that, therefore, Mendel had well-grounded expectations for his experiments on single (and even double) factor trials involving the 5 plant characteristics, since he had seen the parallel results (at much larger sample sizes) for the two pea characteristics a year earlier" (*219–20*). Seidenfeld began with a discussion of the misclassification problem in the 2:1 ratio experiments. It was these experiments and the agreement with the "wrong values" that had triggered Fisher's analysis. He remarked that Fisher had argued against the view that Mendel had a selection bias because: "(i) It does not apply to the trifactorial study, where all plants were classified," and "(ii) It is implausible that bias was equally effective for all five characteristics." He continued, "Like Fisher, I find the coincidences of perfectly offsetting selection biases (the second rebuttal point) more difficult to believe even than the alternative that the data were 'cooked'!" (*222*).

Seidenfeld proposed a new solution to this problem, one he thought compatible with Mendel's own statements. He suggested that when Mendel continued his experiments into subsequent generations, something Mendel stated quite explicitly that he had done, he included an examination of some of the F_4 plants grown from the F_3 plants that showed the dominant trait. If Mendel had tested three such plants, the probability of misclassification was reduced to less than 2×10^{-3}, rather than 0.0563. For only one such plant, the probability is reduced to less than 0.01.[58] Seidenfeld argued, "Thus, one way around Fisher's first objection is to hypothesize that Mendel used an elementary sequential design" (*223*). Seidenfeld pointed out that a sequential experimental design was quite feasible given the size of Mendel's garden.

Seidenfeld also discussed Weiling's attempt to solve this problem either by suggesting that only 8 of 10 seeds germinate, which, as already noted, worsens the problem, or by invoking a hypergeometric distribution.

> Among several difficulties I have with Weiling's statistics, I do not understand the basis for his use of the hypergeometric distribution. It is true, as he writes, that the process of choosing 10 of 30 particular seeds from a plant (as Mendel is posited to have done to make the 10 F_3 offspring per F_2-parent) follows a hypergeometric distribution, with a smaller variance than the i.i.d. Binomial distribution. However, under Mendelian theory, these 30 seeds follow the i.i.d. Binomial distribution. Hence, the net (marginal) distribution for the 10 seeds, chosen from the 30, is again i.i.d. Binomial, not hypergeometric, contrary to what Weiling asserts. The challenge, taken up below, is to justify the claim that the 30 seeds are not an i.i.d. sample from the Binomial distribution. (Seidenfeld 1998, *251n17*)

The question of "too good to be true" data was also discussed by Seidenfeld. He raised several technical objections to Fisher's analysis, particularly the absence of an adequate theory of Fisherian significance testing. Moreover, he discussed the issue of whether Fisher should have used the "corrected" values or Mendel's "incorrect" expectations in computing the χ^2 for the 2:1 ratio experiments. He suggested that it should have been the "corrected" values and calculated the total χ^2 and found, in agreement with Weiling, that the value increased to 48.78 from 41.61, with the probability of a worse fit decreasing from 0.99993 to 0.9975.

Seidenfeld also proposed an alternative model that reduced the variance in Mendel's data, the Correlated Pollen model. This is quite similar to the tetrad-pollen model proposed earlier by Thoday and Beadle.[59] He calculated that this model, under reasonable assumptions, yielded a variance of 74% of that expected for a binomial distribution. This reduced the one-sided probability to approximately 0.96. This was still quite large, but it was not the extraordinary value of 0.99993 obtained by Fisher, and would go some way toward resolving the "too good to be true" aspect of Mendel's data. As noted earlier by Edwards, and discussed in detail below, there are other aspects of Mendel's data in need of explanation.

Seidenfeld remarked that there were two important questions concerning this model. "First, is there evidence to confirm or to refute the speculative genetics that Mendel's peas are not independently distributed within self-fertilizing pods? Second, does it matter to Fisher's analysis if the model of pea genetics is not quite Mendelian but, instead, reflects this alternative distribution of pollen cells? How much of Fisher's .9999 P-value can be explained away with some subtle correlation among the pollen?" (*231*).

With regard to the first question, Seidenfeld noted that Edwards (1986a) had argued that the experiments of Bateson and Kilby (1905) and of Darbishire (1908, 1909) had shown 95% of the variance expected on a binomial model. His own, slightly different, analysis gave 94% of the Mendelian variance. He noted, however, that the number of peas per plant reported by Mendel, 29, was far lower than the 106.5 in the Bateson-Kilby study and the 217.6 reported by Darbishire. A question to be answered, however, is whether peas grown under more severe conditions might exhibit a reduced variance.

Seidenfeld also demonstrated that Mendel's reported result of no more than five recessive seeds in a pod was more likely on the Correlated Pollen model than on a binomial model.[60] He also remarked that the model had no effect on the excellent fit in the gametic experiments, in which artificial fertilization was used.

One important discussion included in Seidenfeld's paper was on the possible model of cheating. In agreement with Edwards, he found that the data "were adjusted (rather than censored) to avoid extreme segregations in the record" (*242*). He presented the results of Edwards's analysis in deciles of probability values (figure 1.3) and noted that there was a deficit in the lowest deciles, as one would expect if extreme values were omitted, but there is no excess at high probability values. There is, rather, a bulge near the median values. This is also true for Seidenfeld's own analysis of Mendel's experiments (figure 1.4).

> What model of cheating, then, can the reader propose that replaces extremely discrepant outcomes with ones clustered about the median of χ^2s? I challenge the reader to try to adjust binomial data from sample sizes in Mendel's experiments, so that the following three features appear in the resulting distribution of P-values from the (1df) [degree of freedom] χ^2s.
>
> 1. There is a significant reduction in the left-tail of the Ps.
> 2. There is no significant departure from uniformity in the right tail of the Ps.
> 3. There is a significant concentration of the Ps about their median, i.e., about .50.
>
> To be fair to Mendel, this exercise should be attempted without the aid of χ^2 Tables, which distribution, the reader recalls, K. Pearson discovered only in 1900! (Seidenfeld 1998, *242–43*)

Seidenfeld concluded that his suggestion of sequential experiments by Mendel removed the problem of Mendel's fit to the "wrong" values in the 2:1 ratio experiments. "Regarding the '3:1' and '2:1' laws of segregation for self-fertilizing hybrids, I propose the Correlated Pollen model, an alternative to the usual Mendelian (i.i.d.) distribution of peas in a pod. The C-P model has the same first moment as the Mendelian model with about 75% of the Mendelian variance. This speculative model is enough to recover the P-values in Mendel's data for each of the two, main Mendelian laws" (*245*). He admitted, however, that there was no support for the model in the large-scale studies of Bateson and Kilby and of Darbishire. On the other hand, one of Darbishire's experiments, which had a smaller yield of peas per plant, closer to that of Mendel, also showed an anomalous small sum of χ^2s. Seidenfeld offered no explanation for the excellent fit in the gametic experiments: "Where do we stand, more than sixty years after Fisher's shocking allegations against the authenticity of data in Mendel's paper? The allegation of misclassification (of hybrids) admits such a straightforward reply that I no longer find merit in that aspect of Fisher's criticism. But, unless some alternative model with reduced variance, like the C-P model, can be justified, I see little hope of explaining away the Ps that are 'too good to be true'" (*244–45*).

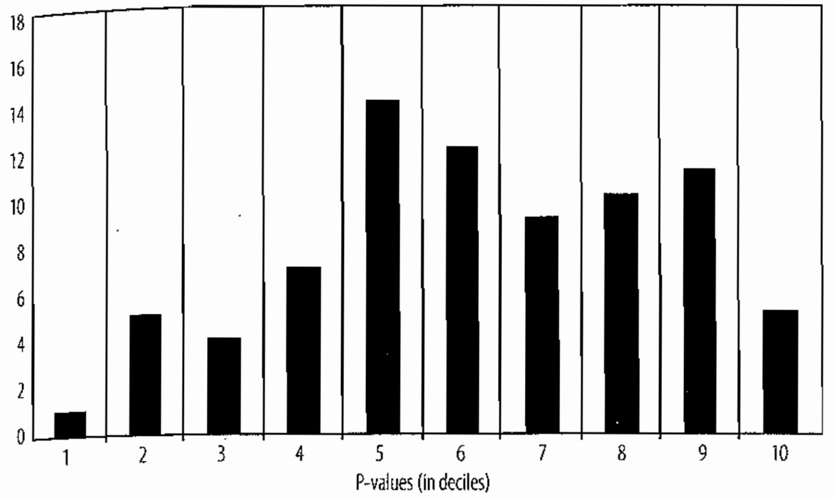


FIG. 1.3 P-values, by deciles, in Edwards's 1-df partition of Mendel's data (N = 78)

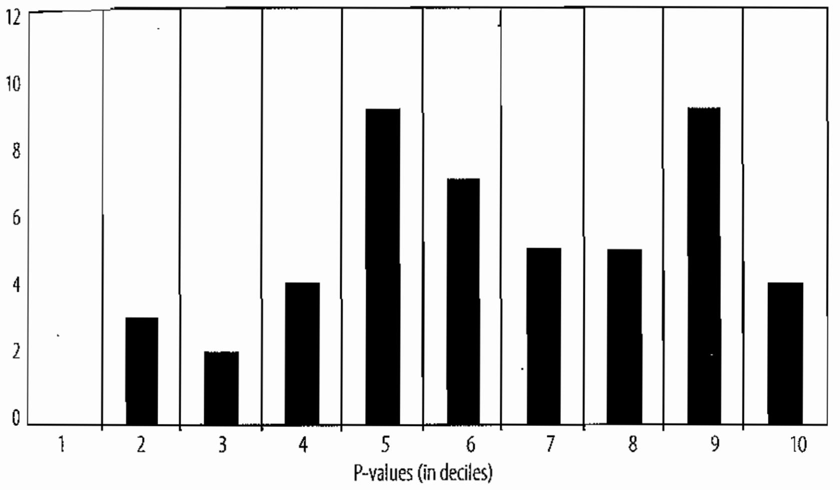


FIG. 1.4 Mendel's data (by experiment) (N = 48)

Seidenfeld also remarked that the C-P model could be subject to field trials. "Regardless the outcome, no matter how peas self-fertilize, I urge the reader to study Mendel's classic paper and Fisher's provocative article. Mendel's work is a standard of clarity and a delight for its intelligent, sequential designs. Fisher, as always, is a brilliant statistician and imposing geneticist. As with many of his other writings, coming to an understanding of how he argues is the key, regardless what the reader thinks, in the end, of his conclusion" (246).

In 2001, Daniel Fairbanks and Bryce Rytting published an extensive review of the Mendel-Fisher controversy. In discussing Mendel's work they noted:

> There is substantial disagreement about his objectives, the accuracy of his presentation, the statistical validity of his data, and the relationship of his work to evolutionary theories of his day. In the following pages we address five of the most contentiously debated issues by looking at the historical record through the lens of current botanical science: (1) Are Mendel's data too good to be true? (2) Is Mendel's description of his experiments fictitious? (3) Did Mendel articulate the laws of inheritance attributed to him? (4) Did Mendel detect but not mention linkage? (5) Did Mendel support or oppose Darwin? (Fairbanks and Rytting 2001, 265)

For the issue I want to address, the most important section of this paper is "Are Mendel's Data Too Good to Be True?" As Fairbanks and Rytting noted, Fisher believed that the fit of Mendel's data to his expectations was so good that it must be questioned. They further commented that Fisher's analysis was based on consistently low χ^2 values, and that his most telling point was Mendel's close fit to an incorrect value in the 2:1 ratio experiments. For those experiments, Fairbanks and Rytting continued, Mendel had reported a total of 399 heterozygous plants to 201 homozygous plants, an excellent fit to the 2:1 ratio expected. As Fisher had pointed out, the correct expectations were 377.5:222.5. These authors asserted that the observation was not, in fact, a statistically significant deviation and that the probability that it was due to chance was 0.0692. They stated that one needs to take half this value (0.0346) to obtain the chance of a deviation of this magnitude, and in the direction toward Mendel's expectation, which agreed with Fisher's estimate of 1 in 29.

Fairbanks and Rytting also addressed Weiling's claim that one could not use Fisher's independent model, but that one should use a model of balls selected from an urn without replacement. They, and Seidenfeld, argued that Weiling was wrong.

> Weiling argued that Mendel sampled ten seeds per plant without replacement in the F_3 progeny tests, and that the sampling, therefore, was not independent. He assumed that the average pea plant in Mendel's experiments had 30 seeds per plant, 23 of which had the dominant phenotype ($0.75 \times 30 = 22.5$, rounded to 23). Based on this assumption, Weiling determined the average probability of misclassification was $23/30 \times 22/29 \times 21/28 \times 20/27 \times 19/26 \times 18/25 \times 17/24 \times 16/23 \times 15/22 \times 14/21 = 0.0381$, instead of 0.0563 as determined by Fisher.

However, although Weiling's estimate is correct for a plant with 30 seeds, 23 of which have the dominant phenotype, it cannot be used to estimate the average probability of misclassification for a population of plants. For any particular number of seeds per plant, the average probability of misclassification must be determined as the sum of the probabilities of misclassification for all possible combinations weighted by the expected frequencies of those combinations according to the binomial distribution. When this is done, the average probability of misclassification is consistently 0.0563. In other words, if Mendel's data are from random seed samples collected from a binomially distributed population, Fisher's estimate of 0.0563 as the probability of misclassification is correct, even when the effect of sampling seeds without replacement is taken into account. (Fairbanks and Rytting 2001, 274)

Fairbanks and Rytting further discussed the possibility of an experimental test of the urn model in peas: "There is currently no empirical evidence to support the urn model in *Pisum*, but it is one that can be empirically tested because it should produce a significant deviation from a binomial distribution for phenotypes of individual seeds from the same pod (or plants grown from those seeds). We have initiated the necessary experiments but do not yet have the results" (279).[61]

The question of whether Mendel had scored exactly 10 plants was also discussed: "The proximity of Mendel's F_3 progeny data to an incorrect expectation is not as questionable as it might seem when viewed in a botanical context" (274). Fairbanks and Rytting suggested that Mendel had probably sown more than 10 seeds to guard against losses due to germination failure. Had Mendel done so, he could have scored seedlings for two of the plant characters: stem length and seed-coat color. The latter was perfectly correlated with axial pigmentation. This would have mitigated the "too good" effect because these two experiments had provided almost half the total deviation of that set of experiments. They pointed out that the same misclassification problem also applied to the trifactorial experiment. They cited Orel and Hartl (1994), who had also presented arguments for Mendel having cultivated more than 10 seeds per plant and that there was sufficient room in Mendel's garden to allow this. Fairbanks and Rytting posited: "When these statistical and botanical aspects of Mendel's F_3 progeny tests are considered, there is no reason for us to question his results from these experiments. *However, we must still account for the bias that is evident when the data for all of the experiments that he reported are compared as a whole*" (276, emphasis added).

They discussed the usual explanations for the goodness-of-fit bias, including the theory that Mendel stopped counting, and found all these arguments insufficient. The most likely explanation of the "goodness of fit," Fairbanks and Rytting suggested, was that Mendel had selected his best data for presentation in a public lecture and its subsequent publication: "We believe that the most likely explanation of the bias in Mendel's data is also the simplest. If Mendel selected for his presentation a subset of his experiments that best represented his theories, χ^2 analyses of these experiments should display a bias. His paper contains multiple references to experiments for which he did not report numerical data, particularly di- and trihybrid experiments" (279–80). They also cited as supporting evidence a letter from Mendel to Nägeli: "In Mendel's second letter to Nägeli, he referred to his paper as 'the unchanged reprint of the draft of the lecture mentioned: thus the brevity of the exposition, as is essential for a public lecture'" (qtd. in Stern and Sherwood 1966, 61). They admitted the bias in Mendel's data but stated, "there are reasonable statistical and botanical explanations for the bias, and insufficient evidence to indicate that Mendel or anyone else falsified the data" (281).

This plausible explanation would help to lessen the overall "too good to be true" character of Mendel's data.[62] It does not, however, address the demonstration, by both Edwards and Seidenfeld, that Mendel's data had not merely been truncated, but adjusted.

Kenneth Weiss (2002) accepted the views of Fairbanks and Rytting on the question of Mendel's research integrity. In discussing the 2:1 ratio experiments, he accepted their calculation that the results for the monohybrid experiments (399:201) are only slightly unlikely and that this exonerated Mendel from the most serious charge of falsification. He did not, however, discuss the question of the overall goodness of fit. Instead, he commented, "I think the whole issue has been greatly overblown in the first place. The reasons to lighten up on the poor monk have to do with his context, with statistical considerations, and with what science is all about" (42). He suggested that because modern statistical tests were unavailable to Mendel that this would excuse the falsification of data. He also agreed that Mendel's statements that he reported only part of his data were evidence against fraud. "And even if he or his assistant selectively pitched (or nibbled) some peas, well, before I put blueberries on my cereal I discard the green ones, but I still say that blueberries are blue" (42). Weiss found no distinction between fraudulent data in scientific experiments and an unwillingness to eat unripe blueberries. In his view, the correctness of Mendel's results provides justification for such behavior. "And he was right: 2:1 makes theoretical sense; 1.88:1.11 makes none" (43).

Weiss's views were criticized by DeGusta (2003a; 2003b). He agreed that Mendel's data were clearly biased, citing Fisher, Fairbanks and Rytting, and Piegorsch. DeGusta noted, "While Mendel was a monk, he was no saint. At least some of the data he published are clearly biased beyond what would be expected by chance. Furthermore, this bias is uniformly in the direction of Mendel's expectations, which has long raised suspicion that its source was something other than innocent error" (2003b, 1). Contrary to Weiss, he argued that the lack of available statistical tests did not provide an excuse for fraudulent data: "The beauty of statistical methods (probably their only beauty) is that the results are independent of any particular context. . . . Of course, as Weiss notes, statistical tests of significance (i.e., the chi-squared test) were not known to Mendel. But such tests cannot *prevent* bias, they can only *detect* it after the fact. So the lack of such tests in Mendel's time does not exculpate any 19th century data manipulators" (2003a, 5). He further remarked that Fairbanks and Rytting had accepted that there was bias and he agreed that their explanation of Mendel's goodness of fit—the selection of that data that best represented his theories—was the best available.

In *The Great Betrayal: Fraud in Science* (2004), Horace Judson discussed the Mendel-Fisher controversy. Citing the work of Corcos and Monaghan, he concluded that Mendel had been vindicated.

Fisher's analysis of the 2:1 ratio experiments was questioned by Edward Novitski (2004). Novitski suggested that Mendel would have accepted as evidence for heterozygosity plants with fewer than 10 progeny, if one of those progeny showed the recessive trait. On the other hand, he would have excluded plants with fewer than 10 progeny if all of them were dominant. "If Mendel demanded 10 progeny for an adequate test only when the 9 or fewer existing progeny failed to reveal the heterozygosity of the parent, we can reasonably assume that when fewer than 10 progeny matured, he would be eliminating primarily *AA* (homozygous) individuals" (1134). This would tend to compensate for the loss of heterozygous plants according to Fisher's correction. Suppose that Mendel excluded all plants with 9 dominant progeny. One-third of the time they would have been homozygous parents. Two-thirds of the time the parent would have been heterozygous. The probability of obtaining 9 dominant progeny from a heterozygous parent is $(3/4)^9 = 0.075$. Thus, the chance of excluding homozygous parents compared with excluding heterozygous parents is $(1/3)/[(2/3)\ (0.075)] = 6.67$. Homozygous plants are preferentially excluded if one excluded sets with 9 progeny. Similar results occur with sets with fewer progeny. This would mitigate the undercounting of heterozygous plants pointed out by Fisher. Novitski assumed that Mendel would have cultivated other sets of 10 progeny to make up for those lost to germination failure. Including those sets would increase the number of heterozygotes counted. In fact, if the failure rate were high enough, then the expected ratio of heterozygous to homozygous plants would be greater than 2:1: "The selective elimination of *AA* [homozygous] individuals would shift the ratio calculated by Fisher from 1.8874 *Aa* to 1.1126 *AA* toward the ideal of 2. In fact, if the failure rate were high enough the ratio might well exceed 2:1" (1134). Novitski presented a numerical example in which one looked at 100 plants with 10 seeds each. For a 2% failure rate, he reported a calculation by C. E. Novitski, who found that 22.5 heterozygous plants would have been excluded by Fisher's correction and 23 added because of the procedure to deal with failures. "The similarity of the loss (22.5) and the gain (23) is purely fortuitous and not to be taken too literally. The essential point is that the two values are of similar magnitude and of opposite effect. They do show, however, how these two complicating factors, working in opposite directions, could give Mendel a final ratio closer to the 2:1 theoretically expected, Fisher's analysis notwithstanding" (1134).

There are, however, two serious problems with this scenario. First, Mendel did not report any such procedure. Second, Mendel's failure rate was, as discussed earlier, 5.6%, which would raise the corrected ratio to larger than 2:1. Recent work by Daniel Hartl and Daniel Fairbanks (Hartl and Fairbanks 2007; see also postscripts to chapters 5 and 7 in this volume) presented a detailed calculation supporting this point.

With regard to the trifactorial experiment, Novitski remarked, "It can be argued, however, that Mendel was almost certainly using the correct expectation, and it is Fisher who was using the incorrect one" (1135). Novitski argued that Fisher had misread Mendel's paper and that the third characteristic used in this experiment was seed-coat color rather than flower color. Novitski remarked that the hybrid seed coats are spotted even when that trait was absent in the parent and that "it is reasonable to assume that the maternal plant might be classified unambiguously as either a homozygote or a heterozygote" (1135). In other words, for this characteristic, the genotype could often be inferred from the phenotype. (Hartl and Fairbanks [2007] suggest that this is probably incorrect.)

Novitski noted, however, that his analysis did "not alter Fisher's conclusions that, overall, Mendel's results are closer to theory than expected on a chance basis" (1136). He cited the results of both Edwards (1986a) and C. E. Novitski (1995) as support for this conclusion. He further suggested that Mendel might have selected his data, or even altered it, to make his results more understandable to his audience and noted that Mendel expected and welcomed repetitions of his experiments. He concluded: "Fisher's criticism of Mendel's data—that Mendel was obtaining data too close to false expectations in the two sets of experiments involving the determination of segregation ratios [2:1]—is undoubtedly unfounded" (1136).

In an adjoining paper, Charles Novitski (2004) presented the details of his calculations. Using the binomial distribution, Fisher's correction, and a failure rate of 2%, he found that R, the ratio of those counted as heterozygotes to those counted as homozygotes was 2.068, which was in reasonable agreement with Mendel's expectation of 2. By minimizing the χ^2 and allowing the failure rate to float, he found a failure rate of 1.54% and R = 1.975. I note, however, that for the more accurate estimate of Mendel's failure rate, 5.6%, the problem is exacerbated. R would then be considerably greater than 2, and would, in fact, be closer to 3.

In an article on the use of statistical techniques to detect fraud, Michael O'Kelly (2004) cited Fisher's study of Mendel as the pioneering use of such techniques to investigate potential fraud. Similarly, in "A Beginner's Guide to Scientific Misconduct," Bob Montgomerie and Tim Birkhead (2005) stated, "Gregor Mendel might well be called the father of scientific misconduct, not because he was necessarily a wrongdoer, but because his published work sparked more than a century of controversy about the validity of his data" (17). After a very brief discussion of the controversy, they cited Fairbanks and Rytting (2001) and concluded, "More recent analyses appear to have exonerated Mendel in any wrongdoing, though the details are complex and not entirely convincing" (17).

Finally, after describing Mendel's achievement as "one of the most brilliant in the entire history of science," Yongsheng Liu (2005, 314) went on to state, "It is well known that truth is the essence of science. Mendel's gravest mischief was that he cooked his experimental data" (315). After briefly discussing the controversy, Liu summed up, "The controversy over the 'too good to be true' data remains unresolved and has been the subject of passionate scientific debate" (315).

Recently, Fairbanks has reported experimental results on peas (see postscript to chapter 7 in this volume). In a letter to this author, he explained: "All analyses (chi-square analysis and linear regression) show no evidence of a bias toward expectation for distributions of phenotypes within pods, and thus no evidence that the tetrad-pollen model (sampling without replacement) is valid. In fact, the pattern that emerged conforms to the one expected with completely random sampling of gametes with replacement. In other words, the pea plant is, as Edwards put it, an excellent randomizer and there is no reason to suspect that gamete sampling is responsible for the bias in Mendel's experiments." Seidenfeld also reported that an experiment to investigate whether pea plants grown under severe conditions exhibit reduced variability has, unfortunately, not yielded usable results (see the tables in the appendix to chapter 6 in this volume). It is interesting to note that one of the preliminary experiments yielded an excellent fit to the expected 9:3:3:1 ratios for that experiment. The data were 156:54:54:18. The probability of a worse fit is 0.9919. Likewise, in the summed data of his test of the tetrad-pollen model, Fairbanks examined a total of 5,204 F_2 plants for flower/axillary pigmentation and scored 3,899 as pigmented and 1305 as not pigmented. The probability of a worse fit to a 3:1 ratio is 0.8981. Extraordinarily good results do occur.

Hartl and Fairbanks (2007 and postscripts to chapters 5 and 7 in this volume) recently argued, quite persuasively, that Fisher's analysis of the 2:1 ratio experiments is incorrect. They suggested that there is no factual basis for suspecting any tampering with the data from the first series of experiments on the 2:1 ratio. They showed that examination of Mendel's data reveals no significant deviation from Fisher's expectations, and, that for the one experiment repeated by Mendel (Experiment 5 on pod color), the pooled results of both experiments fit Fisher's expectation almost exactly. They also noted that Mendel's contemporaries agreed that he was a superb gardener who would certainly have known that in some of the experiments fewer than 10 seeds would germinate. They suggested that Mendel planted 2–3 seeds per hill and thinned them sometime after germination or that he planted more than 10 seeds in the greenhouse and then from the resulting seedlings selected 10 to transplant to his garden. In this way, Mendel could guarantee 10 progeny per parent plant.

They also offered two plausible explanations of the trifactorial experiment, which also involves the 2:1 ratio. They suggested that Mendel used axillary pigmentation rather than flower color, as stated by Fisher. Although these characteristics are perfectly correlated, axillary pigmentation can be easily identified in seedlings, whereas flower color cannot. Thus, the entire experiment could be completed within a single growing season. They further suggested that Mendel had planted more than 10 seeds in order to guarantee at least 10 seedlings and that he might have scored all of the seedlings for axillary pigmentation. They calculated that if Mendel had scored as few as 11 seedlings for 70% of the progeny plants and only 10 seedlings for the remaining 30%, then any significant discrepancy from Fisher's expectations disappears. Hartl and Fairbanks noted that this was consistent with what Mendel had written. He did not state that the experiment had been done in the same way as the monohybrid experiments, but referred only to "further investigations." The second explanation was that Mendel might have been able to identify genotypes by observing the phenotypes. This had been suggested by both Wright (1966) and by E. Novitski (2004). Novitski had suggested that the observed trait was seed-coat color and that this trait could be used to distinguish heterozygotes from homozygotes in plants with the dominant phenotype, which Hartl and Fairbanks argued, on botanical grounds, is probably not correct. Nevertheless, they showed that even a 2% observation rate of such a character would eliminate any significant discrepancy between Mendel's data and the correct expectations.

Hartl and Fairbanks concluded, "Let us hope against all experience that Fisher's allegation of deliberate falsification can finally be put to rest, because on closer analysis it has proved to be unsupported by convincing evidence" (2007, 13).

Conclusions

It has now been 140 years since the publication of Mendel's seminal work, 70 years since Fisher's analysis, and more than 40 years since the beginning of the Mendel-Fisher controversy. Is the issue still unresolved? Perhaps more importantly, should it be? I think the answer is no.

There are, I believe, several conclusions that are supported by both this reexamination of the controversy, as well as the articles included in this volume. They are: (1) Mendel was not guilty of deliberate fraud in the presentation of his experimental results; (2) the problem of the 2:1 ratio experiments has been solved; (3) Fisher's analysis of the "too good to be true" data is still correct; and (4) Fisher would have been quite unhappy with those who used his work to diminish Mendel's achievement.

Let us consider first the question of whether Mendel committed deliberate fraud. I believe the evidence is overwhelmingly against such a conclusion. Mendel published data that disagreed with his expectations in both his original paper on plant hybridization and in his paper on *Hieracium*. He also made it clear that he had not included significant amounts of data. In addition, he sent 140 packets of seeds to Nägeli in the hopes that his experiments would be replicated. These are not the actions of someone guilty of fraud. Perhaps, more importantly, Seidenfeld's challenge to provide a method of cheating that would reproduce the oddities in Mendel's data shown by Edwards and himself is still unanswered. I also believe that these oddities argue against other possible explanations of Mendel's results such as unconscious bias, stopping the count when results agreed with expectations, or fraud by another.

With regard to Fisher's analysis of the 2:1 ratio experiments, I believe that Seidenfeld's explanation that Mendel used sequential experiments in subsequent generations provides an adequate solution. This is consistent with Mendel's statement that he had continued the experiments for several generations. In addition, as Seidenfeld showed, it would take examination of only very few such plants to reduce the misclassification problem to negligible proportions. I also believe that the calculations of Hartl and Fairbanks for both the monohybrid and the trifactorial experiments argue strongly against Fisher's analysis of these experiments. Both explanations may, in fact, be correct. Either argues persuasively against Fisher's analysis of the 2:1 ratio experiments. It is also possible, as suggested by many authors including Orel and Hartl (chapter 5) and Fairbanks and Rytting (chapter 7), that Mendel sowed more than 10 plants in these experiments. That explanation is inconsistent with what Mendel explicitly stated in his report of the early monohybrid experiments, but it is a plausible, although I believe a less probable, explanation of the 2:1 ratios in the trifactorial experiment. Thus, both sets of 2:1 ratio experiments, the experiments that had initially triggered Fisher's suspicions, can be explained without any fraud.

The issue of the "too good to be true" aspect of Mendel's data found by Fisher still stands, however. No one has yet raised any valid criticism of Fisher's analysis, or of the later analyses by Edwards and Seidenfeld. There is also no empirical evidence to support a botanical explanation of reduced variability such as the tetrad or urn model. In fact, good evidence has been provided by both Edwards and Fairbanks that peas are good randomizers. The analyses by Seidenfeld and by Hartl and Fairbanks argue that Fisher's use of 2:1 as the ratio in the monohybrid and trifactorial experiments is correct. In addition, as Weiling showed, there is little difference to the goodness-of-fit problem whether one uses 2:1 or 1.7:1 for the expected ratio in the monohybrid and trifactorial experiments.

Finally, it seems clear that Fisher would have been quite unhappy with those who have used his work to diminish Mendel's achievement. As we have seen, Fisher was unstinting in his praise of both Mendel's methods and his conclusions. As Fisher himself said in describing Mendel's experiments, they are "experimental researches conclusive in their results, faultlessly lucid in presentation, and vital to the understanding not of one problem of current interest, but of many" (Fisher 1936, *139*).

It is time to end the controversy.

NOTES

I want to thank my collaborators Anthony Edwards, Daniel Fairbanks, Daniel Hartl, and Teddy Seidenfeld for valuable discussions, for their comments and always constructive and gentle criticism, and most importantly, for their work, which gave me a sufficient understanding of the issues involved in the controversy so that I could write this essay. I am grateful to Anthony Edwards for both his wonderful hospitality during my research trip to Cambridge and for sharing his files on the Mendel-Fisher controversy with me. Our discussions were invaluable.

1. All page references to papers that are included in this volume refer to those versions contained in this book, and are noted in italics for ease of use. For all quoted material, all italics are from the original source unless otherwise noted.

2. The translation of Mendel's 1865 paper reprinted in this volume is contained in Bateson 1909, which is the same translation used by Fisher.

3. Not all of the experiments described in the same section of Mendel's paper necessarily occurred at the same time.

4. In modern notation, self-fertilization of a heterozygous plant is genetically equivalent to the cross $Aa \times Aa \rightarrow AA + 2Aa + aa$. Both the AA and Aa plants will display the dominant character, whereas an aa plant will display the recessive character. Mendel used A, Aa, and a, respectively, to denote genotypes we currently symbolize as AA, Aa, and aa. For a detailed discussion of this see chapters 5 and 7. I will use Mendel's notation of A and a for dominant and recessive characters, respectively, rather than the modern AA and aa.

5. Mendel noted that care must be taken in these experiments. He stated, "These two experiments are important for the determination of the average ratios, because with a smaller number of experimental plants they show that very considerable fluctuations may occur. In counting the seeds, also, especially in Expt. 2, some care is requisite, since in some of the seeds of many plants the green colour of the albumen is less developed, and at first may be easily overlooked. . . . In luxuriant plants this appearance was frequently noted. Seeds which are damaged by insects during their development often vary in colour and form, but, with a little practice in sorting, errors are easily avoided" (*87*).

6. This rather unexpected result is discussed in some detail in Seidenfeld (1998), which is reproduced in this volume as chapter 6.

7. This is true for the plant characteristics. The seed characteristics would appear in the same generation.

8. The F_3 generation experiments are the first experiments on the 2 to 1 ratio that would be of concern to Fisher. A second instance is the trifactorial hybrid experiment, discussed below.

9. There is considerable discussion in the ensuing controversy over whether the German phrase *von jeder 10 Samen angebot* should be translated as "10 seeds were sown" or "10 seeds were cultivated." In their translation, Stern and Sherwood (1966) used "sown." The choice of 10 seeds is also of significance. Fisher would later argue, as discussed below, that because only 10 seeds were planted, that there is a 5.6% probability that heterozygous plants would be classified as homozygous and thus be undercounted. Because of this, he argued that the ratio should be 1.8874 to 1.1126, or about 1.7 to 1, rather than 2 to 1.

10. This is a reasonable reading of what Mendel wrote, and it was Fisher's interpretation. There is considerable discussion in the later literature about whether Fisher was correct. This is discussed in some detail below and in the other chapters in this volume.

11. A modern statistician would not regard this as a significant deviation.

12. Mendel did not present all of his data from several of his other experiments. In ad-

dition, as discussed below and in chapter 6, this continuation of the experiments is an important point in the discussion of the 2 : 1 ratio experiments.

13. The remainder of this section of Mendel's paper discusses the reversion to parental forms, which is not important for our story.

14. In other experiments, discussed below, Mendel would investigate whether there was any difference in results depending on which characters were associated with the seed and pollen plants, respectively. He would conclude that there was no difference.

15. These results are a good fit to the expected 9 : 3 : 3 : 1 ratio.

16. Although he did not take any notice of this, Mendel obtained results for the failure rates for growing plants from seeds, a quantity which will be of some importance in our later discussion. From 315 round and yellow seeds a total of 301 plants were obtained, a failure rate of 4.4%. For wrinkled and yellow seeds, round and green seeds, and wrinkled and green seeds, the results were 96 plants from 101 seeds (5.0%); 102 plants from 108 seeds (5.6%); and 30 plants from 32 seeds (6.2%), respectively.

17. These experiments are other instances of the 2 : 1 ratio experiments.

18. I interpret this as indicating that Mendel used the same procedure that he had used previously to investigate plant characters, i.e., that he grew 10 plants in the next generation for each plant that showed the dominant gray-brown seed-coat color. The failure rate for growing plants from seeds was 7.0% for this experiment. Hartl and Orel (1992) suggest that Mendel may have meant that he allotted the same amount of space for the experiments and that there was sufficient room for Mendel to have sown more than 10 plants.

19. Notice again the excellent feel that Mendel has for his data. He sees the significant pattern and neglects the small deviations from that pattern. This is true of all of his analyses.

20. The failure rates were 8.2% and 7.4%, respectively.

21. Mendel remarked that this phenomenon had also been emphasized by C. F. Gärtner but gave no reference.

22. Hawkweed was not a good choice for Mendel, although he was not aware of this, at least when he began his hawkweed experiments. It sometimes produces seeds by apomixis, in which seeds are produced from unfertilized ova. In this process, genes are inherited only from the female parent and would not exhibit Mendelian patterns.

23. This was an exaggeration.

24. This was a reference to Pearson (1900), in which the χ^2 test was introduced.

25. In a letter to Karl Pearson, Weldon remarked, "If only one could know whether the whole thing is not a damned lie!" (Karl Pearson Papers, University College London, file 625). Weldon was discussing Mendel's entire scheme, however. He does not suggest that Mendel's results were fraudulent.

26. Weldon's reservations were too much for Bateson, the arch Mendelian. Within a month of the publication of Weldon's article, Bateson published *Mendel's Principles of Heredity: A Defence* (1902) in which he devoted more than 100 pages to the attempted refutation of Weldon. Citing Weldon's statement that he wished to "suggest fruitful lines of inquiry," Bateson concluded: "In this purpose I venture to assist him, for I am disposed to think that unaided he is—to borrow Horace Walpole's phrase—about as likely to light a fire with a wet dish-clout as to kindle interest in Mendel's discoveries by his tempered appreciation" (Bateson 1902, 208).

This was just the latest salvo in the bitter and nasty battle between the biometicians, headed by Pearson and Weldon, who used statistical methods on large populations and supported Darwinism and blending inheritance, and the Mendelians, led by Bateson, proponents of applying Mendel's laws to small populations. (For details of this controversy see

Provine [1971], Froggatt and Nevin [1971a, b], Farrall [1975], Kevles [1980], and Morrison [2002]). Fisher, himself, remarked that this battle was "one of the most needless controversies in the history of science" (1924, 192) and David Hull (1985) referred to it as "an inexplicable embarrassment."

Fisher himself became a casualty of this battle. In 1916 he wrote a paper entitled, "On the Correlation Between Relatives on the Supposition of Mendelian Inheritance," which showed that the two opposing views could be reconciled. This became one of the founding papers of population genetics and was referred to by Kempthorne as "a work of genius" (qtd. in Norton and Pearson 1976, 151). Fisher originally submitted the paper to the Royal Society of London and it was sent to Karl Pearson, a biometrician, and to Reginald Crundall Punnett, a Mendelian, for refereeing. Both had reservations about the paper. Pearson thought that Fisher's assumptions were not supported by either observational or experimental evidence, and that his assumption was only one of many that might lead to similar results. He suggested that the paper's "publication should depend on whether Mendelians consider its hypotheses of value as actually representing observational facts" (qtd. in Norton and Pearson 1976, 153–54). Punnett, on the other hand, admitted that he could not follow the mathematics of Fisher's paper but did not question its correctness and remarked,

> I do not in any way wish to suggest that the mathematics were not all they should be as I have not the least doubt that the author is perfectly competent on this head. And as a contribution to biometry it may have real value—but I am not qualified to judge it from that point of view. However, whatever its value from the standpoint of statistics & population I do not feel that this kind of work affects us biologists much at present. It is too much of the order of problem that deals with weightless elephants upon frictionless surfaces, where at the same time we are largely ignorant of the other properties of the said elephants and surfaces. (qtd. in Norton and Pearson 1976, 155)

Although the referees' reports did not explicitly recommend rejection, Fisher saw that he was caught between Scylla and Charybdis. He withdrew the paper and it was later published as Fisher (1918).

27. The same charge was made later by DiTrocchio (1991). This has been a point of contention. Fisher, and others, have concluded that Mendel did the experiments as described.

28. Here, Fisher is referencing his previous work, *The Genetical Theory of Natural Selection* (Fisher 1930).

29. Some authors have questioned this. See, for example Olby (1979).

30. For randomly distributed data, the average χ^2 per degree of freedom is 1, giving an expected χ^2 of 26 for 26 degrees of freedom, rather than the 15.3224 calculated.

31. It is interesting to note that Fisher gives the probability of a worse fit to the eight experiments on the 2:1 ratio for monohybrids as 0.74, the lowest for any set of Mendel's experiments. See table 1.14.

32. There is an earlier reference to Fisher's paper in the 1963 biographical memoir written by Frank Yates and Kenneth Mather (1963). In discussing Fisher's paper they remark, "His study of Mendel's experiments (1936) was a delightful example of statistical analysis applied to the better understanding of an important chapter in the history of science." There is no mention of any controversy. Nor is any earlier controversy mentioned in the biography of Fisher written by his daughter Joan Fisher Box (1978).

33. One might also argue for 1966 as the centenary. Mendel's talks were given in 1865 and the journal issue in which his paper was published is dated 1865, but the journal did not appear until 1866.

34. This assumes that Mendel did not already have a good idea of his system before he began his experiments. This is a controversial issue. Fisher, and others, disagree, and

believe that Mendel did have a theory either before he began his experiments, or shortly thereafter. Other scholars agree with Zirkle.

35. The result is only slightly more than one standard deviation from the expected result, a not unlikely result.

36. This is true only for the monohybrid experiments, not for the trifactorial experiment.

37. Thoday is referring to Evans and Philip (1964). Although not explicitly applicable to peas, the experiments by Evans and Philip raised the possibility of such an explanation of reduced variance.

38. Teddy Seidenfeld would later independently suggest the same model (see chapter 6, this volume).

39. Campbell (1976) named the same three assistants.

40. Scholars disagree as to whether Mendel's results should be compared to Fisher's corrected predictions or to Mendel's expectations. As we see, the conclusion doesn't change significantly.

41. This was the value van der Waerden calculated for Mendel's Experiment 5.

42. Van der Waerden also calculated that the probability of Mendel obtaining seven pairs varieties of plants which differed in only one character was 0.97: "The difficulty raised by Bateson does not exist. A simple calculation of probabilities shows that it is not at all unlikely that Mendel had, from the beginning, seven pairs of varieties, each pair differing in only one essential character" (van der Waerden 1968, 277).

43. In the case of Mendel's 3:1 ratio experiments, this would require that $P_1 > P_2$, where P_1 and P_2 are the probabilities observing the dominant trait or the recessive trait, respectively.

44. Piegorsch had translated Weiling (1966). It appeared as Paper BU-718-M, Biometrics Unit, Cornell University, 1980.

45. Piegorsch also noted that "only three experiments in 100,000 attempts would show ratios as close or closer to agreement with Mendel's ratios" (2291). He commented that the probability of getting a worse fit was 0.99997, in contrast to Fisher's result of 0.99993. He attributed this to a lack of precision in either Fisher's algorithm or in his calculating machine. This makes Mendel's results even more unlikely.

46. In one experiment, the students obtained a perfect fit to the expected Mendelian ratio. With a sufficient number of "indeterminants" this is not unexpected.

47. This was based on a fragment of Mendel's writings called the Notizblatt. Weiling (1991) dated this fragment as later than 1874. It also does not specify what type of plants were used in the experiments.

48. My arguments against Root-Bernstein's explanation of Mendel's results should not, however, be taken to imply that I believe Mendel was guilty of fraud. They are merely comments on the inadequacy of his analysis.

49. There is no mathematical justification for this statement.

50. Opinion was not unanimous. Gardner (1981) stated, "Mendel's figures are suspect for just this reason. They are too good to be true. Did the priest consciously fudge his data? Let us be charitable. Perhaps he was guilty only of 'wishful seeing' when he classified and counted his talls and dwarfs" (124). Kohn (1986), on the other hand, citing the arguments given by Pearl and Root-Bernstein concluded, "It seems to me that all insinuations about Mendel's possible unethical behavior should be discounted" (43).

51. Using modern computational techniques, Matthew Stephens (1994) confirmed Edwards's conclusions concerning both Mendel's data and excellent randomizing by peas shown in the data of Darbishire and by Bateson and Kilby. "There seems to be good statis-

tical grounds on which to argue that either Mendel was mistaken in his statement or that the model is incorrect" (20). He further remarked, "It is clear that Mendel's results do not fit the binomial model very well; nor do they agree well with the other two data sets considered, which suggests that it is more than just the model which is wrong" (53). Stephens also calculated that the probability of never having more than five recessives in a pod, under a binomial model is approximately 1 in 2750. This is also discussed by Seidenfeld, as detailed later in this introduction.

52. This claim is not only in disagreement with Edwards's analysis, but also with that of Seidenfeld, as discussed later in this introduction and in chapter 6.

53. This point is discussed in detail by both Seidenfeld and by Fairbanks and Rytting, chapters 6 and 7, respectively. Both disagree with Weiling's analysis. Weiling does not calculate the effect that this would have on the χ^2 obtained. In addition, as discussed later in this introduction, the recent experimental results obtained by Fairbanks argue against an urn model.

54. The question of whether Mendel should have observed linkage is also discussed in Piegorsch (1986), van der Waerden (1968), and in the references cited in those papers. Piegorsch concluded that it was quite plausible that Mendel did not observe linkage.

55. Corcos and Monaghan had repeated the error that Pilgrim had made. In fact, the probability that a χ^2 of 0.00030 for two degrees of freedom is due to chance is 0.0002.

56. Nissani cites no evidence to support this point.

57. This refers to the total χ^2 for all six experiments, not to those for the separate experiments. Recall that Fisher had reported a χ^2 probability of 0.74 for the entire set of eight experiments (see table 1.14). This is a good fit, but certainly nothing extraordinary.

58. Seidenfeld is being quite conservative here. The actual probability is 3×10^{-3}. For a probability of 0.01, only six plants out of 600 would have been misclassified, which is negligible.

59. Seidenfeld's model was proposed independently. When he told me of this model in a private discussion, I informed him of the earlier work and gave him the references.

60. Recall that Stephens (1994, 53n52) had calculated that such a result had a probability of 1/2750 on a binomial model.

61. Recent results, discussed briefly below and in chapter 7, indicate no support for the urn model for peas and that the pea is an excellent randomizer.

62. It is amusing to speculate about how Mendel's exclusion of data might have affected the goodness of fit. It is generally agreed that Mendel excluded at least as much data as he presented. Let us assume that the amount of excluded data was equal to the amount of data that Mendel presented and that it was normally distributed, i.e., that it had a total χ^2 of 84 for 84 degrees of freedom. Adding this set of data to Mendel's published data, which had a χ^2 of 41.6056 (Fisher's value), we get a total χ^2 of 125.6056 for 168 degrees of freedom, which gives a probability of a worse fit of 0.9938. This is still extremely good, and it does nothing to solve the problem of the bulge in the probability distribution. If one assumes, quite plausibly, that the excluded data had a somewhat worse than normal distribution, then the probability goes down further. Suppose we wish to add an equal amount of excluded data such that the total χ^2 would be 168 for 168 degrees of freedom, a reasonable result. This would mean that the excluded data had a χ^2 of 126.39 for 84 degrees of freedom, which has a probability of 0.0018 of arising by chance, a very unlikely result.

REFERENCES

Bateson, W. 1902. *Mendel's Principles of Heredity: A Defence*. Cambridge: Cambridge University Press.

———. 1909. *Mendel's Principles of Heredity*. Cambridge: Cambridge University Press.
Bateson, W., and H. Kilby. 1905. "Experimental Studies in the Physiology of Heredity: Peas." *Royal Society Reports to the Evolution Committee* 2:55–80.
Beadle, G. W. 1967. "Mendelism, 1965." In *Heritage from Mendel*, ed. R. A. Brink, 335–50. Madison: University of Wisconsin Press.
Bennett, J. H., ed. 1965. *Experiments in Plant Hybridisation by Gregor Mendel, with Commentary and Assessment by Sir Ronald A. Fisher*. Edinburgh: Oliver and Boyd.
———, ed. 1983. *Natural Selection, Heredity, and Eugenics: Including Selected Correspondence of R. A. Fisher with Leonard Darwin and Others*. Oxford: Clarendon Press.
Box, Joan Fisher. 1978. *R. A. Fisher: The Life of a Scientist*. Wiley Series in Probability and Mathematical Statistics. New York: John Wiley and Sons.
Broad, W., and N. Wade. 1982. *Betrayers of the Truth*. New York: Simon and Schuster.
Campbell, M. 1976. "Explanations of Mendel's Results." *Centaurus* 20:159–74.
Corcos, A., and F. Monaghan. 1986. "Correction: Chi-Square and Mendel's Experiments." *Journal of Heredity* 77:283.
———. 1993. *Gregor Mendel's Experiments on Plant Hybrids: A Guided Study*. New Brunswick: Rutgers University Press.
Darbishire, A. D. 1908. "On the Results of Crossing Round with Wrinkled Peas, with Especial Reference to Their Starch-grains." *Journal of the Royal Society of London B, Proceedings* 80:122–35.
———. 1909. "An Experimental Estimation of the Theory of Ancestral Contributions in Heredity." *Journal of the Royal Society of London B, Proceedings* 81:61–79.
De Beer, G. 1964. "Mendel, Darwin, and Fisher." *Notes and Records of the Royal Society* 19:192–225.
DeGusta, D. 2003a. "More Digging in Mendel's Garden." Crotchety Comments, Kenneth M. Weiss's Lab on the Web, Department of Anthropology, College of Liberal Art, Penn State University, http://www.anthro.psu.edu/weiss_lab/DeGusta_FULL.doc/, 1–14.
———. 2003b. "More Digging in Mendel's Garden." *Evolutionary Anthropology* 12:1.
DiTrocchio, F. 1991. "Mendel's Experiments: A Reinterpretation." *Journal of the History of Biology* 24:485–519.
Dobzhansky, T. 1967. "Looking Back at Mendel's Discovery." *Science* 156:1588–89.
Dunn, L. C. 1965. "Mendel, His Work and His Place in History." *Proceedings of the American Philosophical Society* 109:189–98.
Edwards, A. W. F. 1972. *Likelihood*. Cambridge: Cambridge University Press.
———. 1986a. "Are Mendel's Results Really Too Close?" *Biological Reviews* 61:295–312.
———. 1986b. "More on the Too-Good-to-Be-True Paradox and Gregor Mendel." *Journal of Heredity* 77:138.
Evans, D. A., and U. Philip. 1964. "On the Distribution of Mendelian Ratios." *Biometrics* 20:794–817.
Fairbanks, D. J., and B. Rytting. 2001. "Mendelian Controversies: A Botanical and Historical Review." *American Journal of Botany* 88:737–52.
Farrall, L. A. 1975. "Controversy and Conflict in Science: A Case Study—The English Biometric School and Mendel's Laws." *Social Studies of Science* 5:269–301.
Fisher, R. A. 1918. "On the Correlation between Relatives on the Supposition of Mendelian Inheritance." *Transactions of the Royal Society of Edinburgh* 52:399–433.
———. 1924. "The Biometrical Study of Heredity." *Eugenics Review* 16:189–210.
———. 1930. *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press.
———. 1935. *The Design of Experiments*. Edinburgh: Oliver and Boyd.
———. 1936. "Has Mendel's Work Been Rediscovered?" *Annals of Science* 1:115–37.

———. 1965a. "Introductory Notes on Mendel's Paper." In *Experiments in Plant Hybridisation by Gregor Mendel, with Commentary and Assessment by Sir Ronald Fisher*, ed. J. H. Bennett, 1–6. Edinburgh: Oliver and Boyd.

———. 1965b. "Marginal Comments on Mendel's Paper." In *Experiments in Plant Hybridisation by Gregor Mendel, with Commentary and Assessment by Sir Ronald Fisher*, ed. J. H. Bennett, 52–58. Edinburgh: Oliver and Boyd.

Focke, W. O. 1881. *Die Pflanzen-mischlinge ein Beitrag zur Biologie der Gewächse*. Berlin.

Froggatt, P., and N. C. Nevin. 1971a. "Galton's 'Law of Ancestral Heredity': Its Influence on the Early Development of Human Genetics." *History of Science* 10:1–27.

———. 1971b. "The 'Law of Ancestral Heredity' and the Mendelian Ancestrian Controversy in England, 1899–1906." *Journal of Medical Genetics* 8:1–36.

Gardner, M. 1981. *Science: Good, Bad, and Bogus*. Buffalo: Prometheus Books.

Hartl, D. L., and D. J. Fairbanks. 2007. "Mud Sticks: On the Alleged Falsification of Mendel's Data." *Genetics* 175:975–79.

Hartl, D. L., and V. Orel. 1992. "What Did Gregor Mendel Think He Discovered?" *Genetics* 131:245–53.

Judson, H. F. 2004. *The Great Betrayal: Fraud in Science*. New York: Harcourt.

Kevles, D. J. 1980. "Genetics in the United States and Great Britain, 1890–1930: A Review with Speculations." *Isis* 71:441–55.

Kohn, A. 1986. *False Prophets*. Oxford: Basil Blackwell.

Lamprecht, H. 1968. *Die Grundlagen der Mendelschen Gesetze*. Berlin: Paul Parey.

Leonard, T. 1977. "A Bayesian Approach to Some Multinomial Estimation and Pretesting." *Journal of the American Statistical Association* 72:869–74.

Lindley, D. V., and W. F. Scott. 1984. *New Cambridge Elementary Statistical Tables*. Cambridge: Cambridge University Press.

Liu, Y. 2005. "Darwin and Mendel: Who Was the Pioneer of Genetics?" *Rivista di Biologia/Biology Forum* 98:305–22.

Magnello, E. 2004. "The Reception of Mendelism by the Biometricians and the Early Mendelians (1899–1909)." In *A Century of Mendelism in Human Genetics*, ed. M. Keynes, A. W. F. Edwards, and R. Peel, 19–32. Boca Raton: CRC Press.

Mendel, G. 1870. "Uber einige aus kunstlicher Befruchtung gewonnen Hieracium-Bastarde." *Verhandlungen des naturforschenden Vereines in Brunn* 8:26–31.

Monaghan, F., and A. Corcos. 1985. "Chi-Square and Mendel's Experiments: Where's the Bias?" *Journal of Heredity* 76:307–9.

Montgomerie, B., and T. Birkhead. 2005. "A Beginner's Guide to Scientific Misconduct." *International Society for Behavioral Ecology* 17:16–24.

Morrison, M. 2002. "Modelling Populations: Pearson and Fisher on Mendelism and Biometry." *British Journal for the Philosophy of Science* 53:39–68.

Nissani, M. 1994. "Psychological, Historical, and Ethical Reflections on the Mendelian Paradox." *Perspectives in Biology and Medicine* 37:182–96.

Norton, B., and E. S. Pearson. 1976. "A Note on the Background to, and Refereeing of, R. A. Fisher's 1918 Paper, 'On the Correlation between Relatives on the Superposition of Mendelian Inheritance.'" *Notes and Records of the Royal Society* 31:151–62.

Novitski, C. E. 1995. "Another Look at Some of Mendel's Results." *Journal of Heredity* 86:62–66.

———. 2004. "Revision of Fisher's Analysis of Mendel's Garden Pea Experiments." *Genetics* 166:1139–40.

Novitski, E. 2004. "On Fisher's Criticism of Mendel's Results With the Garden Pea." *Genetics* 166:1133–36.

O'Kelly, M. 2004. "Using Statistical Techniques to Detect Fraud: A Test Case." *Pharmaceutical Statistics* 3: 237–46.
Olby, R. C. 1966. *Origins of Mendelism*. New York: Schocken Books.
———. 1979. "Mendel No Mendelian?" *History of Science* 17:55–72.
———. 1985. *Origins of Mendelism*. Chicago: University of Chicago Press.
Orel, V. 1968. "Will the Story on 'Too Good' Results of Mendel's Data Continue?" *BioScience* 18:776–78.
———. 1996. *Gregor Mendel: The First Geneticist*. Oxford: Oxford University Press.
Orel, V., and D. L. Hartl. 1994. "Controversies in the Interpretation of Mendel's Discovery." *History and Philosophy of the Life Sciences* 16:423–64.
Pearl, R. 1940. *Introduction to Medical Biometry and Statistics*. Philadelphia: W. B. Saunders.
Pearson, K. 1900. "On the Criterion that a Given System of Deviations from the Probable in the Case of a Correlated System of Variables Is Such that It Can Be Reasonably Supposed to Have Arisen from Random Sampling." *Philosophical Magazine* 50:157–75.
Piegorsch, W. 1983. "The Question of Fit in the Gregor Mendel Controversy." *Communications in Statistics: Theory and Methods* 12:2289–304.
———. 1986. "The Gregor Mendel Controversy: Early Issues of Goodness-of-Fit and Recent Issues of Genetic Linkage." *History of Science* 24:173–82.
———. 1990. "Fisher's Contributions to Genetics and Heredity, with Special Emphasis on the Gregor Mendel Controversy." *Biometrics* 46:915–24.
Pilgrim, I. 1984. "The Too-Good-to-Be-True Paradox and Gregor Mendel." *Journal of Heredity* 75:501–2.
———. 1986. "Rebuttal." *Journal of Heredity* 77:138.
Provine, W. 1971. *The Origins of Theoretical Population Genetics*. Chicago: University of Chicago Press.
R. A. Fisher Digital Archive. The University of Adelaide Digital Library. University of Adelaide, North Terrace, Adelaide, Australia. http://digital.library.adelaide.edu.au/coll/special/fisher/.
Robertson, T. 1978. "Testing For and Against an Order Restriction on Multinomial Parameters." *Journal of the American Statistical Association* 73:197–202.
Root-Bernstein, R. S. 1983. "Mendel and Methodology." *History of Science* 21:275–95.
Sapp, J. 1990. "The Nine Lives of Gregor Mendel." In *Experimental Inquiries*, ed. H. E. LeGrand, 137–66. Dordrecht: Kluwer Academic Publishers.
Seidenfeld, T. 1998. "P's in a Pod: Some Recipes for Cooking Mendel's Data." PhilSci Archive, Department of History and Philosophy of Science and Department of Philosophy, University of Pittsburgh, http://philsci-archive.pitt.edu/view/subjects/confirmation-induction.html.
Stephens, M. "The Results of Gregor Mendel: An Analysis and Comparison with the Results of Other Researchers." Diploma in Mathematical Statistics Thesis, University of Cambridge.
Stern, C., and E. R. Sherwood, eds. 1966. *The Origin of Genetics: A Mendel Source Book*. San Francisco: W. H. Freeman and Co.
Sturtevant, A. H. 1965. *A History of Genetics*. New York: Harper and Row.
Thoday, J. M. 1966. "Mendel's Work as an Introduction to Genetics." *Advancement of Science* 23:120–24.
Van der Waerden, B. L. 1968. "Mendel's Experiments." *Centaurus* 12:275–88.
Weiling, F. 1965. "Die Mendelschen Ebversuche in biometrischer Sicht/Zum 100/Jahrestag des ersten Mendelschen Vortrages vor dem Naturforschenden Verein in Brunn am 8.2.1865." *Biometrische Zeitschrift* 7:230–62.

———. 1966. "Hat J. G. Mendel bei seinen Versuchen 'zu genau' gearbeitet? Der Chi-2 Test und seien Bedeutung fur die Beurteilung genetischer Spaltungsverhaltnisse." *Der Zuchter* 36:359–65.
———. 1971. "Mendel's 'Too Good' Data in Pisum-Experiments." *Folia Mendeliana* 6:75–77.
———. 1986. "What about R. A. Fisher's Statement of the 'Too Good' Data of J. G. Mendel's Pisum Paper?" *Journal of Heredity* 77:281–83.
———. 1989. "Which Points Are Incorrect in R. A. Fisher's Statistical Conclusion? Mendel's Experimental Data Agree Too Closely with His Expectations." *Angewandte Botanik* 63:129–43.
———. 1991. "Historical Study: Johann Gregor Mendel 1822–1884." *American Journal of Medical Genetics* 40:1–25.
Weiss, K. 2002. "Goings on in Mendel's Garden." *Evolutionary Anthropology* 11:40–44.
Weldon, W. F. R. 1902. "Mendel's Laws of Alternative Inheritance in Peas." *Biometrika* 1:228–54.
Wright, S. 1966. "Mendel's Ratios." In Stern and Sherwood 1966, 173–75.
Yates, Frank, and Kenneth Mather. 1963. "Ronald Aylmer Fisher." *Biographical Memoirs of Fellows of the Royal Society of London* 9:91–120.
Zirkle, C. 1964. "Some Oddities in the Delayed Discovery of Mendelism." *Journal of Heredity* 55 (1964): 65–72.