Arch. Hist. Exact Sci. 58 (2004) 323–379
Digital Object Identifier (DOI) 10.1007/s00407-003-0076-z

Doing Much About Nothing

ALLAN FRANKLIN

Communicated by J.Z. BUCHWALD

One of the interesting problems concerning experimental results is what happens when an experiment gives a null result, when the phenomenon expected is not observed. Is it because the experimental apparatus and the associated analysis procedures cannot detect or measure the phenomenon in question or is it because the phenomenon is not present? This is a real problem in the practice of physics. In the Michelson-Morley experiment, one of the most famous experiments in modern physics, the experimenters expected to detect a fringeshift caused by the motion of the earth relative to the ether. They found no such fringeshift. Was it because the apparatus was faulty or because the earth's velocity relative to the ether was zero?[1] In this paper I will discuss the Michelson-Morley experiment as well as other cases from the history of modern physics in which null results played a crucial role. I will also examine the strategies and arguments used by physicists to demonstrate that their experiment would have detected the phenomenon had it been present.

Perhaps the most typical strategy is the use of a surrogate signal. As we will see, the production of an adequate surrogate can involve considerable ingenuity by physicists. Some commentators on science have questioned whether the adequacy of such surrogate signals is examined in sufficient detail.[2] I will discuss the arguments offered by physicists for the adequacy of such signals. I will also discuss an episode, that of the report of a possible magnetic monopole, in which the creation and use of a surrogate signal helped to cast doubt on an observation. Another strategy is the use of a computer simulation of a surrogate signal.

Often, because of previous experience with a particular type of experimental apparatus, there is, in fact, no real question as to whether the experimental can detect and measure the phenomenon.

---

[1] Other explanations were later offered for this null result. They included the ether drag hypothesis, that the earth dragged a layer of the ether along with it; the Lorentz-Fitzgerald contraction, that an object shrank in the direction of its motion relative to the ether; and the ballistic theory, that the velocity of light was constant relative to the source of the light. Eventually, all of these alternatives were rejected on the basis of experimental evidence.

[2] See for example (Collins 1985). I have argued elsewhere that Collins is wrong (Franklin 1997).

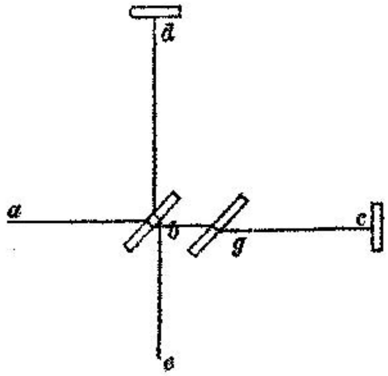

Fig. 1. Schematic view of the Michelson-Morley experiment. From Michelson (1881)

A. The Michelson-Morley experiment

The Michelson-Morley experiment (1887) is perhaps the most famous experiment in modern physics. It was designed to measure the velocity of the earth relative to the luminferous Ether, the presumed medium through which light and electromagnetic radiation were propagated. The title of the Michelson-Morley paper was, in fact, "On the Relative Motion of the Earth and the Luminiferous Ether." The experiment, discussed in detail below, was designed to measure the change in an interference pattern produced by two light beams, one traveling parallel to the velocity of the earth relative to the ether, and the other perpendicular to that velocity, when the experimental apparatus was rotated through 90°. The predicted effect, assuming the velocity of the earth relative to the ether was the earth's orbital velocity, was four tenths of a fringe. Michelson and Morley observed no such change. "The actual displacement was certainly less than a twentieth of this, and probably less than the fortieth part (Michelson and Morley 1887, p. 341)." This essentially null result led to the conclusion that "the relative velocity of the earth and the ether is probably less than one sixth of the earth's orbital velocity and certainly less than one fourth (p. 341)."

The design of the experimental apparatus is shown schematically in Fig. 1. Light from the source at a strikes a half-silvered mirror at b, with half the light being transmitted to mirror e, and the other half reflected to mirror d. The glass plate g "was interposed in the path of the ray bc to compensate for the thickness of the glass b (Michelson 1881, p. 122)." The two reflected beams again strike b, and some of the light from d is transmitted and some from c reflected toward the observer e, who will observe an interference pattern. If the earth is moving relative to the ether this pattern should change when the apparatus was rotated through 90°.

The derivation of the change in the interference pattern is as follows (I follow the derivation given in (Michelson and Morley 1887). I use c for the velocity of light, rather than Michelson's V.) Assume that bc is parallel to the earth's velocity relative to the ether and bd is perpendicular. Let $D = bc$ or bd, c is the velocity of light, v is the velocity of the earth relative to the ether. The time, T_1, for light to travel back and forth along bc is

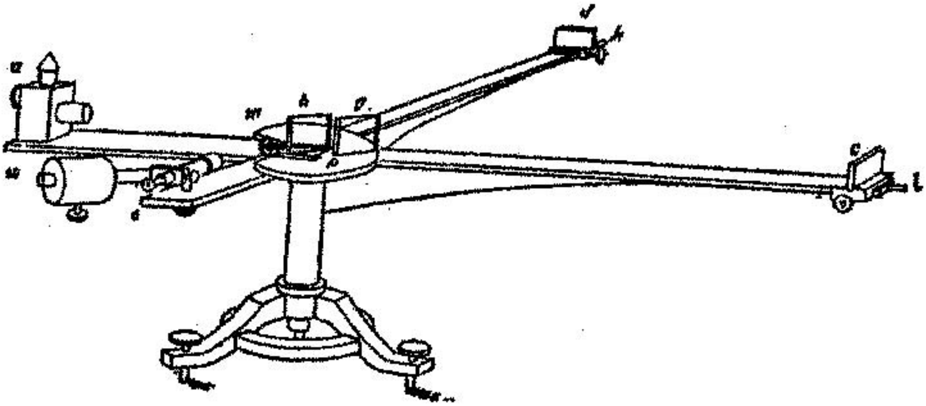

Fig. 2. Michelson's 1881 apparatus. From Michelson (1881)

$$T_1 = D/(c - v) + D/(c + v) = 2Dc/(c^2 - v^2).$$

The distance d_1 traveled by the light in this time is

$$d_1 = 2Dc^2/(c^2 - v^2) \approx 2D(1 + v^2/c^2) \text{ to first order in } v^2/c^2.$$

For the perpendicular path, remembering that the mirror is moving while the light is traveling from b to d, the distance d_2 traveled by the light is

$$d_2 = 2D\sqrt{(1 + v^2/c^2)} \approx 2D(1 + v^2/2c^2).$$

The path difference $\Delta = d_1 - d_2 = Dv^2/c^2$.

When the apparatus is rotated through 90° the difference $\Delta' = d_2 - d_1 = Dv^2/c^2$. Thus, one expects a total fringe shift of $2Dv^2/c^2$.

1. Michelson's 1881 experiment

I will begin with a discussion of Michelson's 1881 experiment, the precursor to the Michelson-Morley experiment of 1887. The problem Michelson faced was to build an apparatus sufficiently sensitive to detect the small effect expected, yet robust enough to be immune to background effects caused by vibration or by the rotation of the apparatus, which might mimic or mask the predicted effect. Michelson's initial apparatus (Fig. 2) had $D = 1.2$ m "or in wavelengths of yellow light 2 000 000. Then in terms of the same unit $2Dv^2/c^2 = 4/100$ (Michelson, 1881, p. 121)".[3]

---

[3] Michelson actually expected to observe a shift of 0.08 fringes ($v/c \approx 10^{-4}$). As pointed out by Lorentz and by Potier, Michelson had made an error in his calculation. He had neglected the motion of the mirror d. This reduces the predicted effect by a factor of two. Michelson was also not consistent in stating the size of his interferometer. In one place he states that $D = 1.2$ m or 2×10^6 wavelengths of yellow light. Elsewhere he used $D \approx 1$ m or 1.7×10^6 wavelengths.

In this 1881 experiment Michelson had great difficulty in performing the experiment. The background effects were substantial.

> The apparatus as above described was constructed by Schmidt and Haensch of Berlin. It was placed on a stone pier in the Physical Institute, Berlin. The first observation showed, however, that owing to the extreme sensitiveness of the instrument to vibrations, the work could not be carried on during the day. The experiment was next tried at night. When the mirrors were placed half-way on the arms the fringes were visible, but their position could not be measured till after twelve o'clock, and then only at intervals. When the mirrors were moved out to the end of the arms, the fringes were only occasionally visible.
>
> It thus appeared that the experiments could not be performed in Berlin, and the apparatus was accordingly removed to the *Astrophyicalisches Observatorium* in Potsdam. Even here the ordinary stone piers did not suffice, and the apparatus was again transferred, this time to a cellar whose circular walls formed the foundation for the pier of the equatorial.
>
> Here, the fringes under ordinary circumstances were sufficiently quiet to measure, but so extraordinarily sensitive was the instrument that the stamping of the pavement about 100 meters from the observatory, made the fringes disappear entirely (p. 124).

There were further difficulties, particularly the effects of temperature and of the rotation of the instrument.

> The principal difficulty which was to be feared in making these experiments, was that arising from changes of temperature of the two arms of the instrument. These being of brass whose coefficient of expansion is 0.00019 and having a length of about 1000 mm. or 1 700 000 wave-lengths, if one arm should have a temperature only one-hundredth of a degree higher than the other, the fringes would thereby experience a displacement three times as great as that which would result from the rotation. On the other hand, since the changes of temperature are independent of the direction of the arms, if these changes were not too great their effect could be eliminated. [Michelson covered the arms *bc* and *bd* with long paper boxes to guard against such changes in temperature].
>
> It was found, however, that the displacement on account of bending of the arms during rotation was so considerable that the instrument had to be returned to the maker, with instructions to make it revolve as easily as possible. It will be seen from the tables, that notwithstanding this precaution a large displacement was observed in particular direction. That this was due entirely to the support was proved by turning the latter through 90°, when the direction in which the displacement appeared was also changed 90° .
>
> On account of the sensitiveness of the instrument to vibration, the micrometer of the observing telescope could not be employed, and a scale ruled on glass was substituted. The distance between the fringes covered three scale divisions, and the position of the center of the dark fringe was estimated to fourths of a division, so that the separate estimates were correct to within 1/12 (p. 125).

Michelson persevered and found that the average displacement of the fringes when the beams were in the N-S and E-W directions and then rotated was 0.022 in fringe units. When the beams were initially in the NE-SW directions the shift was 0.034. "The former is too small to be considered as showing a displacement due to the simple change in direction, and the latter should have been zero (p. 127)."

Michelson's data actually showed a substantial linear drift. The measured result after a rotation of 360° did not agree with that obtained initially, as it should have. Michelson fitted a straight line to his data points and measured the deviation from that line. His

Fig. 3. Michelson's 1881 experimental result. "The dotted curve is drawn of the supposition that the displacement to be expected is one-tenth of the distance between the fringes,... From Michelson (1881)

new results were -0.004 fringe units for the N-S, E-W directions and -0.015 for the NE-SW directions. According to Michelson these were "simply error of experiment." "The results obtained are, however, more strikingly shown by constructing the actual curve [Fig. 3] that should have been found if the theory had been correct (p. 128)."

Michelson concluded, "The interpretation of these results is that there is no displacement of the interference bands. The result of the hypothesis if a stationary ether is thus shown to be incorrect, and the necessary conclusion follows that the hypothesis is erroneous (p. 128)."

Michelson's measurement and result attracted little attention. (For details see (Swenson 1972, pp. 73–74)). No questions seem to have been raised concerning the ability of his apparatus to measure the predicted effect.

2. The Michelson-Morley experiment (1887)

Michelson was encouraged to continue his work by both Lorentz and by Rayleigh. In a letter to Rayleigh, Michelson remarked, "I have never been fully satisfied with my Potsdam experiment, even taking into account the correction which H.A. Lorentz points out.... I have repeatedly tried to interest my scientific friends in this experiment without avail, and the reason for my never publishing the correction was (I am ashamed to confess it) that I was discouraged at the slight attention the work received and did not think it worthwhile. (Michelson to Rayleigh, March 6, 1887, full text in (Shankland 1964, p. 29))."

Michelson and Morley (1887) admitted the error that Lorentz had noted, the omission of the motion of the apparatus relative to the ether, in their 1887 report of their experiment and pointed out that the predicted effect was reduced by a factor of two and that Michelson's 1881 conclusion might well be questioned.

> In deducing the formula for the quantity to be measured, the effect of the motion of the earth through the ether on the path of the ray at right angles to this motion was overlooked. The discussion of this oversight and of the entire experiment forms the subject of a very searching analysis by H.A. Lorentz, who finds that this effect can by no means be disregarded. *In consequence the quantity to be measured had in fact but one-half the value supposed, and as it was already barely beyond the limits of errors of experiment, the conclusion drawn from the result might well be questioned* (Michelson and Morley 1887, pp. 334–335, emphasis added).

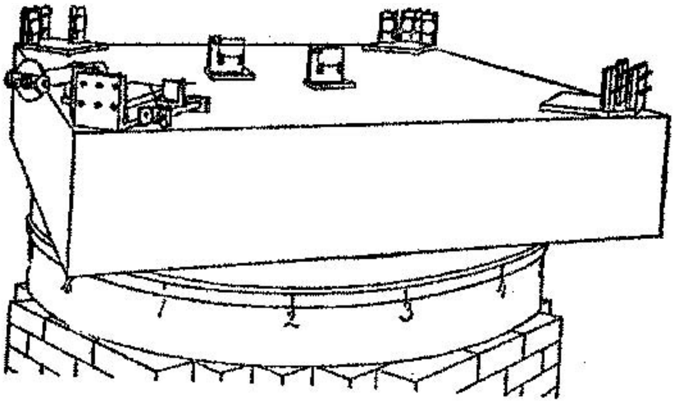

Fig. 4. The 1887 Michelson-Morley apparatus. From Michelson and Morley (1887)

They also noted the problems of vibration and rotation.

> In the first experiment [Michelson's 1881 experiment] one of the principle difficulties encountered was that of revolving the apparatus without producing distortion; another was the extreme sensitiveness to vibration. This was so great that it was impossible to see the interference fringes except at brief intervals when working in the city, even at two o'clock in the morning. Finally, as before remarked, the quantity to be observed , namely, a displacement of something less than a twentieth of the distance between the interference fringes may have been too small to be detected when masked by experimental errors.
>
> The first named difficulties were entirely overcome by mounting the apparatus on a massive stone floating on mercury; and the second by increasing, by repeated reflection, the path of the light to ten times its former value. (pp. 336–337).

The new Michelson-Morley apparatus is shown in Fig. 4. They adjusted the mirrors so that two different light paths were approximately equal and used a sodium light source to make the fringes appear as clear as possible. A white light source was then substituted for the sodium light and the path length adjusted "till the colored interference fringes reappeared in white light. These were now given a convenient width and position (p. 339)."[4] Michelson and Morley were then in a position to take data.

> The observations were conducted as follows: Around the cast-iron trough were sixteen equidistant marks [Fig. 4]. The apparatus was revolved very slowly (one turn in six minutes) and after a few minutes the cross wire of the micrometer was set on the clearest of the interference fringes at the instant of passing one of the marks. The motion was so slow that this could be done readily and accurately. The reading of the screw-head on the micrometer was noted, and a very slight and gradual impulse was given to keep up the motion of the stone; on passing the second mark the same process was repeated, and this was continued until the apparatus had completed six revolutions (p. 339).

---

[4] As Miller later stated, "White light fringes were chosen for the observations because they consist of a small group of fringes having a central, sharply defined black fringe which forms a permanent zero reference mark for all readings (Miller 1933, p. 210)."

Once again, although it is not stated in the paper, there was a significant linear drift in the data (Table 1 gives the uncorrected data. The figures in the table are the averages for six revolutions. For details see (Handschy 1982).) Michelson and Morley fitted a straight line to the data and took the differences between the data and the fitted line as their residuals (Fig. 5). There is no obvious displacement of the fringes. (Note that in the figure the dotted curve is one eighth of the predicted signal).

> It seems fair to conclude from the figures that if there is any displacement due to the relative motion of the earth and the luminiferous ether, this cannot be greater than 0.01 of the distance between the fringes.
>
> Considering the motion of the earth in its orbit only, this displacement should be $2Dv^2/c^2 = 2D \times 10^{-8}$. The distance D was about eleven meters, or 2×10^7 wavelengths of yellow light; hence the displacement was expected to be 0.4 fringe. The actual displacement was certainly less than the twentieth part of this, and probably less than the fortieth part. But since the displacement is proportional to the square of the velocity, the relative velocity of the earth and the ether is probably less than one sixth of the earth's orbital velocity, and certainly less than one fourth (pp. 340–341).

3. Discussion

Despite the improvements in the Michelson-Morley experiment and the increased sensitivity of their result, that result attracted little attention. (For details see (Swenson 1972, Chap. 5) and (Buchwald 1988)). For most physicists, Hertz's classic experiment in 1886 demonstrating the existence of electromagnetic waves, and by implication, the ether, was far more important. The Michelson-Morley result was considered to be a minor problem. In 1889, two years after the publication of the Michelson-Morley result, Rowland remarked, "... the luminferous ether is, to-day, a much more important factor in science than the air we breathe (quoted in (Buchwald 1988, p. 55))."

The Michelson-Morley result was taken seriously by both Fitzgerald and Lorentz who independently proposed a contraction hypothesis, that an object shrank in the direction of its motion relative to the ether by exactly the right factor so that Michelson and Morley would observe a null result. Lodge, an ardent supporter of the ether described Michelson's "remarkable experiment" and suggested that "This experiment might have to be explained away (Lodge 1893, p. 753)."

There was some later criticism of the Michelson-Morley result. Interestingly, it was suggested that Michelson and Morley had, in fact, observed a positive result. Hicks (1902) presented a very detailed technical analysis of the Michelson-Morley experiment and also reanalyzed their data. He also corrected for the linear drift of that data, which he attributed to a temperature effect. Based on his reanalysis he concluded

> The preceding attempt to get rid of the temperature effect is not proposed as one which gives an accurate result. The object is to show that the observations of Michelson and Morley do give an affirmative answer to the question "Is there a drift of the aether past the earth?" The argument is sufficient to show that the experiment should be repeated with extreme care to eliminate temperature errors, and if possible *in vacuo* (Hicks 1902, p. 38)

Table 1
Noon Observations

	16	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
July 8	44.7	44.0	43.5	39.7	35.2	34.7	34.3	32.5	28.2	26.2	23.8	23.2	20.3	18.7	17.5	16.8	13.7
July 9	57.4	57.3	58.2	59.2	58.7	60.2	60.8	62.0	61.5	63.3	65.8	67.3	69.7	70.7	73.0	70.2	72.2
July 11	27.3	23.5	22.0	19.3	19.2	19.3	18.7	18.8	16.2	14.3	13.3	12.8	13.3	12.3	10.2	7.3	6.5
Mean	43.1	41.6	41.2	39.4	37.7	38.1	37.9	37.8	35.3	34.6	34.3	34.4	34.4	33.9	33.6	31.4	30.8
Mean in Wavelengths	0.862	0.832	0.824	0.788	0.754	0.762	0.758	0.756	0.706	0.692	0.686	0.688	0.688	0.678	0.672	0.626	0.616
	0.706	0.692	0.686	0.688	0.688	0.678	0.672	0.626	0.616								
Final Mean	0.784	0.762	0.755	0.738	0.721	0.720	0.715	0.692	0.661								
Evening Observations																	
July 8	61.2	63.3	63.3	68.2	67.7	69.3	70.3	69.8	69.0	71.3	71.3	70.5	71.2	71.2	70.5	72.5	75.7
July 9	26.0	26.0	28.2	29.2	31.5	32.0	31.3	31.7	33.0	35.8	36.5	37.3	38.8	41.0	42.7	43.7	44.0
July 12	66.8	66.5	66.0	64.3	62.2	61.0	61.3	59.7	58.2	55.7	53.7	54.7	55.0	58.2	58.5	57.0	56.0
Mean	51.3	51.9	52.5	53.9	53.8	54.1	54.3	53.7	53.4	54.3	53.8	54.2	55.0	56.8	57.2	57.7	58.6
Mean in Wavelengths (Minus one)	0.026	0.038	0.050	0.078	0.076	0.082	0.086	0.074	0.068	0.086	0.076	0.084	0.100	0.136	0.144	0.154	0.172
	0.068	0.086	0.076	0.084	0.100	0.136	0.144	0.154	0.172								
Final Mean (Minus one)	0.047	0.062	0.063	0.081	0.088	0.109	0.115	0.114	0.120								

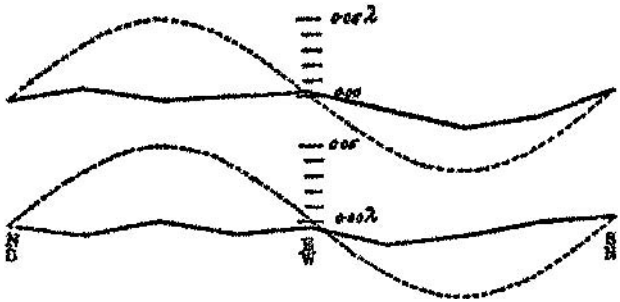

Fig. 5. The 1887 results of Michelson and Morley. The upper curve is for the noon observations and the lower for the evening observations. "The dotted curves represent *one-eighth* of the theoretical displacements. From Michelson and Morley (1887)

Miller (1933) later performed another reanalysis of the Michelson-Morley data and concluded that Michelson and Morley had actually observed a velocity of the earth relative to the ether of approximately 8 km/s. This was in agreement with positive results reported by Miller and Morley in the early 20th century. (See (Miller 1933) for details).[5] It was also approximately equal to the upper limit that had been set by Michelson and Morley in their 1887 paper. Notice, however, that neither Hicks nor Miller questioned the correctness of the Michelson-Morley data, they questioned its analysis.

Michelson and Morley included no explicit discussion of the ability of their apparatus to measure the expected effect. There is no discussion of how accurately the position of the fringe could be measured. This was, in all probability, unnecessary. The width of the observed fringes in the 1887 experiment was approximately 50 screw divisions of the mirror micrometer adjustment. The predicted effect was 0.4 fringes, or 20 such divisions. An effect of that size would have been obvious. Michelson and Morley reported their observations to 0.1 screw divisions, and as seen in Table 1 they measured changes in fringe position of the order of three or four divisions.[6] For further indirect evidence see Fig. 6. This is a photograph taken of the fringes observed with Miller's interferometer. Because his experimental apparatus was three times larger than that of Michelson and Morley, Miller claimed his apparatus was three times as sensitive. He described his observation procedure, which was quite similar to that used by Michelson and Morley, as follows.

> The reading is determined by instantaneous visual estimation; it is quite impracticable to use any kind of scale in the field of view because the width of the fringes is subject to slight variation. That this method is sufficient is shown by the uniformly consistent and systematic periodic curves representing the observations. The numerical quantity used as the result of a "single observation" is the average of forty such readings... and it approached an accuracy of a hundredth of a fringe (Miller 1933, p. 213)."

From examination of the photograph, it seems clear that a shift of 0.4 fringes would have been clearly visible in the Michelson-Morley experiment. Assuming Miller's

---

[5] These positive results were later explained as due to a temperature effect (Shankland, McCuskey et al. 1955).

[6] This change was actually in the average of the measurements.

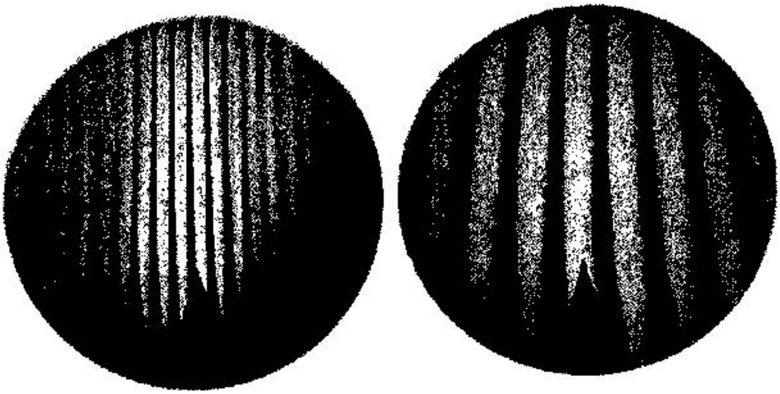

Fig. 6. Interference fringes seen in Miller's interferometer. From Miller (1933)

estimate of his sensitivity and uncertainty is correct, the Michelson-Morley uncertainty would have been approximately three-hundredths of a fringe, clearly far less than the 0.4 fringe shift expected. In this case analysis of the experiment indicates that the effect should have been observed.

B. Surrogate signals

1. The 17-keV neutrino

The use of a surrogate signal to demonstrate that an effect would have been observed by an experimental apparatus is clearly illustrated in the complex history of the 17-keV neutrino. This episode not only involves the use of a surrogate signal, but also the question of whether the proper analysis procedures needed to observe the phenomenon in question were used. In the discussion that follows I will concentrate on the strategies used to demonstrate that an effect, if present, would have been observed. (For a more complete history see (Franklin 1995)).

The 17-keV neutrino was first reported by Simpson (1985). He had searched for a heavy neutrino by looking for a kink in the energy spectrum, or in the Kurie plot,[7] at an energy equal to the maximum allowed decay energy minus the mass of the heavy neutrino, in energy units. The fractional deviation in the Kurie plot value $\Delta K/K \sim R[1 - M_2^2/(Q-E)^2]^{1/2}$, where M_2 is the mass of the heavy neutrino, R is the intensity of the second neutrino branch, Q is the total energy available for the transition, and E is the energy of the electron. Simpson's result is shown in Fig. 7. A kink is clearly seen at an energy of 1.5 keV, corresponding to a 17 keV neutrino. "In summary, the β spectrum of tritium

[7] In a normal beta-decay spectrum the quantity $K = N(E)/[f(Z,E)(E^2-1)^{1/2}E^{1/2}]^{1/2}$ is a linear function of E, the energy of the electron. A plot of that quantity as a function of E, the energy of the decay electron, is called a Kurie plot.

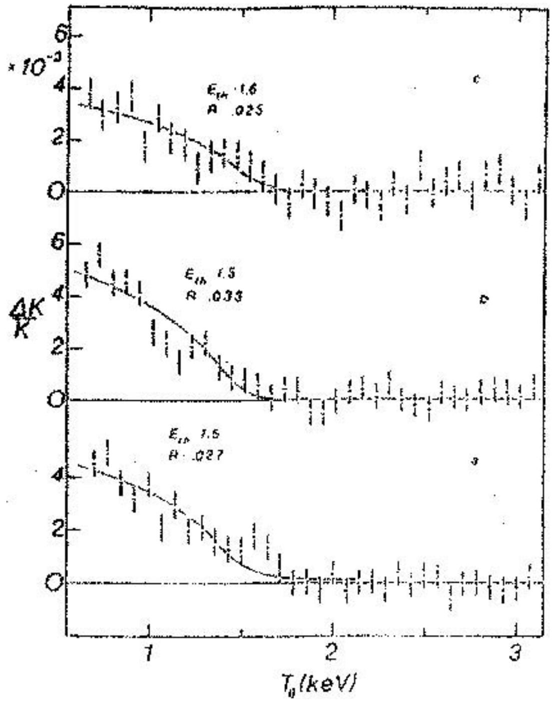

Fig. 7. The data of three runs presented as $\Delta K/K$ (the fractional change in the Kurie plot) as a function of the kinetic energy of the β particles. E_{th} is the threshold energy, the difference between the endpoint energy and the mass of the heavy neutrino. A kink is clearly seen at $E_{th} = 1.5$ keV, or at a mass of 17.1 keV. Run *a* included active pileup rejection, whereas runs *b* and *c* did not. *c* was the same as *b* except that the detector was housed in a soundproof box. No difference is apparent. From Simpson (1985)

recorded in the present experiment is consistent with the emission of a heavy neutrino of mass about 17.1 keV and a mixing probability of about 3% (Simpson 1985, p. 1893)."

Simpson's positive result for the 17-keV neutrino was published in April, 1985. By the end of the year the results of five other experimental searches for the particle had appeared in the published literature (Altzitzoglou, Calaprice et al. 1985; Apalikov, Boris et al. 1985; Datar, Baba et al. 1985; Markey and Boehm 1985; Ohi, Nakajima et al. 1985). All of them were negative. The experiments set limits of less than one percent for a 17-keV branch of the decay, in contrast to Simpson's value of three percent. The typical results of Ohi et al. are shown in Fig. 8 and should be compared to Simpson's result shown in Fig. 7. No kink of any kind is apparent.

Each of these later experiments examined the beta-decay spectrum of ^{35}S, and searched for a kink at an energy of 150 keV, 17 keV below the endpoint energy of 167 keV. Three of the experiments, those of Altzitzoglou et al., of Apalikov et al., and of Markey and Boehm, used magnetic spectrometers. Those of Datar et al. and of Ohi et al. used Si(Li) detectors, the same type used by Simpson. In the latter two cases, however, the source was not implanted in the detector, as Simpson had done, but was separated from it.

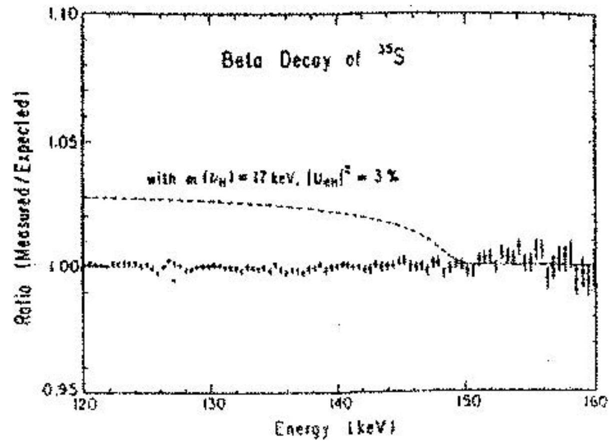

Fig. 8. The ratio of the measured ^{35}S beta-ray spectrum to the theoretical spectrum. A three percent mixing of a 17-keV neutrino should distort the spectrum as indicated by the dashed curve. From Ohi et al. (1985)

What makes this episode so intriguing is that both the initial positive claim, as well as all subsequent positive claims, were obtained in experiments using one type of apparatus, namely those incorporating a solid-state detector, whereas much of the negative evidence resulted from experiments using another type of detector, a magnetic spectrometer. One might worry, and physicists did worry, that the discordant results were due to some crucial difference between the types of apparatus or to different sources of background that might mimic or mask the signal. As Schwarzschild would later remark, "On one thing everyone seems to agree. After six years, the experimenters must begin to resolve the stubborn discrepancy between the two different styles of beta-decay experiment [solid-state detectors and magnetic spectrometers] (Schwarzschild 1991, p. 19)."

The question of whether any of the negative experiments would, in fact, have detected a heavy neutrino, had one been present, was raised at the time. There were two problems. The first was the energy range used in the analysis of the data. Simpson questioned the negative results reported in the five experiments on ^{35}S. He argued that the type of analysis used, which fitted the beta-decay spectrum over a rather large energy range, would tend to minimize the effect due to a heavy neutrino. He commented that 45 percent of the effect occurred within 2 keV of the neutrino threshold. "... in trying to fit a very large portion of the β spectrum, the danger that slowly-varying distortions of a few percent could bury a threshold effect seems to have been disregarded. One cannot emphasize too strongly how delicate is the analysis when searching for a small branch of a heavy neutrino, and how sensitive the result may be to apparently innocuous assumptions (Simpson 1986b, p. 576)." Simpson reanalyzed the results of the experiment of Ohi et al. and argued that they, in fact, showed statistically significant effects that agreed with his tritium results.[8] His reanalysis of Ohi et al.'s result is shown in Fig. 9.[9]

---

[8] Simpson also argued that the other negative results were inconclusive.
[9] See also (Simpson 1986a).

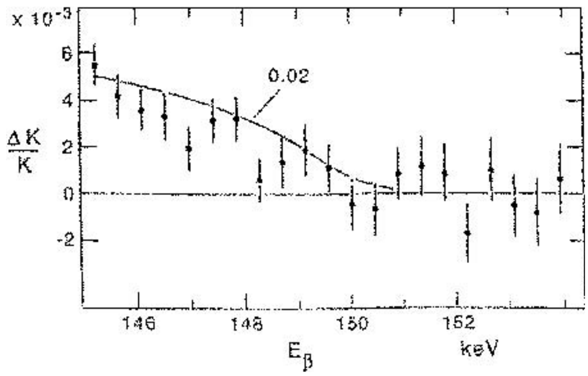

Fig. 9. $\Delta K/K$ for the ^{35}S spectra of (Ohi and others 1985) as recalculated by Simpson. From Simpson (1986a)

The reader might wonder how the same data could be used to reach such different conclusions concerning the existence of the 17-keV neutrino. The answer lies in the analysis procedures used. Morrison later showed that the positive result obtained by Simpson in his reanalysis of Ohi's data was due to his choice of energy range. "The question then is, How could the apparently negative evidence of Fig. [8] become the positive evidence of Fig. [9]? The explanation is given in Fig. [10], where a part of the spectrum near 150 keV is enlarged. Dr. Simpson only considered the region 150 keV $\pm$ 4 keV (or more exactly $+4.1$ and -4.9 keV). The procedure was to fit a straight line, shown solid, through the points in the 4 keV interval above 150 keV, and then to make this the base-line by rotating it down through about 20° to make it horizontal. This had the effect of making the points in the interval 4 keV below 150 keV appear above the extrapolated dotted line. This, however, creates some problems, as it appears that a small statistical fluctuation between 151 and 154 keV is being used: the neighboring points between 154 and 167, and below 145 keV, are being neglected although they are many standard deviations away from the fitted line. Furthermore, it is important, when analyzing any data, to make sure that the fitted curve passes through the end-point of about 167 keV, which it clearly does not (Morrison 1992, p. 600)."[10] Experimental results may be quite sensitive to the analysis procedures used.

Other experimenters took Simpson's criticism of the analysis procedures quite seriously. In a later experiment, Hetherington et al. (1987) used both a narrow and a wide energy range in their analysis of their data. Their conclusions, for both the wide-scan and narrow-scan spectra, agreed. "The shape of the plot and the reduced χ^2 value clearly rule out this large a mixing fraction [3%] for the 17 keV neutrino (p. 1510)." They set an upper limit of 0.3% for the mixing probability of the 17 keV neutrino. They did, however, offer

[10] The effect seen by Simpson was quite sensitive to the energy interval chosen. In general, an experimental result should be robust against such changes. See the comments of Hetherington and others below concerning the danger of mistaking a statistical fluctuation for a physical effect. For a more extensive discussion see (Franklin 1998).

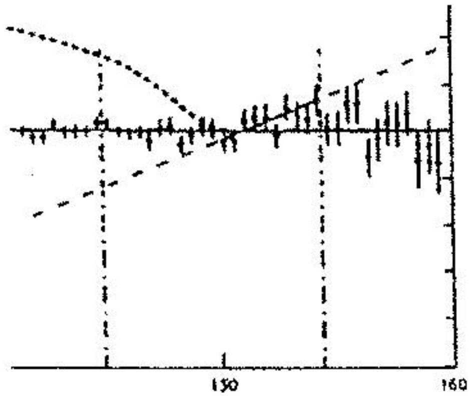

Fig. 10. Morrison's reanalysis of Simpson's reanalysis of Ohi's result. From Morrison (1991)

a note of caution concerning Simpson's analysis. "It has been argued [by Simpson] that in order to avoid systematic errors, only a narrow portion of the beta spectrum should be employed in looking for the threshold effect produced by heavy neutrino mixing. If one accepts this argument, our data in the narrow scan region set an upper limit of 0.44%. However, we feel that concentrating on a narrow region and excluding the rest of the data is not warranted provided adequate care is taken to account for systematic errors. The rest of the spectrum plays an essential role in pinning down other parameters such as the endpoint. Furthermore, concentrating on too narrow a region can lead to misinterpretation of a local statistical anomaly as a more general trend which, if extrapolated outside the region, would diverge rapidly from the actual data (p. 1512)." As one can see from Morrison's later discussion, given above, this was a prescient comment.

There were also difficulties in calculating the expected spectrum shape that was to be compared with the experimental data. Despite the best efforts of the group, "it was found in the analysis that a shape 'correction' of the form $S = (1 + \alpha E)$ was required in order to obtain a good fit. This is probably caused by uncertainties in the instrumental corrections e.g. window absorption, penetration through the edges of the counter slits, electrostatic effects on transmission, etc. (Hetherington, Graham et al. 1987, p. 1508)."[11] This shape correction factor was needed in all of the magnetic spectrometer experiments. There was a question as to whether such a correction factor could mask the presence of a 17-keV neutrino.

The subsequent history of heavy neutrino experiments was quite complex. The year 1991 was the high point in the life of the 17-keV neutrino. Although the evidence for its existence was, at the time, far from conclusive, its existence had been buttressed by recent results from several groups (the most persuasive was that of Hime and Jelley (1991)).[12]

[11] "The penalty paid for having an unknown shape correction is that its interdependence with $|U_{e2}|^2$ raises the error in that parameter (Hetherington et al. 1987, p. 1508)." $|U_{e2}|^2$ gives the probability of the emission of a heavy neutrino.

[12] Other positive results were (Hime and Simpson 1989; Simpson and Hime 1989; Norman, Sur et al. 1991; Sur, Norman et al. 1991; Zlimen, Ljubicic et al. 1991).

From this point on, however, the evidence would be almost exclusively against it. Not only would there be high-statistics, extremely persuasive negative results, but serious questions would be raised about its strongest support.

The question of the ability of the negative experiments to detect the proposed 17-keV neutrino was explicitly raised and answered. Simpson's view of the early negative magnetic spectrometer results was strongly supported by Bonvicini's work (published first as a 1992 CERN report (CERN-PPE/92-54) and later as Bonvicini (1993)). In this work Bonvicini discussed the question of whether a kink in the energy spectrum due to an admixture of a 17-keV neutrino could be masked by the presence of unknown distortions, such as the shape correction factors used in magnetic spectrometer experiments. "Most urgent in this discussion is why experiments where the β^- energy is measured calorimetrically tend to see the effect, and those which use spectrometers do not. My analysis... shows that *large continuous distortions in the spectrum can indeed mask or fake a discontinuous kink*. In the process I point to some deep inconsistencies in all the spectrometer experiments considered here (Bonvicini 1993, p. 97, emphasis added)." He performed a detailed analysis and Monte Carlo simulation of what were then generally regarded as best experiments on either side of the 17-keV neutrino issue: the positive result from ^{35}S by Hime and Jelley (1991), and the negative result from ^{63}Ni by Hetherington and others (1987). He also analyzed several other experiments.

Bonvicini concluded that the positive Hime and Jelley result was statistically sound. He cautioned, however, that the electron response function (the efficiency for electron detection) in this experiment had been only partially measured, and that this might be a possible problem. Bonvicini's analysis of the experiment of Hetherington et al. concluded that although their use of a 2.5 percent shape correction factor was certainly acceptable when searching for a 3 percent kink, when one looked for a 0.8 percent kink more work was needed (Simpson's later work had reduced the size of the observed effect.). His summary of the overall situation was as follows: "A look at the published data seems to indicate that the statistical criteria listed above would eliminate all the negative experiments considered here, but it is left to the authors to look at their data (p. 114)." As far as the positive experiments were concerned, Bonvicini believed that only the Hime-Jelley result was credible. "The ^{35}S result of Hime and Jelley is statistically sound, as they have run the checks suggested in this paper,.... Thus there is only one experiment at this time and in my knowledge where one could say that a kink is certainly there (p. 116)."

Bonvicini's work argued quite strongly that the negative results of the previous magnetic spectrometer experiments were inconclusive and suggested the design of experiments which either used no shape correction factor or had such overwhelming statistical accuracy that a kink would always be visible. In his review of the subject, Hime (1992) made similar suggestions, including the criterion that the experimenters demonstrate that a heavy neutrino signal would be observed in their experiment. He observed that "Given the obvious disagreement between magnetic spectrometer searches on the one hand and the positive results with solid state detectors on the other it is now generally agreed that insight into the discrepancy could be made if the sensitivity of a magnetic spectrometer to uncover a heavy neutrino signal could be experimentally demonstrated. Proposals include measurements with a mixed source such as (99%^{35}S + 1%^{14}C), or artificially invoking energy loss in part of the spectrum at some predetermined level.

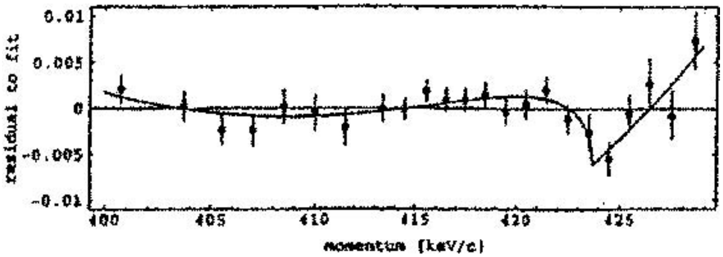

Fig. 11. Synthetic kink induced in the beta spectrum of ^{35}S by a 17 μm aluminum foil. The solid curve is the spectrum expected with a 2.5% admixture of a 15.6-keV neutrino. From Radcliffe et al. (1992)

This latter approach was suggested by the Caltech group and has been implemented in their program (p. 1310)." Experiments of this type were, in fact, performed and were decisive in answering the question as to whether or not the 17-keV neutrino existed.

The Caltech experiment Hime referred to was that of Radcliffe et al. (1992). This was the first attempt to demonstrate that a magnetic spectrometer experiment could detect a 17-keV neutrino. The experiment investigated the ^{35}S spectrum with a magnetic spectrometer. They took data in two different runs: a wide energy range, 130–167 keV; and a narrow scan of 10 keV around the kink expected at 150 keV for the 17-keV neutrino. Both runs were consistent with no heavy neutrino and excluded a 17-keV neutrino with a 0.85% mixing probability at the 99.3% confidence level and at the 99.9% confidence level for the wide and narrow scan runs, respectively.

An interesting feature of this experiment was the simulation of a kink in the spectrum. The experimenters shielded 10% of their detector with a 17 micron aluminum foil. The electrons would lose energy in passing through the foil and they expected this energy loss to produce a kink in the spectrum that would simulate a heavy neutrino with a 1% admixture. Their results with the foil in place are shown in Fig. 11. A kink is clearly visible and gave a best fit for a mass of 15.6 keV with a mixing factor of 2.5%, thus demonstrating that a magnetic spectrometer experiment was sensitive enough to detect a 17-keV neutrino, at least at that level. Looking at the figure, one might legitimately wonder whether the apparatus was sensitive enough to detect a heavy neutrino with 1% mixing. The shape of the spectrum distortion produced was also different from that expected for a heavy neutrino.

The problem of conclusively demonstrating that an experiment could detect a heavy neutrino signal was successfully solved by a group at Argonne National laboratory led by Freedman (Mortara, Ahmad et al. 1993). The experiment used a solid-state, Si(Li), detector, the same type used by Simpson and an external ^{35}S source. They also made a more extensive measurement of their detection efficiency than had been done previously. "The present experiment requires that we know the electron response function between 120 and 167 keV. Measurements of the conversion lines of ^{139}Ce at 127, 160, and 167 keV are the principal constraint on the model of the electron response function (p. 395)." Previous ^{35}S experiments had used an electron response function extrapolated from the lower energy ^{57}Co lines. Finally, and perhaps most importantly, this experiment required no arbitrary shape correction factor.

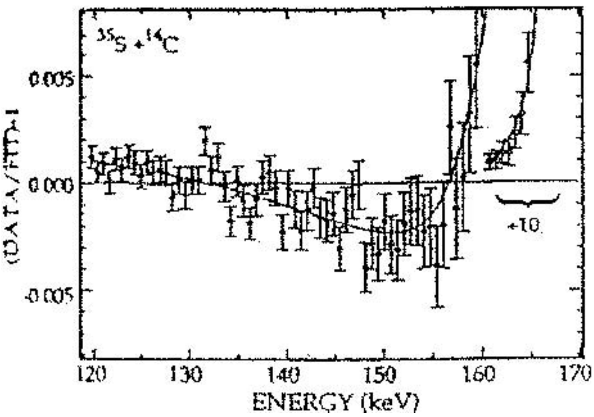

Fig. 12. Residuals from fitting the beta spectrum of a mixed source of ^{14}C and ^{35}S with a pure ^{35}S shape; the reduced χ^2 of the data is 3.59. The solid curve indicates residuals expected from the known ^{14}C contamination. The best fit yields a mixing of (1.4 ± 0.1)% and reduced χ^2 of 1.06. From Mortara et al. (1993)

The experimenters also demonstrated the sensitivity of their apparatus to a possible 17-keV neutrino.

> To assess the reliability of our procedure, we introduced a known distortion into the ^{35}S beta spectrum and attempted to detect it. A drop of ^{14}C-doped valine ($E_o - m_e \sim 156$ keV) was deposited on a carbon foil and a much stronger ^{35}S source was deposited over it. The data from the composite source were fitted using the ^{35}S theory, ignoring the ^{14}C contaminant. The residuals are shown in Fig. [12]. The distribution is not flat; the solid curve shows the expected deviations from the single component spectrum with the measured amount of ^{14}C. The fraction of decays from ^{14}C determined from the fit to the beta spectrum is $(1.4 \pm 0.1)\%$. This agrees with the value of 1.34% inferred from measuring the total decay rate of the ^{14}C alone while the source was being prepared. This exercise demonstrates that our method is sensitive to a distortion at the level of the positive experiments. Indeed, the smoother distortion with the composite source is more difficult to detect than the discontinuity expected from the massive neutrino (p. 396).

Their final result, shown in Fig. 13, was $\sin^2\theta = -0.0004 \pm 0.0008$ (statistical) (±)0.0008 (systematic), for the mixing probability of the 17-keV neutrino. "In conclusion, we have performed a solid-state counter search for a 17 keV neutrino with an apparatus with demonstrated sensitivity. We find no evidence for a heavy neutrino, in serious conflict with some previous reports (p. 396)."

This experiment was clearly convincing. It met all the criteria previously suggested by Hime and Bonvicini along with a demonstrated ability to detect a kink in the spectrum had one been there. Along with the extremely high statistics Tokyo experiment discussed below, it provided very strong evidence against the existence of the 17-keV neutrino

The magnetic spectrometer experiment on ^{63}Ni by the Tokyo group also provided strong evidence against the existence of the 17-keV neutrino (Kawakami, Kato et al. 1992; Ohshima, Sakamoto et al. 1993; Ohshima 1993). The experimenters noted some of the problems of experiments that used wide energy regions and commented that, "We

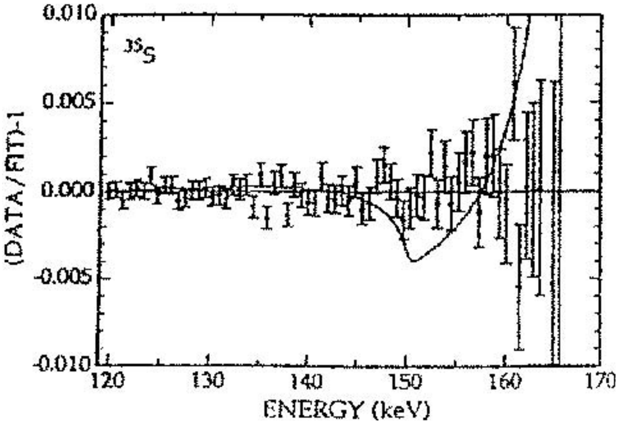

Fig. 13. Residuals from a fit to the pile-up corrected ^{35}S data assuming no massive neutrino; the reduced χ^2 for the fit is 0.88. The solid curve represents the residuals expected for decay with a 17-keV neutrino and $\sin^2 \theta = 0.85\%$; the reduced χ^2 of the data is 2.82. From Mortara et al. (1993)

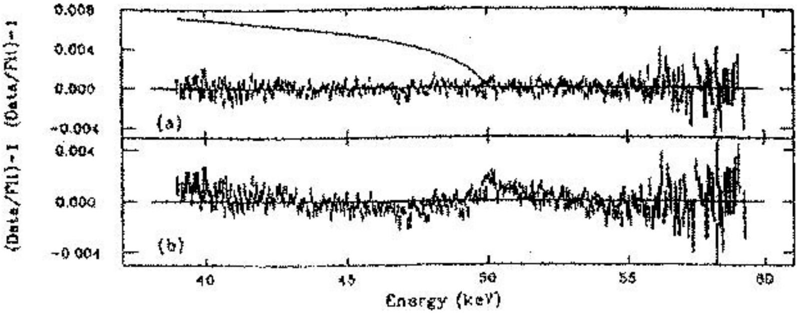

Fig. 14. Deviations from the best global fit with $|U|^2$ free (a) and fixed to 1% (b). The curve in (a) indicates the size of a 1% mixing effect of the 17-keV neutrino. From Ohshima (1992)

have concentrated on performing a measurement of high statistical accuracy, in a narrow energy region, using very fine energy steps. Such a restricted energy scan . . . also reduced the degree of energy-dependent corrections and other related systematic uncertainties (Kawakami 1992, p. 45)." The data were taken over three overlapping energy ranges; 41.2–46.3 keV, 45.7–51.1 keV, and 50.5–56.2 keV [the threshold for a 17-keV neutrino occurs at approximately 50 keV]. The results of their experiment are shown in Fig. 14, for (a) the mixing probability allowed to be a free parameter, and (b) with the probability fixed at 1%. The effect expected for a 17-keV neutrino with a 1% mixing probability is also shown in (a). No effect is seen. Their best value for the mixing probability of a 17-keV neutrino was $(-0.011 \pm 0.033$ (statistical) ± 0.030 (systematic))%, with an upper limit for the mixing probability of 0.073% at the 95% confidence level. This was the most stringent limit yet. "The result clearly excludes neutrinos with $|U|^2 \geq 0.1\%$ for the mass range 11 to 24 keV (Ohshima 1992, p. 1128)."

Although the experiment's narrow energy range was designed to minimize the dependence of the result on the shape correction, the experimenters also checked on the sensitivity of their result to that correction. They normalized their data in the three energy regions using the counts in the overlapping regions, and divided their data into two parts: (A) below 50 keV, which would be sensitive to the presence of a 17-keV neutrino, and (B) above 50 keV, which would not. They then fit their data in (B) and extrapolated the fit to region (A). The resulting fit was far better than one that included a 1% mixture of the 17-keV neutrino, demonstrating that the shape correction was not masking a possible effect of a heavy neutrino. Bonvicini noted that this experiment, with its very high statistics, had convincingly answered his criticism of magnetic spectrometer experiments. "Thus, I conclude that this experiment could not possibly have missed the kink and obtain[ed] a good χ^2 at the same time, in the case of an unlucky misfit of the shape factor (Bonvicini 1993, p. 115)."

In this episode there was, at least for the Argonne experiment, no question concerning the adequacy of the surrogate signal. (Recall the earlier discussion by Hime). The question was whether the experimental apparatus, along with its associated analysis procedures, could detect the signal. Both the Tokyo and Argonne groups demonstrated, albeit in different ways, that they would have detected a heavy neutrino signal had it been present, and set very stringent limits on the presence of a 17-keV neutrino. This combined with problems found with the results of Hime and Jelley (1991) and of Sur et al. (1991), the two most persuasive positive results, convinced the physics community that there was no 17-keV neutrino. (For details see (Franklin 1995)). This episode has also demonstrated the importance of including analysis procedures when examining experimental results.

2. Early attempts to detect gravity waves

In the case of the search for the 17-keV neutrino, discussed in the preceding section, there was no question concerning the adequacy of the surrogate signal used in the experiments. The question was whether the analysis procedures used could detect the signal. This was definitely not the case in the early attempts to detect gravitational radiation (gravity waves). In this episode questions were raised concerning both the adequacy of the surrogate signal and about the analysis procedures used to process the data. As we shall see, these two issues were inextricably intertwined.

Beginning in the late 1960s attempts were made to detect gravity waves. Such waves are predicted by Einstein's General Theory of Relativity. Just as an accelerated, electrically-charged particle will produce electromagnetic radiation (light, radio waves, etc.), so should an accelerated mass produce gravitational radiation. Such radiation can be detected by the oscillations produced in a large mass when it is struck by gravity waves. Because the gravitational force is far weaker than the electromagnetic force, a large mass must be accelerated to produce a detectable gravity wave signal. The difficulty of detecting a weak signal is at the heart of this episode.

In 1969 Joseph Weber claimed to have detected such radiation. Weber used a massive aluminum alloy bar, or antenna, which was supposed to oscillate when struck by gravitational radiation (Fig. 15). The oscillation was to be detected by observing the

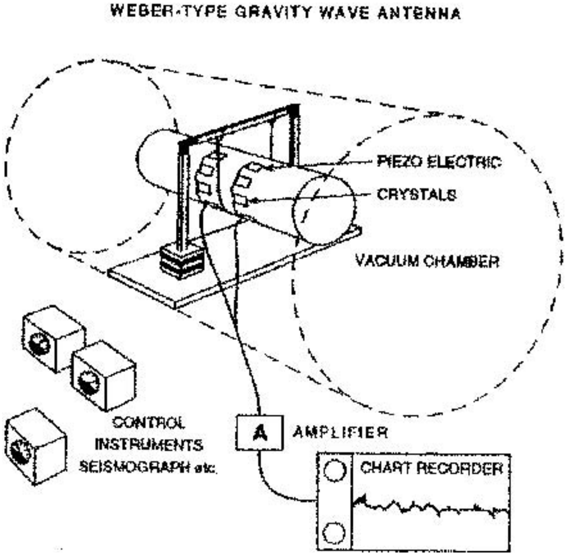

Fig. 15. A Weber-type gravity-wave detector. From Collins (1985)

amplified signal from piezo-electric crystals attached to the antenna. The signals were expected to be quite small and the bar had to be insulated from other sources of noise such as electrical, magnetic, thermal, acoustic, and seismic forces. Because the bar was at a temperature different from absolute zero, thermal noise could not be avoided, and to minimize its effect Weber set a threshold for pulse acceptance. Weber claimed to have observed above-threshold pulses (in excess of those that are to be expected above the threshold from thermal noise).[13] In 1969, Weber claimed to have detected approximately seven pulses/day due to gravitational radiation.

The problem was that Weber's reported rate was far greater than that expected from calculations of cosmic events (by a factor of more than 1000), and his early claims were met with skepticism. During the late 1960s and early 1970s, however, Weber introduced several modifications and improvements that increased the credibility of his results (Weber, Lee et al. 1973). He claimed that above-threshold peaks had been observed simultaneously in two detectors separated by 1,000 miles. Such coincidences were extremely unlikely if they were due to random thermal fluctuations. In addition, he reported a 24 hour periodicity in his peaks,[14] the sidereal correlation, that indicated a single source for the radiation, perhaps near the center of our galaxy. These results increased the plausibility of his claims sufficiently so that by 1972 three other experimental groups had not only built detectors, but had also reported results. None was in agreement with Weber. By 1975 it was generally agreed that Weber's claim was unacceptable.

[13] Given any such threshold there is a finite probability that a noise pulse will be larger than that threshold. The point is to show that there are pulses in excess of those expected statistically.

[14] It was noted by a critic that the periodicity should, in fact, be only twelve hours. This periodic effect later disappeared.

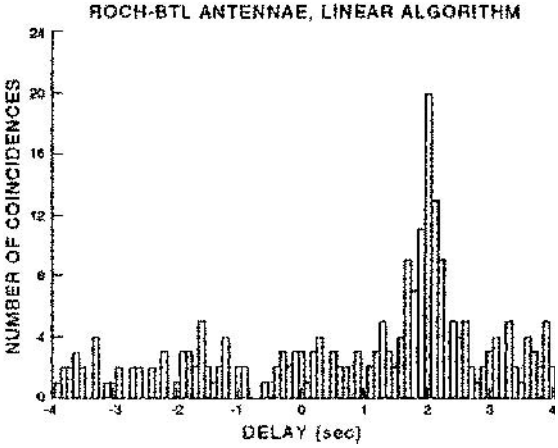

Fig. 16. A plot showing the calibration pulses for the Rochester-Bell Laboratory collaboration. The peak due to the calibration pulses is clearly seen. From Shaviv and Rosen (1975)

Because these experiments used a new type of apparatus to detect a hitherto unobserved phenomenon the question of whether the apparatus could detect a gravity wave signal was crucial. The problem of determining whether or not there is a signal in a gravitational wave detector, or whether or not two such detectors have fired simultaneously is not simple. There are several difficulties. One is due to energy fluctuations in the bar from thermal, acoustic, electrical, magnetic, and seismic noise, etc. When a gravity wave strikes the antenna its energy is added to the existing energy. This may change either the amplitude or the phase, or both, of the signal emerging from the bar. It is not simply a case of observing a larger signal from the antenna after a gravitational wave strikes it.

Weber's critics attempted to show that their experiments could detect gravity waves by using surrogate signals. These were pulses of acoustic energy injected into the antenna to simulate the effect of gravity waves and to test whether the apparatus was working properly. Weber's critics could detect such signals (Fig. 16), whereas Weber could not. Weber admitted that his experiments were twenty times less efficient at detecting the surrogate signals than were those of his critics.[15] How could this happen? The difference was due to a choice of analysis procedure, or algorithm. The nonlinear, or energy, algorithm preferred by Weber was sensitive only to changes in the amplitude of the signal. The linear algorithm, preferred by everyone else, was sensitive to changes in both the amplitude and the phase of the signal.

On the other hand, Weber claimed to have detected gravity wave signals (Fig. 17, upper graph), whereas his critics did not (Fig. 18). How could Weber's experiment fail to detect the calibration signal and yet detect a "real" signal. Weber had an explanation. He suggested that although the linear algorithm was better for detecting calibration pulses,

[15] Harry Collins has questioned whether an acoustic signal was an adequate surrogate for a gravity wave signal. No such questions were raised at the time. In addition, one might argue that a force is a force. Weber's critics's use of both analysis algorithms, discussed below, made the question irrelevant.

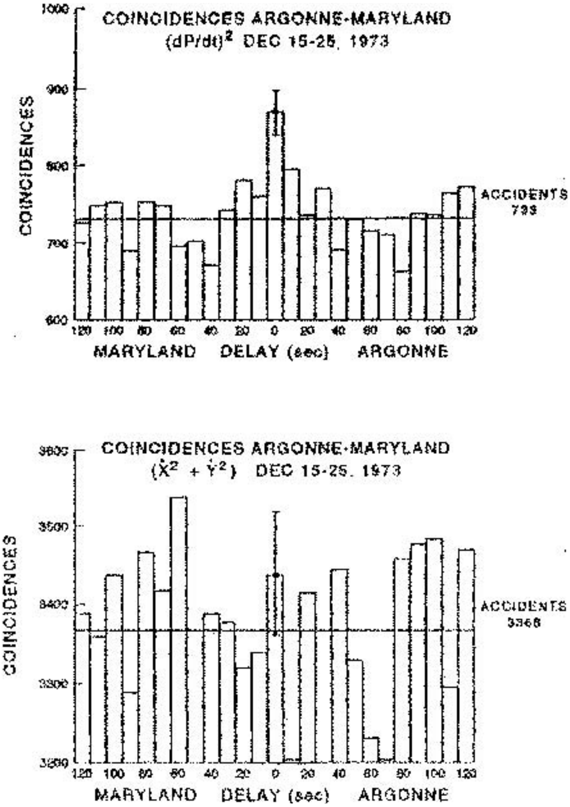

Fig. 17. Weber's time-delay data for the Maryland-Argonne collaboration for the period 15–25 December 1973. The top graph uses the non-linear algorithm, whereas the bottom uses the linear algorithm. From Shaviv and Rosen (1975)

which were short, the real signal of gravitational waves was a longer pulse than most investigators thought. He argued that the nonlinear algorithm that he used was better at detecting these longer pulses. (Note that Weber did not detect a gravity wave signal using the linear algorithm (Fig. 17, lower graph)).

Weber's critics responded by using Weber's preferred non-linear algorithm to analyze their data. They failed to detect both the gravity wave signal and the calibration signal. (Fig. 19).[16] In addition, Drever, one of Weber's critics, reported that he had looked at the sensitivity of his apparatus with arbitrary waveforms and pulse lengths. Although he found a reduced sensitivity for longer pulses, he did analyze his data explicitly to look for such pulses. He found no effect with either the linear or non-linear analysis.[17]

---

[16] There was considerable cooperation between the experimental groups, They exchanged both data and analysis programs.

[17] Drever summarized the situation in June 1974 as follows.

> Perhaps I might just express a personal opinion on the situation because you have heard about Joseph Weber's experiments getting positive results, you have heard about three other

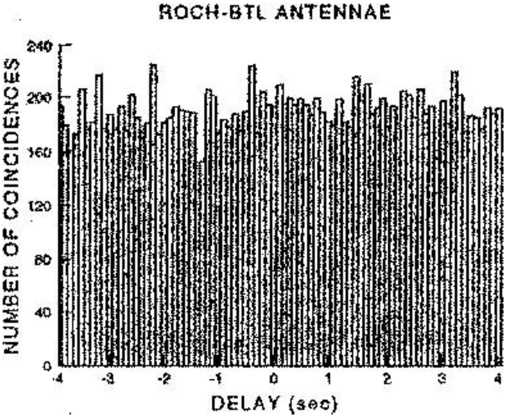

Fig. 18. A time-delay plot for the Rochester-Bell Laboratory collaboration using the linear algorithm. No sign of a zero-delay peak is seen. From Shaviv and Rosen (1975)

These analyses, combined with the critics's ability to detect the surrogate signals and Weber's failure to do so, would, under ordinary circumstance, have been decisive. Because these experiments used a new type of apparatus to detect a previously unobserved phenomenon, other arguments were both needed, and provided. The details of these arguments is beyond the scope of this paper (See (Franklin 1994) for details), but they involved questions concerning Weber's apparatus, possible selectivity in his analysis procedures, and errors in his analysis programs. Nevertheless we have seen

experiments getting negative results and there are others too getting negative results, and what does this all mean? Now, at its face value there is obviously a strong discrepancy but I think it is worth trying hard to see if there is any way to fit all of these apparently discordant results together. I have thought about this very hard, and my conclusion is that in any one of these experiments relating to Joe's one, there is always a loophole. It is a different loophole from one experiment to the next. In the case of our own experiments, for example, they are not very sensitive for long pulses. In the case of the experiments described by Peter Kafka and Tony Tyson, they used a slightly different algorithm which you would expect to be the most sensitive, but it is only the most sensitive for a certain kind of waveform. In fact, the most probable waveforms. But you can, if you try very hard, invent artificial waveforms for which this algorithm is not quite so sensitive. So it is not beyond the bounds of possibility that the gravitational waves have that particular kind of waveform. However, our own experiment would detect that type of waveform; in fact, as efficiently as it would the more usually expected ones, so I think we close that loophole. I think that when you put all these different experiments together, because they are different, most loopholes are closed. It becomes rather difficult now, I think, to try and find a consistent answer. But still not impossible, in my opinion. One cannot reach a really definite conclusion, but it is rather difficult, I think to understand how all the experimental data can fit together (Drever, in (Shaviv and Rosen 1975, pp. 287–8)).

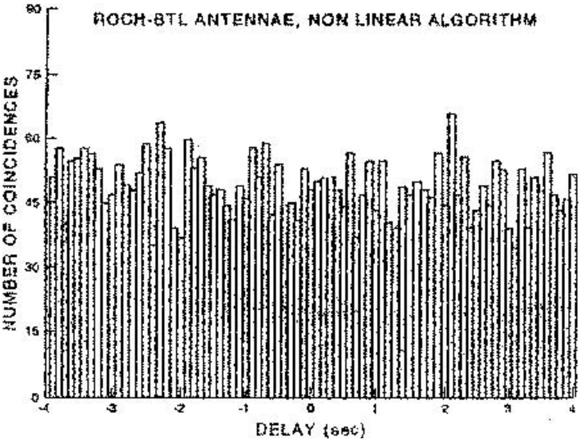

Fig. 19. A time-delay plot for the Rochester-Bell Laboratory collaboration using the non-linear algorithm. No sign of a zero-delay peak is seen. From Shaviv and Rosen (1975)

another example of the use of surrogate signal and the arguments used to demonstrate their adequacy.

3. Other examples

Under ordinary circumstances the presence of background that might mimic or mask the effect one wishes to measure is a serious problem. Sometimes, however, measuring the background can demonstrate that the experimental apparatus can measure that effect. This is illustrated in one of the attempts to measure the branching ratio of $\mu \rightarrow e + \gamma$ (the fraction of all muons that decay into an electron and a γ ray). The motivation for the experiment was "the apparent absence of the decay $\mu \rightarrow e + \gamma$ (unaccompanied by neutrinos), although it is not forbidden by any known selection rules (Bartlett, Devons et al. 1962, p. 120)."

The question of the identity of the muon and electron neutrinos had been raised earlier. It was made more pressing by the failure to observe the decay $\mu \rightarrow e + \gamma$. Ordinary muon decay was thought to be $\mu \rightarrow e + \nu + \bar{\nu}$. If the neutrinos were identical then the neutrino and the antineutrino could annihilate one another before the decay, resulting in the decay $\mu \rightarrow e + \gamma$.[18] The ratio $R = (\mu \rightarrow e + \gamma)/(\mu \rightarrow e + \nu + \bar{\nu})$ could be calculated. The theoretical values ranged from $R = 10^{-3}$ to $R = 10^{-6}$, depending on the choice of theoretical assumptions. As of the beginning of 1962 the measured value of R was $< 2 \times 10^{-6}$ (90% confidence level) (Berley, Lee et al. 1959). Although the result was in disagreement with the most plausible theoretical estimate of R, 10^{-4}, more experimental work was needed to exclude other possibilities. In their study, Bartlett et al. remarked, "Alternatively, suggestions have been advanced for a new selection rule or conservation law to explain the absence of $\mu \rightarrow e + \gamma$ decays. Additive and mul-

[18] In this case we would speak of the annihilation of the virtual neutrino and antineutrino.

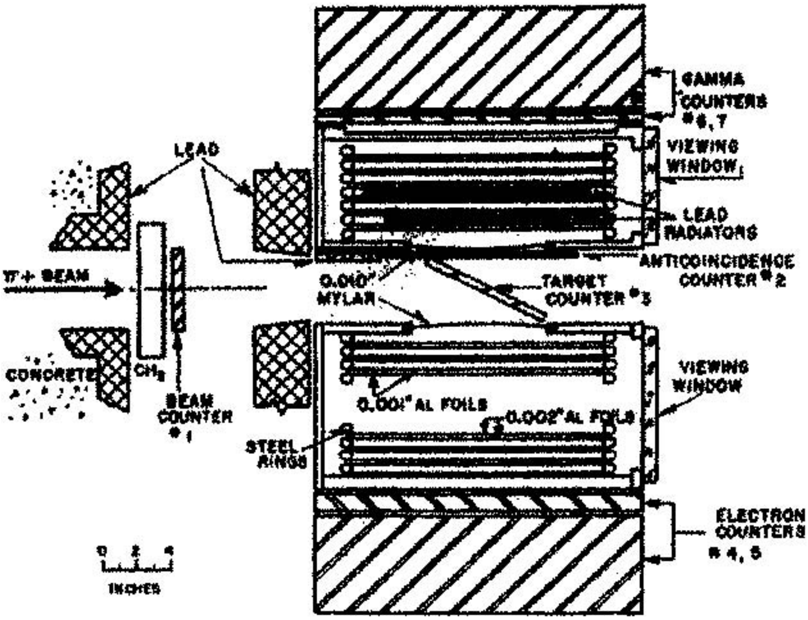

Fig. 20. Plan view of the experimental apparatus for the experiment to search for the decay $\mu \to e + \gamma$. The lower spark chamber and counter assembly detects electrons and the upper one detects γ rays. From Bartlett et al. (1962)

tiplicative types of conservation law have been proposed, both involving two sorts of neutrinos, one associated with electrons, the other with muons. Additional support for *such a radical interpretation* would be provided by a still smaller experimental limit on the $\mu \to e + \gamma$ process (Bartlett, Devons et al. 1962, p. 120, emphasis added)." The limit they provided was $R < 6 \times 10^{-8}$. In an adjoining paper Frankel et al. found a value of $R < 1.9 \times 10^{-7}$ (1962). Both of these results were lower than any existing theoretical estimates. There was clearly a problem. One solution was two neutrinos, an electron neutrino and a muon neutrino. If this were the case then the neutrino and the antineutrino could not annihilate into a γ ray.

A question one might legitimately ask was whether the experiment of Bartlett et al. would have detected the decay $\mu \to e + \gamma$, had it been present. The experimental apparatus is shown in Fig. 20. Positive π mesons were stopped in counter 3. The positrons from the sequential decay $\pi \to \mu \to e$ passed through a thin-plate spark chamber and were detected by counters 4 and 5. γ rays emitted in the opposite direction produced electron-positron pairs in two lead converters located inside a second spark chamber. These pairs were detected by counters 6 and 7. An anticoincidence counter 2 in front of the γ ray chamber discriminated against charged particles. Conservation of momentum requires that the positron and the γ ray from the decay at rest of $\mu \to e + \gamma$ be emitted at 180° relative to one another (back to back). This is not the case for the positron and the γ ray emitted in the internal bremsstrahlung process, $\mu \to e + \gamma + \nu + \bar{\nu}$. They can be emitted at any angle relative to one another and those emitted near 180° will simulate those expected from $\mu \to e + \gamma$.

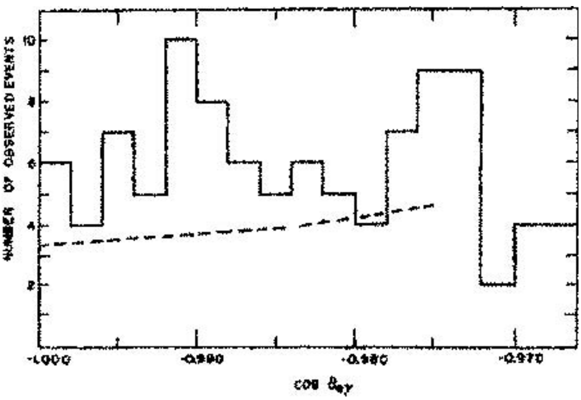

Fig. 21. Number of events as a function of cos $\theta_{e\gamma}$. No peak is seen at 180° (cos $\theta_{e\gamma} = -1$). From Bartlett et al. (1962)

The angular distribution for measurements with $\Theta_{e\gamma}$, the angle between the positron and the γ ray, between 165° and 180° is shown in Fig. 21. Events due to $\mu \rightarrow e + \gamma$ should be seen as a peak at 180° (Cos$\Theta_{e\gamma} \approx -1$). (Because of multiple scattering of the positrons one expects the peak within 3° of 180°. No significant peak can be seen. "From the average rate of events between 165 and 180 deg, one would expect 4.2 events in the significant region. The actual number is 5, consistent with a zero rate for the process $\mu \rightarrow e + \gamma$. (Bartlett, Devons et al. 1962, p. 121)." The experimenters were able to set an upper limit of 6×10^{-8} for the branching ratio.

The demonstrated ability of the experimental apparatus to detect the coincident positrons and γ rays from the background internal bremsstrahlung process $\mu \rightarrow e+\gamma+\nu+\bar{\nu}$ also argued that it would detect such pairs from $\mu \rightarrow e + \gamma$. There was no dependence of the detection efficiency as a function of $\Theta_{e\gamma}$. This was shown by the good fit between the observed distribution and the theoretical calculation. In this case the background served as a surrogate signal.

Although the experiment of Bartlett et al. had suggested the existence of two types of neutrino, more evidence was required. That was provided by experiments that searched for the reactions: (1) $\nu_\mu + p \rightarrow \mu^+ + n$ and (2) $\nu_\mu + p \rightarrow e^+ + n$. If there were two different neutrinos then Reaction (2) would not be observed. Reaction (1) required a neutrino energy greater than 100 MeV.

A solution to the technical problem of how to create a beam of high energy neutrinos with sufficient intensity was proposed independently by Pontecorvo (1960) and by Schwartz (1960). Both of them proposed using high-energy pions produced in the collision of high-energy accelerator protons with a metal target. The pions would then decay into a muon and a neutrino [$\pi^\pm \rightarrow \mu^\pm + (\bar{\nu}|\nu)$]. The neutrino energy would be a significant fraction of the pion energy and would be emitted along the pion direction.

> That night it came to me. It was incredibly simple. All one had to do was use neutrinos [to study high-energy weak interactions]. The neutrinos would come from pion decay and a large shield could be constructed to remove all background consisting of the strongly and electromagnetically interacting particles and allow only neutrinos through. . . They [T.D. Lee and C.N. Yang] also pointed out that this experiment could resolve the long standing

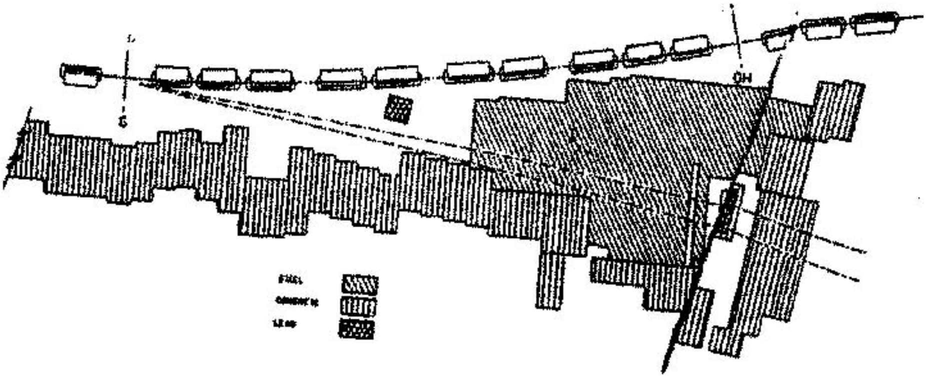

Fig. 22. Plan view of the two-neutrino experiment. From Danby et al. (1962)

> puzzle of the missing decay of the muon into electron and gamma. There were clear-cut theoretical predictions, in contradiction to the experiments, that $\mu \rightarrow e + \gamma$ should take place in one in every 10^5 muon decays, unless there is a new quantum number to forbid it. Indeed it became increasingly clear that the only way in which this absence could be explained required that there be two neutrinos, one associated with the electron and the other associated with the muon. In this case, making use of neutrinos from the decay ($\pi \rightarrow \mu + \nu$) we would only see muons produced, never electrons. Estimates at that point (Schwartz 1960) indicated that with 10 tons of detector we might obtain an event per day, if the new Alternate Gradient Synchrotron at Brookhaven accelerated as much as 10^{11} protons per second. (Schwartz 1972, pp. 82–83).

The type of beam proposed by Pontecorvo and Schwartz was used by Schwartz, Lederman, Steinberger, and their collaborators in an experiment to test the two-neutrino hypothesis at the Alternate Gradient Synchrotron (Danby, Gaillard et al. 1962).The plan view of the experiment is shown in Fig. 22. Pions were produced by the 15 GeV protons striking a beryllium target at G. The entire flux of particles struck 13.5 m of steel[19] in front of a 10-ton spark chamber that served as a detector. The shielding removed virtually all of the beam particles except neutrinos. The group obtained a total of 113 pictures of which 34 contained single tracks (Fig. 23), which, if interpreted as muons, had momenta greater than 300 MeV/c, 22 were "vertex" events which had more than one track, and 8 were "showers," that were "in general single tracks, too irregular in structure to be typical of μ mesons, and more typical of electron or photon showers."

The experimenters offered arguments that the observed events were not produced by the possible backgrounds of cosmic rays or neutrons. The cosmic ray background was estimated by operating the experimental apparatus with the accelerator off. They found a total of 5 ± 1 events that could be attributed to cosmic rays. Neutrons were eliminated as a possible cause of the observed events because there was no measured attrition of the observed events as a function of distance in the detector that would be

[19] The steel used came from the armor plate of a battleship. Although it wasn't beating swords into plowshares, it was cutting armor into shielding, a peaceful use. Other high-energy physics experiments use barrels from naval guns, cutting cannon into collimators.

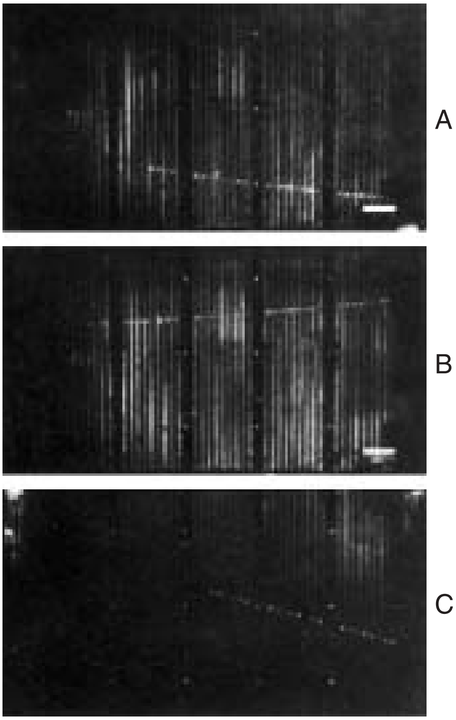

Fig. 23. Single muon events. (A) $p_\mu > 540$ MeV/c and ray indicating direction of motion (neutrino beam incident from the left; (B) $p_\mu > 700$ MeV/c; (C) $p_\mu > 440$ MeV/c with δ ray. From Danby et al. (1962)

expected if they were caused by neutrons, because the shielding reduced the calculated number of such events by a factor of ten thousand, and by checking that the event rate remained unchanged when four feet of iron was removed from the shield wall. If the observed events were due to neutrons then this change would have increased the number of events by a factor of one hundred. No such increase was observed. In addition, if the 29 single-track events (34 observed - 5 background) were neutron induced then 15 neutral pions should have been observed. None were found.

The group also presented arguments that the single particles observed were muons and that they were due to the decay products of pions and K mesons, i.e. neutrinos. For the former they argued that the single tracks traversed a total of 820 cm of aluminum without producing a single "clear"nuclear interaction. The interaction length for 400 MeV pions, the alternative explanation of the observed tracks, was less than 100 cm. "We should, therefore, have observed of the order of 8 'clear' interactions [if the tracks were pions]; instead we observed none." The tracks were muons, not pions. In addition, the experimenters moved four feet of iron shielding from the main shield and placed it as close to the target as was feasible. This reduced the distance in which the pion could decay by a factor of eight. The number of events observed fell from 1.46 ± 0.2 to 0.3 ± 0.2 per 10^{16} incident protons. "This reduction is consistent with that which is expected for neutrinos which are the decay products of pions and kaons (p. 42)."

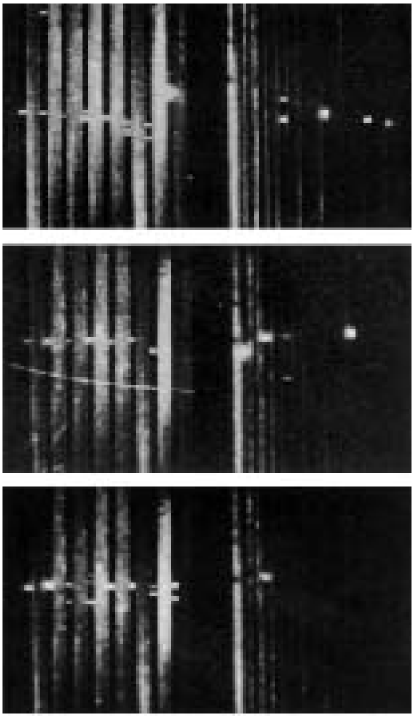

Fig. 24. 400 MeV electrons from the Cosmotron. From Danby et al. (1962)

It was clear that neutrinos were producing muons. The question of whether they were also producing electrons remained. "Are there two kinds of neutrinos? The earlier discussion leads us to ask if the reactions (2) and (3) [(2)$\nu + n \rightarrow p + e^-, \bar{\nu} + p \rightarrow n + e^+$; (3)$\nu + n \rightarrow p + \mu^-, \bar{\nu} + p \rightarrow n + \mu^+$] occur with the same rate. This would be expected if ν_μ, the neutrino coupled to the muon and produced in pion decay, is the same as ν_e , the neutrino coupled to the electron and produced in nuclear beta decay (p. 42)." They noted that the tracks for their muon events (Fig. 23) were quite different from the showers produced in their spark chambers by 400 MeV electrons (Fig. 24). They exposed two of their spark chambers to beams of 400 MeV/c electrons to provide a surrogate signal for the unobserved electron events. In this case the surrogate signal, 400 MeV electrons was identical to the signal expected. "We have observed 34 single muon events of which 5 are considered to be cosmic-ray background. If $\nu_\mu = \nu_e$, there should be of the order of 29 electron showers with a mean energy greater than 400 MeV/c. Instead, the only candidates which we have for such events are six showers' of qualitatively different appearance from those of Fig. [24](p. 42)." The distribution of sparks from the electron events is quite different from that of the "shower" events (Fig. 25). The "shower" events were attributed to either neutron background or to electron neutrinos from kaon decay. They were not showers from neutrino produced electrons. The experimenters demonstrated, using a

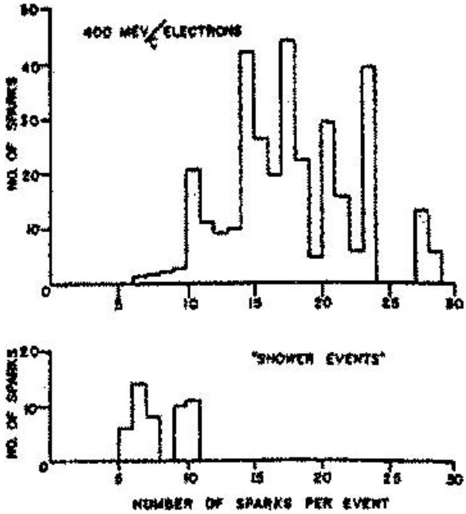

Fig. 25. Spark distribution for 400 MeV electrons normalized to expected number of showers. Also shown are the "shower" events. From Danby et al. (1962)

beam of 400 MeV electrons, that if electrons had been produced they would have been detected.

The experimenters concluded, "However, the most plausible explanation for the absence of the electron showers, and the only one that preserves universality is then that $\nu_\mu \neq \nu_e$; that there are at least two types of neutrinos. This also resolves the problem raised by the forbiddenness of the $\mu^+ \rightarrow e^+ + \gamma$ decay (p. 42)." Now there were two.[20]

Sometimes a surrogate signal may be unavailable, or extremely difficult to use. In such an experiment a computer simulation might be used to provide such a signal. Consider the case of the Fermilab E791 collaboration's search for rare decays of the D meson (Aitala, Amato et al. 1996). There was, in fact, no serious question about the proper operation of the experimental apparatus, but there was also no surrogate signal available. One of the key elements in the experiment was the identification of decay particles as either muons or electrons. The experimental apparatus is shown in Fig. 26. The search for $D^+ \rightarrow \pi^+\mu^+\mu^-$ and $D^+ \rightarrow \pi^+e^+e^-$ decays also required muon and electron identification criteria, respectively, Such criteria can be, and were, set independent of the final result. The muons were identified by scintillation counters located behind 15 interaction lengths of shielding. Muons have a longer range in

[20] The discovery of the muon neutrino led physicists to formulate two separate conservation laws, one for electron family members and one for muon family members. Previously one only required that the number of leptons, or light particles, be conserved. Now the decay of the muon was characterized as $\mu^- \rightarrow e^- + \bar{\nu}_e + \nu_\mu$. This conserved both family numbers. If $\bar{\nu}_e$ is not the antiparticle of ν_μ then there can be no annihilation of the virtual neutrino and antineutrino, and thus no decay $\mu \rightarrow e + \gamma$. Subsequently, physicists discovered a third family: the τ lepton and its neutrino. For details see (Franklin 2000a, Chap. 7).

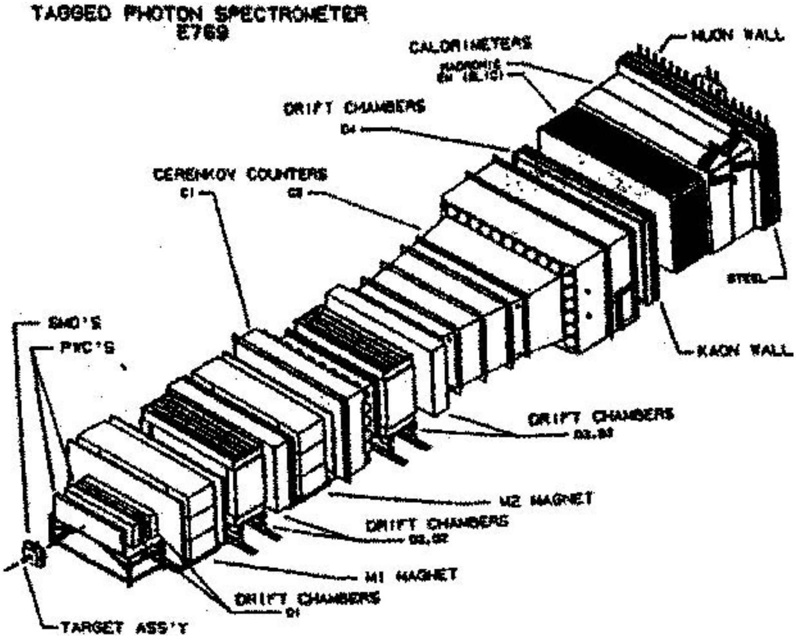

Fig. 26. Experimental apparatus for experiment E791. From Aitala et al. (1996)

matter than either pions or electrons, and the probability of either electrons or pions penetrating the shielding was very low. The muon counter efficiency was measured in special runs using independent muon identification and was found to be $(99 \pm 1)\%$. Electrons were identified by the lead and liquid scintillator calorimeter. The identification was based on energy deposition, shower shape, and position in the calorimeter. "Calorimeter response was studied with topologically identified electron-positron pairs from γ conversions upstream of the tracking, and with pions from kinematically identified $K_S^\circ \to \pi^+\pi^-$ decays (p. 366)." Once again, there was an identity between the surrogate and expected signal. The experimenters found no signal above background. They concluded, "In summary, Fermilab experiment E791 has obtained upper limits on branching fractions B for the three-body FCNC decays $D^+ \to \pi^+\mu^+\mu^-$ and $D^+ \to \pi^+e^+e^-$ that are an order of magnitude below those previously published. At 90% C.L. [confidence level], $B(D^+ \to \pi^+\mu + \mu^- < 1.8 \times 10^{-5}$ and $B(D^+ \to \pi^+e^+e^-) < 6.6 \times 10^{-5}$ (p. 367)." These results were accepted by the Particle Data Group as the definitive limits.[21]

Although it was clear that the apparatus would detect both electrons and muons, it was an open question as to whether it would detect the rare decays had they been present. Although there was no apparent way to provide a physical surrogate signal for these rare decays, the experimenters performed a check by inserting computer simulated events

[21] In fact, the 1998 Review of Particle Physics (Caso, Conforto et al. 1998) cited only the E791 results for these decays.

into their observed distribution. "We have also tested the procedure with ensembles of simulated experiments in which fixed numbers (2–10) of simulated FCNC signal events, drawn randomly from a Gaussian mass distribution, are added to the observed spectrum and *successfully found by the fit* (p. 366, emphasis added)." If the decays had been present they would have been detected by the analysis procedures.

C. Now we see it, now we don't

One interesting problem that occurs in experimental science is when an experiment initially observes an effect and subsequent work fails to replicate that result. One might legitimately ask if that is because the observed effect is rare, or because the experimental apparatus malfunctioned, either in producing the original result or in its subsequent operation. How does one argue that one of the experimental results is more reliable? In a sense we have discussed this earlier in the sections on the 17-keV neutrino and on gravity waves. In those episodes we had discordant results produced by different experimental groups. In this section I will examine cases in which all of the work was done by a single experimental group over a period of time.

1. Time reversal violation?

Perhaps the most exciting experimental result reported at the February, 1969, New York meeting of the American Physical Society was that reporting an observation of the violation of time reversal symmetry by a group from Princeton University.[22] The 1950s and 1960s had seen observation of the violation of other discrete symmetries: parity, or left-right symmetry, by (Friedman and Telegdi 1957; Garwin, Lederman et al. 1957; Wu, Ambler et al. 1957) and of CP symmetry, combined parity and charge conjugation or particle-antiparticle symmetry, by (Christenson, Cronin et al. 1964) (For details see (Franklin 1986, Chaps. 1 and 3)). If CPT symmetry, combined parity, charge conjugation, and time reversal symmetry, a requirement for all local field theories, held then CP violation implied the violation of time reversal symmetry. This violation had never been observed directly.

The Princeton group reported that a preliminary analysis of their data showed a five standard-deviation difference between the angular distributions for $\mathrm{n} + \mathrm{p} \rightarrow \mathrm{d} + \gamma$ and its time-reversed reaction $\gamma + \mathrm{d} \rightarrow \mathrm{n} + \mathrm{p}$. (The probability that a five standard deviation effect is a statistical fluctuation is 5.73×10^{-7}). In his talk at the meeting, Carl Friedberg announced that the group had seen definite evidence of time-reversal violation.[23, 24]

The result generated considerable excitement. The journal *Scientific Research* reported that "Time-symmetry violation in the electromagnetic interaction has been

[22] The group consisted of David Bartlett, Carl Friedberg, Dino Goulianos, Ira Hammerman, and David Hutchinson.

[23] I was present at the talk.

[24] Unfortunately the original graph showing the two distributions is no longer available. As discussed below, further analysis reduced the size of the effect.

indicated in preliminary results reported at the American Physical Society meeting by a group of physicists at Princeton University (17 February, 1969, p. 14)". Saul Barshay, a theoretical physicist who had calculated such an effect wrote from Europe to Bartlett and Goulianos, the group leaders, "The rumors here make it sound like the days of Lee and Yang and Madame Wu and Co. all over again [a reference to the discovery of parity nonconservation]. A spectacular effect overthrowing a fundamental symmetry. Is it true? (S. Barshay, letter to D. Bartlett. 24. Feb. 1969)"

The experimental apparatus is shown in Fig. 27. A neutron beam from the Princeton-Pennsylvania Accelerator impinged on a liquid hydrogen target.[25] The beam of the PPA was bunched so that the energy of the neutron could be determined by measuring the time of flight between a counter placed near the internal platinum target of the accelerator and counter D_1 . The deuterons from the reaction $n + p \rightarrow d + \gamma$ passed through spark chambers S_1–S_5 and were detected in counters D_1 and D_2–D_4 . The momentum of the deuteron was determined by measuring the trajectory of the particle on either side of the bending magnet shown in the figure. The time of flight between counters D_1 and D_2–D_4 was recorded and measured the particle's velocity. The two measurements, of the particle's momentum and velocity, determined the particle's mass, allowing a separation of deuterons from protons.

The γ rays were converted into electron-positron pairs in three lead-plate spark chamber arrays 1-2, 3-4-5, and 6. An anticounter, to ensure that a neutral particle entered the chambers, was placed before each of these spark chambers. This was supplemented with a requirement that the first gap in the chambers did not fire. The charged particles produced in the converters were detected by scintillation counters placed in the chambers. The experimenters had measured the momentum and energy of the neutron and the deuteron, along with the angles relative to the neutron beam of the deuteron and the γ ray.

A major problem for the experimenters was to separate the deuterons from the desired reaction $n + p \rightarrow d + \gamma$ from those produced by the 100 times more likely reaction $n+p \rightarrow d+\pi^\circ$. (The π° meson decays into two γ rays and thus events from the latter reaction can mimic those from the desired reaction). This was done in three stages. First, using the measured neutron and deuteron momenta one can calculate the mass of the missing neutral particle, $M_x(n, d)$, either a γ ray or a π° meson (the mass of the γ ray is zero).

$$M_x^2(n, d) = (E_n + M_p) - E_d)^2 - (\mathbf{p}_n - \mathbf{p}_d)^2$$

where $\mathbf{p}_n$ and $\mathbf{p}_d$ were the vector momenta of the neutron and deuteron, respectively. "In principle, this calculation permits the separation of $n + p \rightarrow d + \gamma$ events ($M_x^2 = 0$) from $n + p \rightarrow d + \pi^\circ$ events ($M_x^2 = M_\pi^2$). In practice, the two mass peaks overlap considerable owing to the poor resolution of the neutron momentum ($\Delta p/p \approx 4\%$ at $T_n = 600$ MeV) (Bartlett, Friedberg et al. 1969, p. 894)." The experimenters made a cut at $M_x^2(n, d) < 0.66M_\pi^2$.

This cut used no information about the γ ray. The second stage used the polar angle, $\theta_{\gamma n}$, of the γ ray. The experimenters assumed that the neutron momentum was unknown

[25] Charged particles and γ rays were eliminated by passing the beam through a lead converter and then a sweeping magnet which bent the charged particles out of the beam.

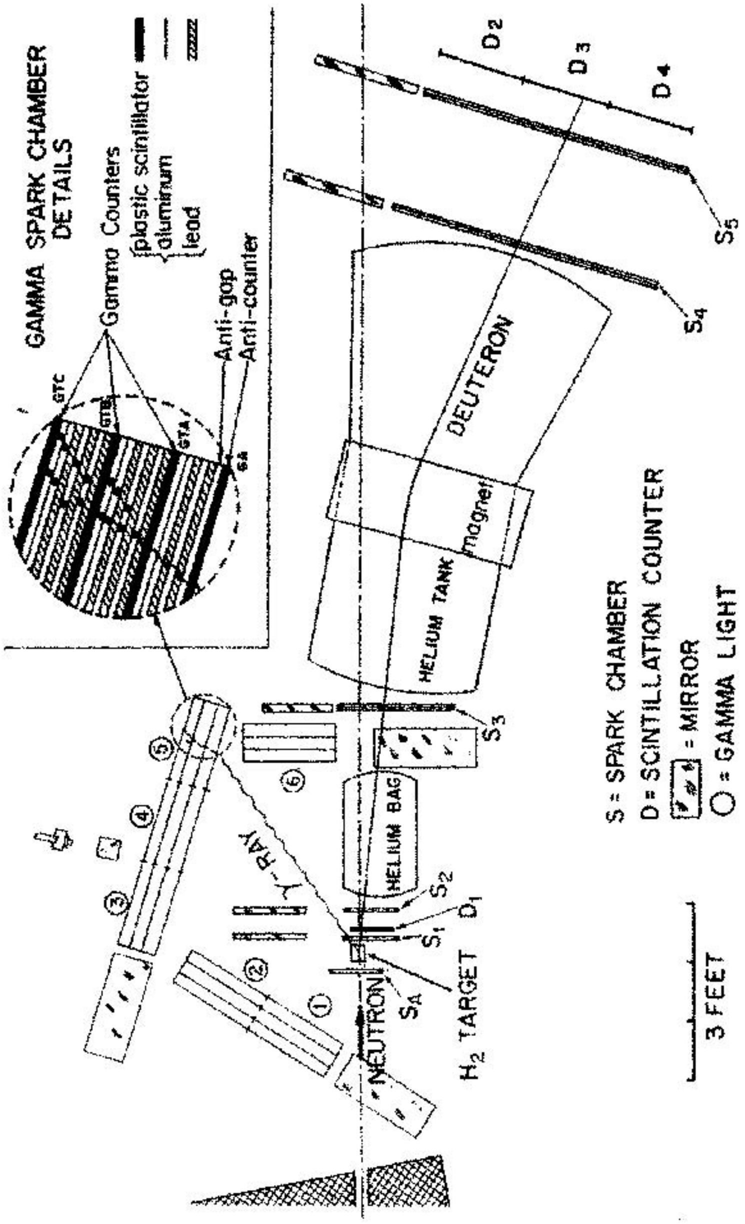

Fig. 27. Experimental apparatus to investigate n + p → d + γ. From Friedberg (1969)

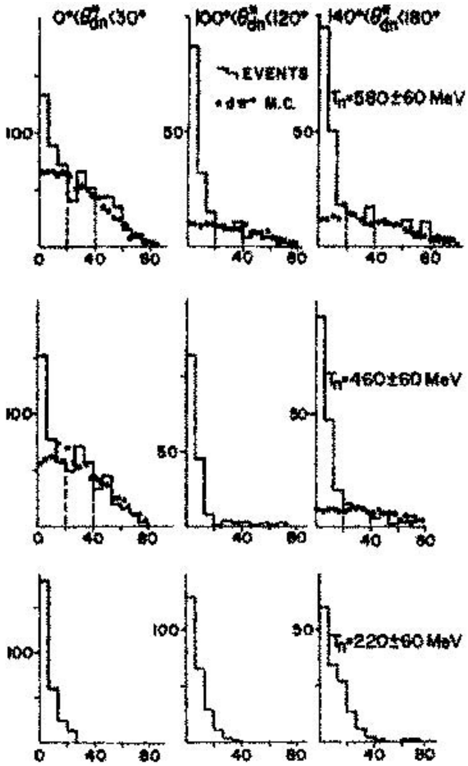

Fig. 28. "Coplanarity" (in mm) of events in three angular intervals. From Bartlett et al. (1969)

and used the angle $\theta_{\gamma n}$ along with the other measured quantities to construct a neutron momentum p_n'. This calculated momentum was used to calculate a missing mass $M_x^2(\gamma, d)$

$$M_x^2(\gamma, d) = (E_n' + M_p - E_d)^2 - (\mathbf{p}_n' - \mathbf{p}_d)^2$$

Events with $M_x^2(\gamma, d) < 0.55M_\pi^2$ were selected for the final stage of analysis.

This final stage made use of the fact that the γ rays from $n + p \rightarrow d + \gamma$ were coplanar with the neutron and the deuteron, whereas γ rays from the decay of the from the π° decay in the reaction $n + p \rightarrow d + \pi^\circ$ are, in general, not coplanar. A plane was defined by the γ-ray conversion point, the deuteron spark in chamber S_2 and the center of the internal platinum target of the accelerator. The experimenters defined 'coplanarity' as the distance in millimeters of the deuteron spark in chamber S_3 from this plane. The coplanarity for various neutron energies and angular intervals is shown Fig. 28 along with the results of a Monte Carlo simulation of $n + p \rightarrow d + \pi^\circ$ (For neutron energies in the range 220 ± 60 MeV no π° mesons can be produced). "At all energies above the $d\pi^\circ$ threshold, more than 98% of the $d\gamma$ signal is confined to coplanarities smaller than 20 mm. An exhaustive Monte Carlo calculation of the background reaction $n + p \rightarrow d + \pi^\circ$ shows that the background falls as the coplanarity increases from 0 mm

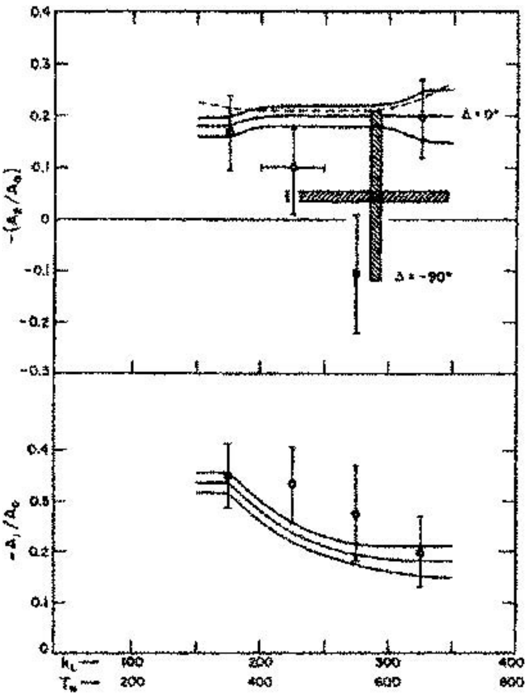

Fig. 29. Comparison of the results of Schrock et al. with the inverse reaction. Circles are the data. The dotted line is the theory and the cross-hatched cross is the prediction of Barshay. From Schrock et al. (1969)

to 40 mm. Based on the number of events observed with coplanarity between 20 mm and 40 mm, the Monte Carlo program predicts the number of background events to be expected between 0 mm and 20 mm. The number of dγ events is then determined by subtracting the predicted number from the number of events actually observed between 0 and 20 mm (Bartlett, Friedberg et al. 1969, p. 895)." As one can see from the figure the Monte Carlo simulation accurately fits the background above 20 mm, giving confidence in the subtraction. Using another Monte Carlo calculation to calculate the efficiency for detecting the $n+p \rightarrow d+\gamma$ reaction in various energy and angular bins, the final angular distributions were obtained.

The Princeton group's result was supported by the results of an experiment performed by (Schrock, Detoeuf et al. 1969). They, too, measured the angular distribution for the reaction $n + p \rightarrow d + \gamma$ and compared it to the inverse reaction. "We find an intriguing discrepancy in the vicinity of the well-known peak in the total cross section of the reaction (1b) [$\gamma + d \rightarrow n + p$] and apparent agreement elsewhere (p. 727)." Their results are shown in Fig. 29. The graphs show the fitted coefficient to a Legendre polynomial fit to the angular distributions. The lines are the results for the inverse reaction. They noted that their result was based on an analysis of twenty percent of their data, and that they "have had a result of this nature for over a year and have found no systematic errors in our treatment that could lead to this discrepancy (p. 733–34)."

Even before the confirmation by Schrock and collaborators, the Princeton group had modified their dramatic conclusion. In a talk given at the American Physical Society meeting in Washington in April 1969, the size of the discrepancy had been reduced. Their final result, published in October 1969 is shown in Fig. 30 (Bartlett, Friedberg et al. 1969). For the neutron energy range 580 ± 60 MeV there is a disagreement with the angular distribution for the inverse reaction obtained by Anderson and collaborators (Anderson, Prepost et al. 1969). As shown in the figure, the distributions agreed at lower energies. "We have also made a direct comparison of our data points to the fit to their [Anderson et al.] data. If the relative normalization is varied to minimize the χ^2 , We find $\chi^2 = 10$ for six degrees of freedom (Bartlett, Friedberg et al. 1969, p. 896)." This fit has a probability of approximately 12 percent, a far cry from their earlier five standard-deviation effect which had a probability of 5.73×10^{-5} percent. What had changed?

There had been a problem with the experimental apparatus that had appeared only with further analysis of the data

> The seventh gap following the H_2 target, located in chamber S5, was known to break down rather frequently. During the run, several steps were taken to remedy this problem, notably removing the chambers for cleaning and reassembly, and flushing argon to attempt to quench the breakdown. However, the problem persisted at what was believed to be a small level. During the scanning of the film by SPASS [an automated spark-chamber film scanner], however, it became apparent that a significant fraction of events were lost because of this breakdown. A careful study of the SPASS error codes has made it possible to untangle this problem. By looking at chamber S4, and extrapolating into S5, it was possible to separate those events (for the purpose of this study) which had tracks in S5 but for which SPASS could not locate a track. A region extending from each side of the chamber about 3 inches was found to be inefficient, as shown in Fig. [31]. This inefficiency was probably due to either an insufficient gas flow, or else to poisoning from vapors from uncured R-T-F sealant which was used in the chamber construction. Unfortunately, this inefficiency was more severe in the SPASS scanned film (a machine is more likely to have a sharp threshold than a human, who could more easily distinguish faint sparks amidst the breakdown). The effect is now well understood, and the correction is accurate to 10%. The largest inefficiency for the $d\gamma$ data is for $T_n = 220$ MeV and $\theta^* = 107°$, with a $15 \pm 2\%$ loss. *However, the effect on the shape of the background coplanarity is dramatic* (Friedberg 1969, p. 89, emphasis added).

The inefficiency of the spark chamber had led to an underestimate of the background due to $d\pi^\circ$ events. The later results, as opposed to the initial report, included this inefficiency. This accounted for the difference.

Nevertheless there was still an intriguing, if small, discrepancy between the angular distributions for $n + p \rightarrow d + \gamma$ and $\gamma + d \rightarrow n + p$. The group decided to continue their work with a new experimental apparatus.

Experimenters rarely just repeat a measurement, they attempt to improve it. The new experiment differed "from the previous one in three important aspects: (a) the number of $n + p \rightarrow d + \gamma$ events collected is about 10 times larger, (b) angles and momenta are measured about 2.5 times more accurately, and (c) data are taken over a wider angular range and extend to higher neutron energies (Bartlett, Friedberg et al. 1971, p. 882)." In addition, "apparatus inefficiencies were *measured*, and errors in these measurements

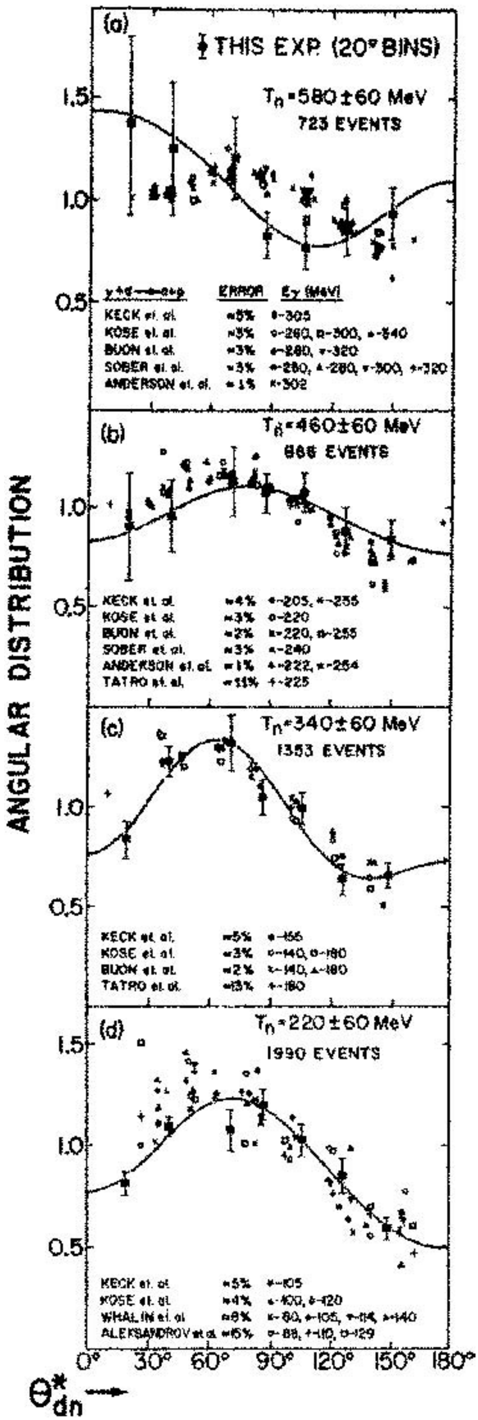

Fig. 30. Angular distributions for $n + p \to d + \gamma$ and $\gamma + d \to n + p$. From Bartlett et al. (1969)

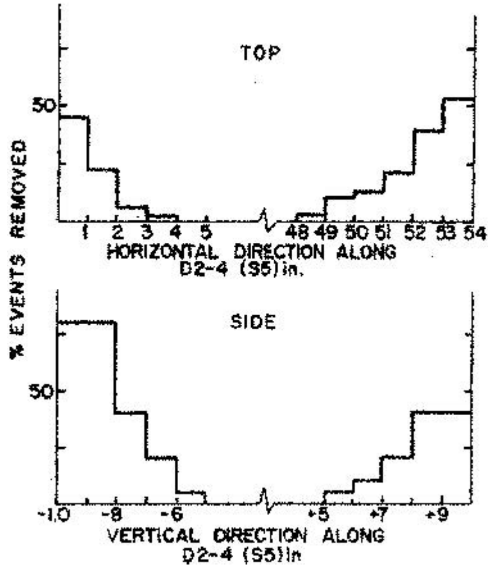

Fig. 31. Percentage of events removed by scanning as a function of position. From Friedberg (1969)

were included in the errors in the angular distribution. The separation of the $n+p \rightarrow d+\gamma$ events from background was performed in two independent ways, neither of which relied on Monte Carlo programs (Goldhagen 1973, p. 4)."

The experimental apparatus is shown in Fig. 32. Photons were converted in lead plates placed in front of two wire spark chambers. An anticounter (GA) in front of the converters guaranteed the neutrality of the incoming particle. Two scintillation counters (G1 and G2) located behind the spark chambers detected the electron shower produced. The time of flight of the γ rays was measured and its direction taken as the line joining the conversion point to the intercept of the extrapolated deuteron trajectory with the hydrogen target. The direction of the deuteron was measured with six wire chambers and its momentum was calculated from the measured time of flight between the counter N and scintillation counters D. (This experiment did not use a magnetic field to determine the deuteron momentum). The neutron momentum was calculated from the measured time of flight between the RF signal of the accelerator and counter N, which detected the deuteron emerging from the hydrogen target.

Because time-of-flight measurements were so crucial to the experiment, small variations in the time of flight due to the position of the particle in the scintillation counters as well as variations due to pulse height in the photomultiplier were measured and included in the analysis. Another experimental improvement was the division of both the γ-ray counters and the deuteron counters into two groups placed symmetrically "up and down" relative to the neutron beam. The event trigger required that an "up" γ ray be associated with a "down" deuteron, and vice versa. This was essentially a loose coplanarity requirement. It also helped to reduce the background π° events which would produce "up-up" and down-down"configurations, in addition to the "up-down" events.

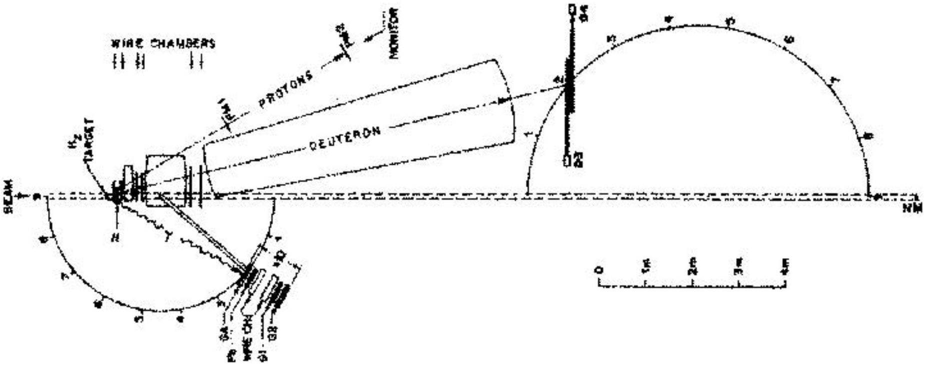

Fig. 32. Experimental apparatus for $n + p \rightarrow d + \gamma$. From Bartlett et al. (1971)

Once again, the crucial part of the analysis involved the separation of $n + p \rightarrow d + \gamma$ from the far more numerous $n + p \rightarrow d + \pi^\circ$ background events. The experimenters used the kinematic variables $p_n, p_d, \theta_d, \theta_\gamma$, and $\phi_\gamma - \phi_d$, where p_n and p_d were the neutron and deuteron momentum, respectively, θ_d and θ_γ were the azimuthal angles of the deuteron and γ rays, and $\phi_\gamma - \phi_d$ the difference in polar angle. The only unmeasured quantity was the γ-ray energy. Using conservation of energy and momentum allowed a three-constraint fit to the hypothesis that an event was from the reaction $n + p \rightarrow d + \gamma$.[26]

First, the experimenters used p_n, p_d, θ_d, and θ_γ to construct a two-constraint χ^2 fit and events with $\chi^2 < 6$ were plotted as a function of coplanarity θ_c , where sin $\theta_c = \sin\theta_d \sin\phi_\gamma - \phi_d$. As seen in Fig. 33 the $n + p \rightarrow d + \gamma$ events form a narrow peak above a broad background. The background was then extrapolated into the peak region and subtracted to give the number of true $n + p \rightarrow d + \gamma$ events. As a check on this background subtraction a different two-constraint χ^2 fit to the same $n + p \rightarrow d + \gamma$ data was done using all variables except θ_γ and a new θ'_γ predicted from this fit. Events with $\chi^2 < 6$ were plotted against $\Delta\theta = \theta_\gamma - \theta'_\gamma$. Now the $n + p \rightarrow d + \gamma$ events appear as a peak over a "background of *different* shape" (Fig. 34). This was also used to calculate the background "Figure [34] shows the histograms of coplanarity and $\Delta\theta$ for the average case and for the case with the largest background. For a particular bin, the difference in the number of events obtained using the two methods is a measure of the systematic error involved in the background subtraction (Bartlett, Friedberg et al. 1971, pp. 883–884)." This was included in the calculation of the overall uncertainty in the results. Note that unlike the initial experiment, the background subtraction in the new experiment was independent of Monte Carlo calculations.

To obtain the angular distribution the number of events in each energy bin was divided by the measured beam intensity and the efficiency of the apparatus, calculated by a Monte Carlo simulation. The final results along with those from the time reversed reaction $\gamma + d \rightarrow n + p$ are shown in Fig. 35. "Reactions [$n + p \rightarrow d + \gamma$] and

---

[26] Conservation of momentum, a vector quantity, and conservation of energy give rise to four equations with only one unknown. This overconstrains the event by three variables.

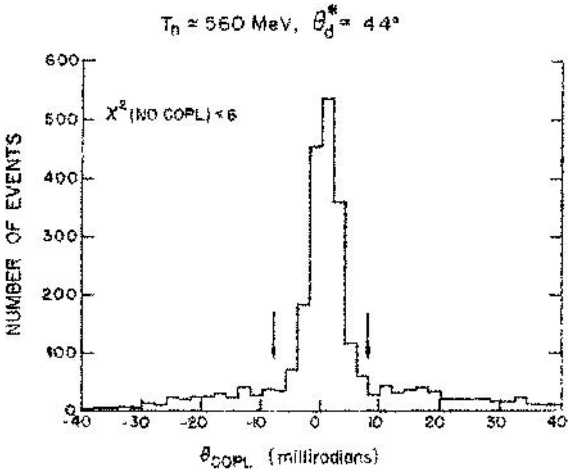

Fig. 33. Number of events as a function of coplanarity angle. From Goldhagen (1973)

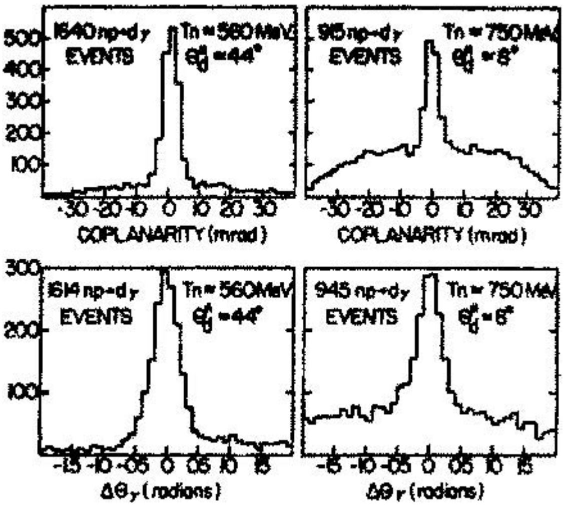

Fig. 34. Coplanarity and $\Delta\theta_\gamma$ for a typical bin (*left*) and for the bin with the largest background (*right*). From Bartlett et al. (1971)

[$\gamma + d \rightarrow n + p$] are in good agreement, as predicted by time-reversal invariance (Bartlett, Friedberg et al. 1971, p. 884)."

Although the Princeton group did not explicitly discuss why their most recent result was more credible than their earlier ones, they did remark, as noted above that they had made several significant improvements to the experiment. These included " (a) the number of $n + p \rightarrow d + \gamma$ events collected is about 10 times larger, (b) angles and momenta are measured about 2.5 times more accurately, and (c) data are taken over a wider angular range and extend to higher neutron energies (Bartlett, Friedberg et al. 1971, p. 882)." Perhaps the most important of these was the increased statistics obtained in the later

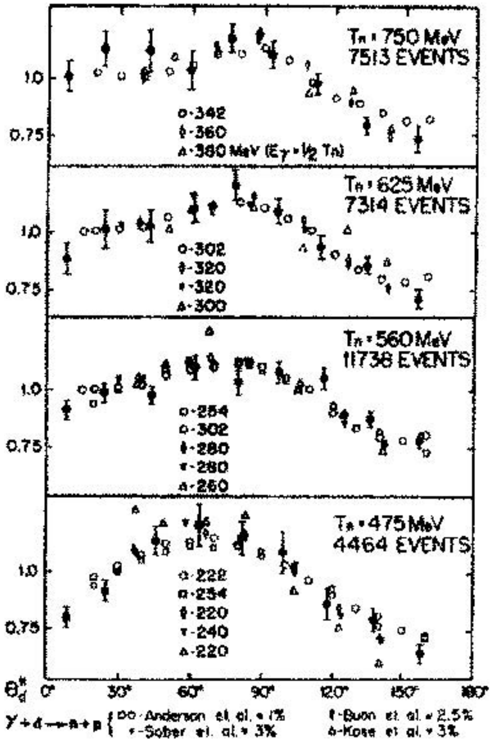

Fig. 35. Angular distributions for n + p → d + γ. From Bartlett et al. (1971)

experiment. In addition, as Goldhagen remarked in his dissertation, the latest experiment had used two independent methods of making the background subtraction due to n + p → d + π° events, which were independent of any Monte Carlo simulation. These were demonstrable improvements. It was clear that the new result was more reliable.

Even before the publication of the last Princeton result, Schrock et al. (1971) had concluded, based on an analysis of all of their data that time-reversal invariance was preserved.[27]

A startling effect had disappeared.

2. Are there magnetic monopoles?

One interesting fact about electromagnetism is that single electric charges, positive and negative, exist, whereas single magnetic charges, magnetic monopoles, do not seem to exist in nature. All known magnetic fields have two poles, north and south.

In 1931 Dirac began a theoretical investigation that led to interesting conclusions about magnetic monopoles, if they existed. Dirac's original intent was to try to provide a reason for the existence of the smallest unit of electric charge, *e*, the charge of the electron.

[27] Recall that their initial result was based on an analysis of twenty percent of their data.

> It will be concerned essentially, not with electrons and protons, but with the reason for the existence of the smallest electric charge. This smallest charge is known to exist experimentally and to have the value e given approximately by $hc/2\pi e^2 = 137$. [h was Planck's constant and c was the speed of light.] The theory of this paper, while it looks at first as though it will give a theoretical value for e, is found when worked out to give a connection between the smallest electric charge and the smallest magnetic pole. It shows, in fact, a symmetry between electricity and magnetism quite foreign to current views (Dirac 1931, p. 62).

Dirac further noted that although physicists believed that quantum mechanics, as usually formulated, was applicable only when there were no isolated magnetic poles, this was not the case.

> The object of the present paper is to show that quantum mechanics does not really preclude the existence of isolated magnetic poles. On the contrary, the present formalism of quantum mechanics, when developed naturally without the imposition of arbitrary restrictions, leads inevitably to wave equations whose only physical interpretation is the motion of an electron in the field of a single pole, This new development requires *no change whatever* in the formalism when expressed in terms of abstract symbols denoting states and observables, but is merely a generalisation of the possibilities of representation of these abstract symbols by wave functions and matrices. Under these circumstances one would be surprised if Nature had made no use of it.
>
> The theory leads to a connection, namely equation (9), between the quantum of magnetic pole and the electric charge (p. 71).

Dirac's equation (9) was $hc/2\pi e\mu_o = 2$, where μ_o was the strength of the magnetic pole. Dirac further stated that "This means that the attractive force between two one-quantum poles of opposite sign is $(137/2)^2 = 46923\frac{1}{4}$ times that between electron and proton. This very large force may perhaps account for why poles of opposite sign have never yet been observed (p. 72)."

In later work Dirac presented a more general theory of the interaction of charged particles and magnetic poles.

> If one supposes that a particle with a single magnetic pole can exist and that it interacts with charged particles, the laws of quantum mechanics lead to the requirement that the electric charge be quantized – all charges must be integral multiples of a unit charge e connected with the pole strength g [the former μ_o] by the formula $eg = hc/4\pi$. Since electric charges are known to be quantized and no reason for this has yet been proposed apart from the existence of magnetic poles, we have a reason for taking magnetic poles seriously. The fact that they have not yet been observed may be ascribed to the large value of the quantum of pole (Dirac 1948, p. 817).

This theoretical work formed the background to searches for magnetic poles, and provided an enabling theory for the experiments by giving an estimate of the strength of the magnetic pole and of the size of the effects that might be observed.

a) Blas Cabrera and the Saint Valentine's day event

One of the interesting searches for magnetic monopoles was conducted by Blas Cabrera and his collaborators in the 1980s and 1990s. In the very first experimental run an event consistent with a monopole was found, but subsequent searches by the group, with improved apparatus, found no similar event. Was the first event an example of very rare magnetic monopoles, or was it an artifact of the experiment?

Cabrera's method of searching for magnetic monopoles was conceptually straightforward although technically difficult, particularly for large area detectors. He used a loop of superconducting wire connected to the superconducting input coil of a SQUID (superconducting quantum interference device) magnetometer. A magnetic monopole passing through a such loop of superconducting wire will produce a change in magnetic flux through the loop of $4\pi g = hc/e$, where g is the magnetic charge of the monopole, h is Planck's constant, c is the speed of light, and e is the charge of the electron. This is twice the flux quantum of superconductivity $\Phi_o = hc/2e$. "Such a detector measures the moving particle's magnetic charge regardless of its velocity, mass, electric charge, or magnetic dipole moment.... In the general case, any trajectory of a magnetic charge g which passes through the ring will result in a flux-quanta change of 2, while one that misses the ring will produce no flux change (Cabrera 1982, p. 1378)."

Cabrera constructed a 20 cm^2 superconducting loop and took data during five experimental runs for a total of 151 days. The loop actually contained four turns so , "The passage of a single Dirac charge through the loop would result in an 8 Φ_o change in flux through the superconducting circuit, comprised of the detection loop and the SQUID input coil (a factor of 2 from $4\pi g = 2\Phi_o$ and of 4 from the turns in the pickup loop) (p. 1379)." The detector was calibrated in three different and independent ways: 1) by measuring the SQUID response to a known current in calibration Helmholtz coils (±4%), 2) by estimating the self-inductance of the superconducting circuit (30%), and 3) by directly observing flux quantization within the superconducting circuit (±10%).[28] The calibrations agreed within the stated uncertainties. Figure [36] shows several intervals of data recording. There are typical small disturbances in the trace due to the daily liquid-nitrogen transfer and weekly liquid helium transfers. These disturbances are far smaller than that observed for the possible monopole event. "A single large event was recorded [Fig. 36b]. It is consistent with the passage of a single Dirac charge within a combined uncertainty of ±5%.... It is the largest event of any kind in the record (p. 1379)." This event was recorded on 14 February 1982 when the laboratory was unoccupied, alowing for the possibility of a transient apparatus malfunction, or even a human intervention.

Other flux changes were recorded. There were 27 events exceeding a threshold of 0.2 Φ_o, after exclusion of known disturbances such as liquid helium and liquid nitrogen transfers. An event was defined as a sharp offset with well-defined stable levels for one hour before and after. No other event was within a factor of four of the signal from the single large event, or that expected for a Dirac monopole.

Cabrera devoted considerable effort to searching for possible spurious detector responses that might have caused the possible monopole signal. Neither line voltage fluctuations or rf interference from the motor brushes of a heat gun caused detectable

---

[28] Later versions of the experimental apparatus would include calibration devices and techniques that would more closely approximate the signal due to a magnetic monopole. As we shall see, the fact that one could artificially produce such a signal cast doubt on the reality of the initially observed event.

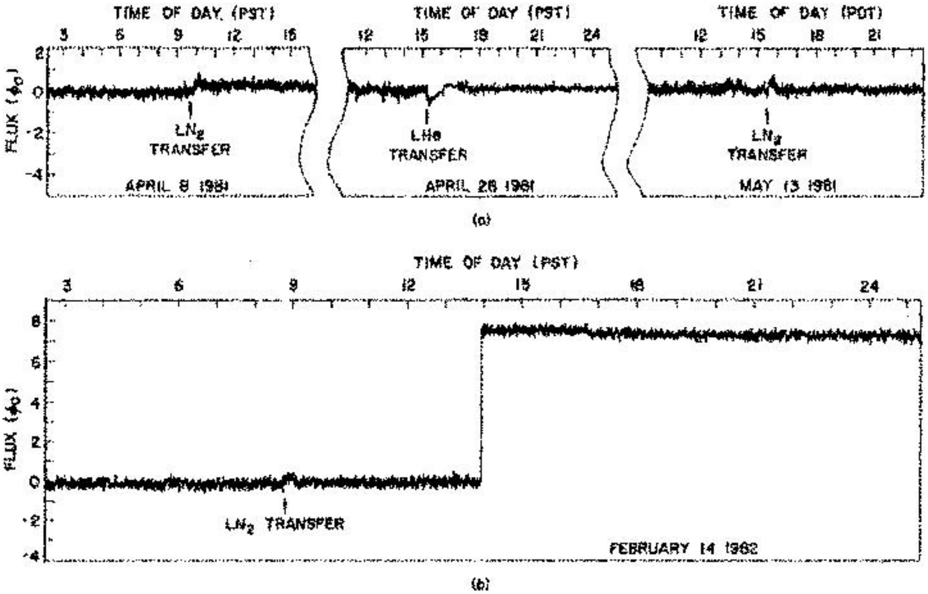

Fig. 36. Data records showing (**a**) typical stability and (**b**) the candidate monopole event. From Cabrera (1982)

effects. No seismic disturbance, which might have shaken the apparatus and produced such a signal was observed on the date of the event. External magnetic fields, ferromagnetic contaminants, critical current quenching of the superconducting loop, and cosmic rays were also eliminated.

There was, however, one possible alternative explanation of the signal that could not be conclusively eliminated.

> Mechanically induced offsets have been intentionally generated and are probably caused by shifts of the four-turn loop-wire geometry which produce inductance changes. Sharp raps with a screw driver handle against the detector assembly cause such offsets. On two occasions out of 25 attempts these have exceeded $6\Phi_o$ (75% of the shift expected from one Dirac charge); however, drifts in the level were seen during the next hour (p. 1380).

Cabrera did think that this was a likely cause of the observed signal, but did not feel that he could completely eliminate it as a possibility.

> A spontaneous and large external mechanical impulse is not seen as a possible cause for the event; however, the evidence presented by this event does not preclude the possibility of a spontaneous internal stress release mechanism. Regardless, to date the experiment has set an upper limit of $6.1 \times 10^{-10}\,\text{cm}^{-2}\,\text{sec}^{-1}\,\text{sr}^{-1}$ for the isotropic distribution of any moving particles with magnetic charge greater than $0.06g$ (pp 1380–81).

In a later comment to Kent Staley, Cabrera remarked, "It was a striking event, because it was exactly the right step size [for a Dirac monopole], but I was not convinced because of the other possible although improbable mechanism (Staley 1999, p. 221). Cabrera made no discovery claim for a magnetic monopole, but it remained a possibility.

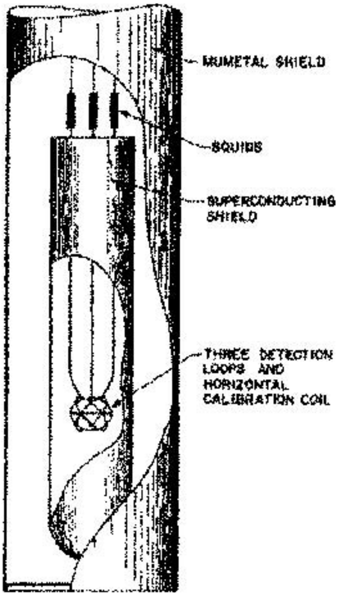

Fig. 37. Schematic diagram of the three-loop detector. From Cabrera et al. (1983)

b) The three-loop detector

Cabrera and his collaborators continued the search for magnetic monopoles with an improved and larger experimental apparatus (Fig. 37) (Cabrera, Taber et al. 1983).[29] The detector consisted of three superconducting loops of two turns each. The area of the new apparatus was 476 cm^2, a loop area of 70.5 cm^2, with a near-miss area of 405 cm^2 . (Monopoles striking the near-miss area would also register in the detector). It was twenty times larger than Cabrera's original detector. The experimenters found that a current of 53.2 nA in the calibration coil induced a supercurrent change of 4 Φ_o, in all three loops. This was the signal expected for a Dirac monopole.

One of the most significant improvements to the experiment was the use of coincidence signals between the loops.

> Operation of the original noncoincidence single-loop detector demonstrated the need for discrimination against spurious events. Along with monitoring other known causes of spurious signals (discussed later), the most reliable technique is to use coincidence detection, having two or more uncoupled detectors that will respond in coincidence to a monopole event but not to a spurious event (Gardner, Cabrera et al. 1991, p. 625).

Unlike the original experiment in which some of the possible causes of spurious signals were checked after the data was taken, in this experiment the experimenters monitored such causes as they went along.

---

[29] A longer an more detailed account of the experiment appears in (Gardner, Cabrera et al. 1991).

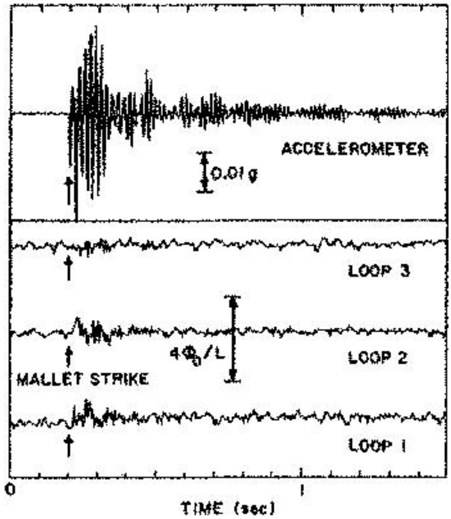

Fig. 38. Detector response to striking the detector with a mallet. From Gardner et al. (1991)

> To guard against spurious signals, additional instruments monitor parameters known to affect the detector. A flux-gate magnetometer monitors the external field variations, a pressure transducer measures the helium gas pressure above the bath in the Dewar, an accelerometer detects any mechanical motion of the apparatus along the vertical axis, and a power-line voltage monitor detects six different line-noise and fault conditions (Cabrera, Taber et al. 1983, p. 1934).

Once again the experimenters checked on the possibility that a mechanical effect could cause a spurious signal. There had been a significant improvement. The signals produced were considerably smaller than they had been in the original experiment and they were detected by the monitoring instruments.

> Superconducting offsets can be generated by tapping on the Dewar with a mallet, but these signals also show up on the accelerometer data as shown in Fig. [38]. We suspect that these offsets are the result of motion of the trapped flux in the SQUID's due to the acoustic wave pulse in the superconductors which make up the SQUID sensors or the motion of the pickup loops in the ambient magnetic field; in any case they are rarely larger than a few tenths of Φ_o (Gardner, Cabrera et al. 1991, p. 628)

The analysis of the data used an algorithm that searched for offsets larger than a threshold of $0.1\Phi_o/L$ and ignored slow changes and brief excursions above threshold that returned quickly to their previous values. They required that 100 points of data (500s) remained within $\pm 0.1\Phi_o/L$ of their initial value before and after any transient signal. "Events from known causes, determined by peripheral instrumentation and logged annotations are not included (Cabrera, Taber et al. 1983, p. 1936)." No events satisfied the double-coincidence requirement. Using these data the experimenters set an upper limit for the uniform flux of magnetic monopoles of $3.7 \times 10^{-11}\ \text{cm}^{-2}\ \text{s}^{-1}\ \text{sr}^{-1}$ at the 90 percent confidence level. "No large or spurious signals were seen, casting no light on the origin of the previously reported candidate. However, these data lower that previous flux

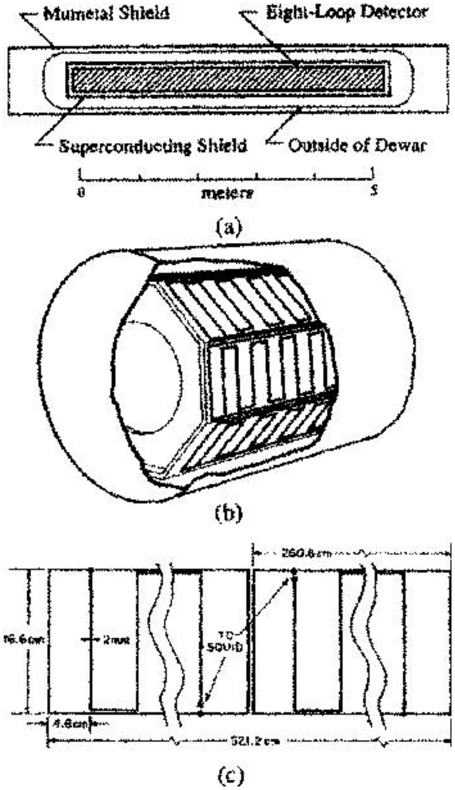

Fig. 39a–c. Schematic representation of the eight-loop detector. From Huber et al. (1990)

limit by a factor of 38, increasing the probability of a spurious cause for that event (Cabrera, Taber et al. 1983, p. 1936)." The experimenters remarked that they planned to continue operating the detector for at least a year and were also designing a larger detector.

c) The eight-loop detector

Cabrera's group continued their search for magnetic monopoles using a further improved eight-loop detector (Fig. 39) (Huber, Cabrera et al. 1990).[30] They noted that a number of groups, including their own, had been searching for monopoles for several years without observing any convincing candidates.

They reported on the first 547 days of operation of their new eight-loop detector, which consisted of eight gradiometer loops each located on the face of an octagonal prism. The total usable area of the detector was 1.1 m^2 , "the largest superconducting detector to date." They required a coincidence signal from two of the loops.

[30] A longer and more detailed account appears in (Huber, Cabrera et al. 1991).

> A feature of this geometry is that a monopole can induce a signal in at most two loops and, for most of the cross section, no fewer than two loops. In contrast, offsets in more than two loops must be the result of electrical or mechanical disturbances and are rejected as monopole candidates (Huber, Cabrera et al. 1990, p. 835).

As is usually the case in experiments, not everything went as planned. The original area of the detector was 1.5 m^2. "Upon cooling the detector, the conducting NbTi ribbon cracked and opened two pickup-loop circuits, causing those loops sections to be unresponsive to flux changes (p. 837)." This reduced the active sensing area to 1.3 m^2.

Once again the experimenters installed checks to guard against spurious signals. These were more extensive than those used in the three-loop experiment.

> We installed additional instrumentation to monitor parameters known to affect detector operation. This instrumentation includes a strain gauge attached to the exterior of the superconducting lead shield (to detect mechanical motion), a pressure transducer (to monitor the helium pressure above the liquid in the Dewar), and a power line monitor (to detect six different fault conditions). During most of our operating period a flux-gate magnetometer has been used to detect changes in the external field. We did not observe a significant correlation, so we have substituted a wideband rf [radiofrequency] voltmeter to detect changes in the local rf environment which can cause offsets in SQUID's. An ultrasonic motion detector monitors laboratory activity. When we perform activities known to disturb detector stability, we set a "veto" switch to prevent generating large numbers of useless computer events and to aid in calculating our live time (p. 837).

Other changes were introduced to reduce the number of spurious signals from known causes. Recall that in the earlier versions of the experiment transfers of liquid helium and liquid nitrogen caused offsets. In this experiment, "A closed-cycle helium liquefier connected to our Dewar eliminates helium transfers and maintains a constant liquid-helium level, so the operation can be extremely stable. Gas-cooled radiation shields eliminate the need for liquid nitrogen (p. 837)."

A new calibration system was installed that more closely approximated the signal expected for a magnetic monopole. These consisted of narrow, toroidally-wound coils. Each calibration coil coupled to two adjacent gradiometers simultaneously. "A current of 0.19μA through the coil... produces a flux equivalent to that of a Dirac monopole (Fig. [40]) (Huber, Cabrera et al. 1991, p. 638)." Note the similarity of the calibration signal to that observed in the early monopole candidate event (Fig. 36). A typical data recording is shown in Fig. 41. The top eight rows of the graphs show the signals from the SQUIDs. Rows S and F contain data from the strain gauge and the flux-gate magnetometer. Rows P and H record the pressure monitor and helium level sensor. The last rows contain data from a cosmic-ray channel (unused), power line monitor, ultrasonic motion detector, and event veto. Part (a) of the figure shows a single loop event (Loop 5) that was detected at 9:41 on 17 July 1987. It is not a monopole candidate because the signal was observed only in a single loop. Note, however, that the SQUID signal is correlated with signals in the strain gauge, the magnetometer, and the motion detector (Rows S, F, and U) and would have been rejected on these grounds as a monopole candidate.

The group detected 43 single channel events, which did not correlate with any disturbances in the monitors. Based on that frequency of occurrence, they calculated that accidental double coincidences would be detected approximately once in 800 years.

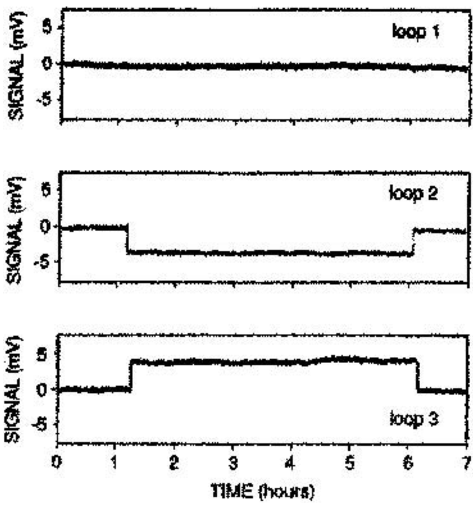

Fig. 40. Calibration signal equivalent to one Dirac charge in inductors Number 2 and 3. From Huber at al. (1991)

> The incidence of double-coincident offsets is much higher than this estimate. Four have already been observed Table [2]; however, the magnitudes are inconsistent with a Dirac charge, and such effects always occur in adjacent detectors. Since adjacent-pane events contribute only ∼0.152 of the total sensing area it is extremely unlikely that we would observe four such events without observing events of any other type. The probability is approximately $(0.152)^4$ or 0.0005. A more likely cause is mutual rf interference between SQUID's coupled through adjacent pickup coils. All four events were recorded in the first 221 days of operation, and none have occurred since the rf excitation frequency for each SQUID was adjusted to avoid mutual resonances. Nevertheless, we have discarded the area contributed by adjacent-panel events, reducing our quoted sensing area to 1.1 m^2 (Huber, Cabrera et al. 1991, p. 648),

The experimenters stated that

> In conclusion, these data set an upper limit of 7.2×10^{-13} cm^{-2} s^{-1} sr^{-1} at 90% confidence level on any uniform flux of magnetic monopoles passing through the Earth's surface at any velocity. This limit is a factor of 2000 below the flux suggested by the single-candidate event seen with the prototype detector. Based on this large factor and based on the noncoincident nature of the prototype detector, we conclude that the entire data set from the prototype detector which contains the single event should be discarded (Huber, Cabrera et al. 1990, p. 838).

The demonstrated improvements in the experimental apparatus and analysis, including the coincidence requirements and the monitoring instruments, along with the failure to reproduce the original effect had persuaded both the experimental group, and the physics community, that the original event was spurious and that magnetic monopoles had not been observed.

In an interesting epilogue, Cabrera, in a later talk at the University of Colorado (I was present), stated that the single monopole candidate was made even less plausible

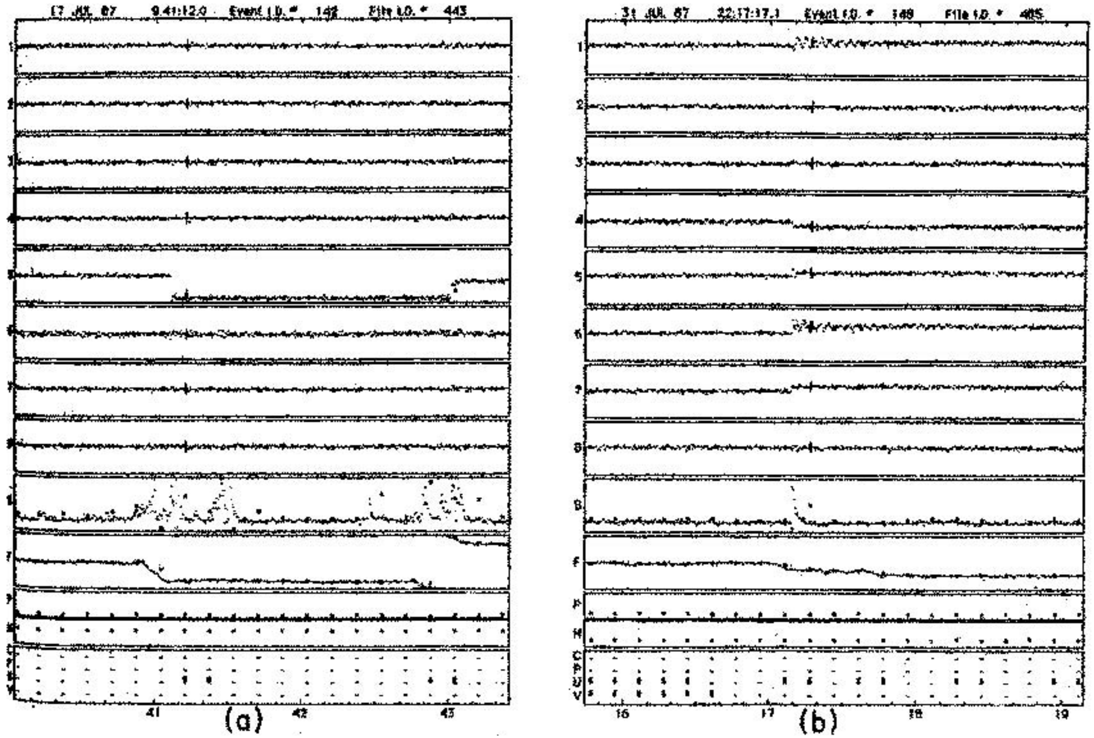

Fig. 41a, b. Summary of an event data file for the eight-loop detector. The top eight rows contain filtered data from the SQUID's. Rows S and F contain filtered data for the strain gauge and the flux-gate magnetometer, respectively. Rows P and H contain unfiltered low-frequency data from the pressure monitor and helium level sensor, respectively. The lasts row contains digital data from a cosmic-ray channel(unused, power line monitor, ultrasonic motion detector, and event veto. (**a**) The event shown occurred at approximately 9:41. Note the coincidence between signals in the SQUID with signals in the strain gauge, magnetometer, and motion detector. (**b**) Event data demonstrating the resolution of SQUID detectors. Note the oscillations in the SQUID output resulting from a sharp disturbance recorded in the strain gauge. From Huber et al. (1991)

Table 2. (From Huber at al. 1991)

Time	Date	Channel	Magnitude (pA)
5:50	24 July 1987	5	342
		6	58
4:47	27 July 1987	4	48
		5	139
18:11	18 August 1987	1	65
		8	149
21:26	1 October 1987	5	395
		6	73

by the fact that the experimenters could artificially generate a similar signal. If they could do it, so could an intruder. Recall that the laboratory was unoccupied when that event occurred. Although it was very unlikely, it was possible that human intervention might have caused that signal. The third version of the experimental apparatus, which contained a motion sensor, eliminated that possibility in the last run, and would have done so, had it been present, in the prototype experiment.[31]

D. Conclusion

The examples presented above by no means exhaust the types of strategies used to argue for the correctness of a null result, but they do provide us with a variety of those strategies. We have seen that the results themselves, the use of surrogate signals, and the observation of background effects may all argue for the correctness of a null result.

In the case of the Michelson-Morley experiment there were, in fact, no questions raised at the time as to whether the 1887 apparatus could measure the expected effect. The measurements themselves, which recorded fringe shifts of the order of three or four micrometer screw divisions, showed clearly that the apparatus was able to detect the twenty-division effect predicted for the motion of the earth relative to the ether. Later evidence provided by Miller confirmed that judgment. Michelson and Morley recognized that the 1881 apparatus, which was approximately ten times less sensitive than the 1887 apparatus, was not capable of measuring the predicted effect. In that case, the experimental uncertainty was approximately the same size as the predicted effect.

The episode of the 17-keV neutrino emphasizes the importance of including analysis procedures, which transform data into an experimental result, in examining those results. The questions of the appropriateness of the shape-correction factor used in the magnetic spectrometer experiments and of the width of the energy range used in the search for the presence of the heavy neutrino were clearly crucial. Physicists questioned whether the shape corrections or a too-wide energy range could mask or mimic the presence of

[31] Although searches have continued, no convincing monopole candidate has been reported.

a heavy neutrino. In the experiment of Mortara et al. we saw the use of an extremely clever surrogate signal, namely the superposition of two different β-decay energy spectra obtained from ^{35}S mixed with one percent ^{14}C. This combined spectrum had a kink because the two spectra had different endpoint energies. The observation of that kink demonstrated that the experimenters would have detected a kink of the appropriate size in the pure ^{35}S spectrum, had one been present. The experimenters also found that their fit to the superposed spectra gave the same value for the amount of ^{14}C present as that obtained by the direct measurement of the ^{14}C intensity when the source was being prepared. This provided further support for their null result. The Tokyo group had so much data and their analysis procedure was so detailed that everyone agreed that a kink could not have been missed (See discussion at the end of Sect. B.1).

A more straightforward surrogate signal was used in the search for two neutrinos. In this case the question was whether the apparatus could detect 400 MeV electrons. The experimenters calibrated their apparatus by exposing their chambers to a beam of 400 MeV electrons and showed that their signal was due to muons, not electrons, and that their background events were not caused by electrons.

In the disagreement between Weber and his critics concerning the existence of gravity waves, questions were raised concerning the adequacy of the surrogate signal and of the analysis procedures used. Under ordinary circumstances, Weber's failure to detect the calibration pulses, combined with the critics's ability to detect them would have been decisive. Because this was a case in which a new type of experimental apparatus was being used to detect a hitherto unobserved phenomenon, it wasn't. Weber's argument that the calibration pulses were too short to be an adequate surrogate signal for gravity waves was taken seriously. The critics responded by explicitly changing their analysis procedures to search for longer pulses and by applying Weber's preferred analysis algorithm to their own data. In both instances no signal was seen. These arguments against Weber's positive result argued for the null results obtained by his critics. Not only were other arguments against Weber's result provided (see (Franklin 1994) for details), but there was an overwhelming preponderance of evidence from the critics's results. Gravity waves had not been observed.

The use of background signal to demonstrate that an apparatus could detect the phenomenon of interest was shown in the experiment that searched for the decay $\mu \rightarrow e+\gamma$. In this episode events from the background decay $\mu \rightarrow e + \nu + \bar{\nu} + \gamma$, which contained both an electron and a γ ray, were observed, giving confidence that the apparatus would have detected the same particles from the decay $\mu \rightarrow e + \gamma$.

Another strategy involved the use of computer simulation. In the E791 experiment at Fermilab, the experimenters were searching for the decays $D^+ \rightarrow \pi^+\mu^+\mu^-$ and $D^+ \rightarrow \pi^+e^+e^-$. Using straightforward surrogate signals the experimenters showed that their apparatus would detect muons and electrons. The question was whether it would detect the decays. They injected computer simulated events for both decays into their data and demonstrated that their analysis program would detect such decays.

In the final two episodes discussed, the experiments each initially observed a rather startling phenomenon, whereas later attempts at replication failed to detect the same phenomena. Here, the arguments were that the later experiments were better, more credible, and that the phenomena in question did not exist. In the experiments that investigated time-reversal symmetry the experimenters showed that a considerable portion, but not all, of their initially observed effect was due to an apparatus malfunction. When this was taken into account the effect was considerably reduced. Their second experiment included clear improvements such as increased statistics, improved measurements of angles and momenta, and the measurement, rather than the calculation, of backgrounds. Such improvements made their later null result more believable.

Blas Cabrera and his collaborators, using a single-loop detector, initially found a single event consistent with a magnetic monopole. Their later experiments, first with a three-loop detector and then with an eight-loop detector, failed to find any similar events. The later experiments required a coincidence signal between two loops, to reduce any spurious signal. They also included instrumentation, which was absent in the first experiment, to monitor parameters known to affect their detector, and which might mimic a monopole signal. In the eight-loop experiment such instrumentation was more extensive than in the three-loop experiment. In addition, the experimenters used a calibration system that more closely approximated the signal expected for a monopole, to demonstrate that their apparatus would have detected a monopole signal had it been present. The analysis procedures were also improved. These demonstrated improvements in the experimental apparatus itself, in the monitoring instrumentation, and in the analysis procedures argued strongly that the null results of the later experiments were more credible than that of the first experiment and that the monopole signal originally observed was spurious.

These examples illustrate some of the strategies used by experimenters to argue that when nothing is observed it is because the phenomenon is not present, not because the apparatus cannot detect it. In these cases no signal is a real signal.

Acknowledgement. I would like to thank David Bartlett for providing both original documentation and valuable discussions of the time-reversal experiments.

References

Aitala, E. M., S. Amato, J. C. Anjos, et al. 1996. "Search for the Flavor-Changing Neutral Current Decays $D^+ \to \pi^+\mu^+\mu^-$ and $D^+ \to \pi^+e^+e^-$." *Physical Review Lettters* 76: 364–367.

Altzitzoglou, T., F. Calaprice, M. Dewey, et al. 1985. "Experimental Search for a Heavy Neutrino in the Beta Spectrum of ^{35}S." *Physical Review Letters* 55: 799–802.

Anderson, R. L., R. Prepost and B. H. Wiik 1969. "Photodisintegration of the Deuteron from 220–340 MeV." *Physical Review Letters* 22: 651–654.

Apalikov, A. M., S. D. Boris, A. I. Golutvin, et al. 1985. "Search for Heavy Neutrinos in β Decay." *JETP Letters* 42: 289–293.

Bartlett, D., S. Devons and A. M. Sachs 1962. "Search for the Decay Mode: $\mu \to e+\gamma$." *Physical Review Letters* 8: 120–123.

Bartlett, D. F., C. E. Friedberg, P. E. Goldhagen, et al. 1971. "New Results on $n + p \to d + \gamma$, and Time-Reversal Invariance." *Physical Review Letters* 27: 881–885.

Bartlett, D. F., C. E. Friedberg, K. Goulianos, et al. 1969. "$n + p \to d + \gamma$ and Time-Reversal Invariance." *Physical Review Letters* 23: 893–897.

Berley, D., J. Lee and M. Bardon 1959. "Upper Limit for the Decay Mode $\mu \to e+\gamma$." *Physical Review Letters* 2: 357–359.

Bonvicini, G. 1993. "Statistical Issues in the 17-keV Neutrino Experiments." *Zeitschrift fur Physik A* 345: 97–117.

Buchwald, J. Z. 1988. "The Michelson Experiment in the Light of Electromagnetic Theory Before 1900.".

Cabrera, B. 1982. "First Results from a Superconductive Detector for Moving Magnetic Monopoles." *Physical Review Letters* 48: 1378–1381.

Cabrera, B., M. Taber, R. Gardner, et al. 1983. "Upper Limit on Flux of Cosmic-Ray Monopoles Obtained with a Three-Loop Superconductive Detector." *Physical Review Letters* 51: 1933–1936.

Caso, C., G. Conforto, A. Gurtu, et al. 1998. "Review of Particle Physics." *European Physical Journal* 3: 1–794.

Christenson, J. H., J. W. Cronin, V. L. Fitch, et al. 1964. "Evidence for the 2π Decay of the K_2^o Meson." *Physical Review Letters* 13: 138–140.

Collins, H. 1985. *Changing Order: Replication and Induction in Scientific Practice*. London: Sage Publications.

Danby, G., J.-M. Gaillard, K. Goulianos, et al. 1962. "Observation of High-Energy Neutrino Interactions and the Existence of Two Kinds of Neutrinos." *Physical Review Letters* 9: 36–44.

Datar, V. M., C. V. K. Baba, S. K. Bhattacherjee, et al. 1985. "Search for a heavy neutrino in the β-decay of ^{35}S." *Nature* 318: 547–548.

Dirac, P. A. M. 1931. "Quantified Singularieties in the Electromagnetic Field." *Proceedings of the Royal Society* (*London*) 133: 60–72.

Dirac, P. A. M. 1948. "The Theory of Magnetic Poles." *Physical Review* 74: 817–830.

Frankel, S., J. Halpern, L. Holloway, et al. 1962. "New Limit on the $e + \gamma$ Decay Mode of the Muon." *Physical Review Letters* 8: 123–125.

Franklin, A. 1986. *The Neglect of Experiment*. Cambridge: Cambridge University Press.

Franklin, A. 1994. "How to Avoid the Experimenters' Regress." *Studies in History and Philosophy of Modern Physics* 25: 97–121.

Franklin, A. 1995. "The Appearance and Disappearance of the 17-keV Neutrino." *Reviews of Modern Physics* 67: 457–490.

Franklin, A. 1997. "Calibration." *Perspectives on Science* 5: 31–80.

Franklin, A. 1998. "Selectivity and the Production of Experimental Results." *Archive for the History of Exact Sciences* 53: 399–485.

Franklin, A. 2000a. *Are There Really Neutrinos? An Evidential History*. Cambridge, MA: Perseus Books.

Friedberg, C. E. (1969). $N + p \rightarrow d + \gamma$: A Test of Time Reversal Invariance. Physics. Princeton, NJ, Princeton University : 157.

Friedman, J. L. and V. L. Telegdi 1957. "Nuclear Emulsion Evidence for Parity Nonconservation in the Decay Chain pi - mu-e." *Physical Review* 105: 1681–1682.

Gardner, R. D., B. Cabrera, M. E. Huber, et al. 1991. "Search for cosmic-ray monopoles using a three-loop superconductive detector." *Physical Review D* 44: 622–635.

Garwin, R. L., L. M. Lederman and M. Weinrich 1957. "Observation of the Failure of Conservation of Parity and Charge Conjugation in Meson Decays: The Magnetic Moment of the Free Muon." *Physical Review* 105: 1415–1417.

Goldhagen, P. E. (1973). A Measurement of the Angular Distribution of $n + p \rightarrow d + \gamma$ as a Test of Time-Reversal Invariance. Physics. Princeton, N.J., Princeton University : 140.

Handschy, M. 1982. "Re-examination of the 1887 Michelson-Morley Experiment." *American Journal of Physics* 50: 987–990.

Hetherington, D. W., R. L. Graham, M. A. Lone, et al. 1987. "Upper Limits on the Mixing of Heavy Neutrinos in the Beta Decay of ^{63}Ni." *Physical Review C* 36: 1504–1513.

Hicks, W. M. 1902. "On the Michelson-Morley Experiment relating to the Drift of the Aether." *Philosophical Magazine* 3: 9–43.

Hime, A. 1992. "Pursuing the 17 keV Neutrino." *Modern Physics Letters A* 7: 1301–1314.
Hime, A. and N. A. Jelley 1991. "New Evidence for the 17 keV Neutrino." *Physics Letters* 257B: 441–449.
Hime, A. and J. J. Simpson 1989. "Evidence of the 17-keV Neutrino in the β Spectrum of ^{3}H." *Physical Review D* 39: 1837–1850.
Huber, M. E., B. Cabrera, M. A. Taber, et al. 1990. "Limit on the Flux of Cosmic-Ray Monoples from Operations of an Eight-Loop Superconducting Detector." *Physical Review Letters* 64: 835–838.
Huber, M. E., B. Cabrera, M. A. Taber, et al. 1991. "Search for a flux of cosmic-ray monopoles with an eight-channel superconducting detector." *Physical Review D* 44: 636–660.
Kawakami, H., S. Kato, T. Ohshima, et al. 1992. "High Sensitivity Search for a 17 keV Neutrino. Negative Indication with an Upper Limit of 0.095%." *Physics Letters* 287B: 45–50.
Lodge, O. J. 1893. "Aberration Problems: A Discussion concerning the Motion of the Ether near the Earth, and concerning the Connexion between Ether and Gross Matter; with some New Experiments." *Philosophical Transactions of the Royal Society of London* 184: 727–804.
Markey, H. and F. Boehm 1985. "Search for Admixture of Heavy Neutrinos with Masses between 5 and 55 keV." *Physical Review C* 32: 2215–2216.
Michelson, A. A. 1881. "The Relative Motion of the Earth and the Luminiferous Ether." *American Journal of Science* 22: 120–129.
Michelson, A. A. and E. W. Morley 1887. "On the Relative Motion of the Earth and the Luminiferous Ether." *American Journal of Science* 34: 333–345.
Miller, D. C. 1933. "The Ether-Drift Experiment and the Determination of the Absolute Motion of the Earth." *Reviews of Modern Physics* 5: 203–242.
Morrison, D. 1992. "Review of 17 keV Neutrino Experiments." *Joint International Lepton-Photon Symposium and Europhysics Conference on High Energy Physics*. S. Hegarty, K. Potter and E. Quercigh. Geneva, Switzerland: World Scientific. 1: 599–605.
Mortara, J. L., I. Ahmad, K. P. Coulter, et al. 1993. "Evidence Against a 17 keV Neutrino from ^{35}S Beta Decay." *Physical Review Letters* 70: 394–397.
Norman, E. B., B. Sur, K. T. Lesko, et al. 1991. "Evidence for the Emission of a Massive Neutrino in Nuclear Beta Decay." *Journal of Physics G* 17: S291–S299.
Ohi, T., M. Nakajima, H. Tamura, et al. 1985. "Search for Heavy Neutrinos in the Beta Decay of ^{35}S. Evidence Against the 17 keV Heavy Neutrino." *Physics Letters* 160B: 322–324.
Ohshima, T. 1993. "0.073% (95% CL) Upper Limit on 17 keV Neutrino Admixture." *XXVI International Conference on High Energy Physics*. J. R. Sanford. Dallas: American Institute of Physics. 1: 1128–1135.
Ohshima, T., H. Sakamoto, T. Sato, et al. 1993. "No 17 keV Neutrino: Admixture $<$ 0.073% (95% C.L.)." *Physical Review D* 47: 4840–4856.
Pontecorvo, B. 1960. "Electron and Muon Neutrinos." *Soviet Physics JETP* 10: 1236–1240.
Radcliffe, T., M. Chen, D. Imel, et al. 1992. "*New Limits on the 17 keV Neutrino.*" Progress in Atomic Physics, Neutrinos and Gravitation: Proceedings of the XXVIIth Rencontre de Moriond, Les Arcs, France, Editions Frontieres.
Schrock, B. L., J. F. Detoeuf, R. P. Haddock, et al. 1969. "*An Experimental Test of Time-Reversal Invariance in the Photodisintgration of the Deuteron.*" High Energy Physics and Nuclear Structure, New York, NY, Penum Press.
Schrock, B. L., R. P. Haddock, J. A. Helland, et al. 1971. "Experimental Test of Time-Reversal Invariance in the Reaction $n + p \to \gamma + d$." *Physical Review Letters* 28: 1659–1662.
Schwartz, M. 1960. "Feasibility of Using High-Energy Neutrinos to Study the Weak Interactions." *Physical Review Letters* 4: 306–307.
Schwartz, M. 1972. "One Researcher's Personal Account." *Adventures in Experimental Physics* α: 82–85.

Schwarzschild, B. 1991. "Four of Five New Experiments Claim Evidence for 17-keV Neutrinos." *Physics Today* 44: 17–19.

Shankland, R. S. 1964. "Michelson-Morley Experiment." *American Journal of Physics* 32: 16–35.

Shankland, R. S., S. W. McCuskey, F. C. Leone, et al. 1955. "New Analysis of the Interferomter Observations of Dayton C. Miller." *Reviews of Modern Physics* 27: 167–178.

Shaviv, G. and J. Rosen, Eds. 1975. *General Relativity and Gravitation: Proceedings of the Seventh International Conference (GR7), Tel-Aviv University, June 23–28, 1974*. New York: John Wiley.

Simpson, J. J. 1985. "Evidence of Heavy-Neutrino Emission in Beta Decay." *Physical Review Letters* 54: 1891–1893.

Simpson, J. J. 1986a. "Is There Evidence for a 17 keV Neutrino in the ^{35}S β Spectrum? The Case of Ohi et al." *Physics Letters* 174B: 113–114.

Simpson, J. J. 1986b. "Evidence for a 17-keV Neutrino in ^{3}H and ^{35}S β Spectra." *'86 Massive Neutrinos in Astrophysics and in Particle Physics: Proceedings of the Sixth Moriond Workshop*. O. Fackler and J. Tran Thanh Van. Gif sur Yvette: Editions Frontieres : 565–577.

Simpson, J. J. and A. Hime 1989. "Evidence of the 17-keV Neutrino in the β Spectrum of ^{35}S." *Physical Review D* 39: 1825–1836.

Staley, K. 1999. "Golden Events and Statistics: What's Wrong with Galison's Image/Logic Distinction?" *Perspectives on Science* 7: 196–230.

Sur, B., E. B. Norman, K. T. Lesko, et al. 1991. "Evidence for the Emission of a 17-keV Neutrino in the β Decay of ^{14}C." *Physical Review Letters* 66: 2444–2447.

Swenson, L. 1972. *The Ethereal Aether*. Austin: University of Texas Press.

Weber, J., M. Lee, D. J. Gretz, et al. 1973. "New Gravitational Radiation Experiments." *Physical Review Letters* 31: 779–783.

Wu, C. S., E. Ambler, R. W. Hayward, et al. 1957. "Experimental Test of Parity Nonconservation in Beta Decay." *Physical Review* 105: 1413–1415.

Zlimen, I., A. Ljubicic, S. Kaucic, et al. 1991. "Evidence for a 17-keV Neutrino." *Physical Review Letters* 67: 560–563.

Department of Physics
University of Colorado
Campus Box 390
Boulder, CO 80309
USA
Allan.Franklin@colorado.edu

(Received September 15, 2003)
Published online December 2, 2003