Arch. Hist. Exact Sci. 53 (1998) 399–485. © Springer-Verlag 1998

Selectivity and the Production of Experimental Results

"Any fool can take data. It's taking good data that counts."
E. Commins (private communication)

ALLAN FRANKLIN

Communicated by J. Z. BUCHWALD

I. Introduction

Experimenters never use all of their data in producing a result. Data may be excluded for many legitimate reasons.[1] Certainly no one would think of using data obtained when the experimental apparatus was not working properly. Even when the apparatus is working properly problems may arise when only selected portions of the data, i.e., "good" data, are used to obtain a result. Selection criteria, usually referred to as "cuts," are applied to either the data themselves or to the analysis procedures[2] and are designed to maximize the desired signal and to eliminate or minimize background that might mask or mimic the desired effect.[3] One might worry that the experimental result is an artifact produced by the cuts, and not a valid result.[4] A further worry may arise if the effect of

[1] I will not deal here with the selectivity that is built in to an experimental apparatus. In the experiment that measured the K^+_{e2} branching ratio, discussed below, the experimenters required that the decay particle give a signal in a Cerenkov counter set to detect positrons. This was designed to exclude events resulting from decay modes such as $K^+_{\mu 2,}K^+_{\pi 2}$, and $K^+_{\mu 3}$ that did not include a positron.

[2] One should distinguish between experimental data and an experimental result. They are usually different. (See discussion in note 12). What I mean by "analysis procedures" are those processes that transform data into an experimental result. These processes may involve computer analysis and simulation, making cuts on the data, and other procedures. This distinction will be illustrated in the episodes discussed below.

[3] Selectivity may also be applied to the conditions under which the experiment is performed.

[4] By valid, I mean that the experimental result has been argued for in the correct way, using epistemological strategies. These strategies include: (1) experimental checks and calibration; (2) the reproduction of artifacts that are known in advance to be present; (3) intervention, in which the experimenter manipulates the object under observation; (4) independent confirmation using independent experiments; (5) elimination of plausible sources of error and alternative explanations of the result (the Sherlock Holmes strategy); (6) the use of the results themselves to argue for their validity; (7) the use of an independently well-corroborated theory of the phenomena to explain

the cuts on the experimental result is known in advance. Is the experimenter tuning the cuts to produce a desired outcome?

This is not a purely philosophical or methodological issue. It is an issue in the concrete practice of science. Consider, for example, the recent controversy concerning the possible existence of low-mass electron-positron states, or particles, produced in high-energy heavy ion-atom collisions. Early experiments found evidence that positrons with discrete energies, or positron lines, were produced in such collisions. Later experiments also detected an electron produced in coincidence with the positron and found peaks in the sum-energy spectrum ($E_{e+} + E_{e-}$), suggestive evidence for the existence of low-mass electron-positron states. The fact that both effects, the positron lines and the sum-energy peaks, appeared for various pairs of projectile and target nuclei, as well as in three different detectors, gave the results credibility. This led others to further investigate the phenomenon. These later experiments produced conflicting claims. Some experiments found evidence for such states whereas others did not. Part of the difficulty was that the results were not reproducible. Even when later experiments were done under quite similar conditions to the original ones the effects did not always appear, and when they did the states had different energies from those observed initially. The failure to reproduce the effects cast doubt on the original results themselves. It also raised the question of how similar must conditions be for one experiment to be considered a replication of another experiment.

A recent analysis by one of the experimental groups, EPOS II, has suggested that the results might be an artifact of applying selection criteria. Applying certain cuts to one half of the data and tuning those cuts to produce a maximal effect, the group found evidence for a peak in the sum-energy spectrum. Applying identical cuts to the other half of the data showed no such evidence. This failure cast doubt on the result and also on the statistical procedures used. The effect originally seen was significant, amounting to a five standard-deviation effect. If such a large effect is real, and not an artifact caused by the cuts, it is very improbable that it could be made to disappear by altering the selection criteria.

Because selection cuts are invariably present in modern physics experiments the question of how one argues that an observed effect is not an artifact produced by the cuts is often of crucial importance in establishing the validity of experimental results. Experimenters use several strategies to argue that their result is not an artifact. These will be illustrated in the episodes discussed below, which will illuminate the different ways in which selectivity is used in producing experimental results, and how arguments for the validity of those results are made despite the application of these cuts.

The first of these strategies involves robustness. Experimenters will vary the selection criteria over reasonable limits and observe whether the result is stable under such variations. If it is, then the result is taken to be real. In the experiment to measure the K^+_{e2} branching ratio, the fraction of all K^+ mesons that decay into a positron plus a neutrino, the experimenters varied the values of a range cut, as well as their track matching criteria, and found that the branching ratio remained constant. Similarly, in the case of early attempts to detect gravity waves, one of the selection criteria involved a choice of analysis

---

the results; (8) the use of an apparatus based on a well-corroborated theory; and (9) the use of statistical arguments. See Franklin (1986, Ch. 6 and 1990, Ch. 6) for details.

algorithm. One group of experimenters used both proposed algorithms and found no change in their result. In the case of the 17-keV neutrino there was a choice to be made concerning the energy range to be used in the analysis of the data. Several experiments used both a wide-and a narrow-energy range and showed the result to be constant.

If the result is, on the other hand, sensitive to variations in the selection criteria, then this suggests, although it does not prove, that the result is an artifact. There are indeed cases, such as resonant scattering of light, in which the result is highly sensitive to the experimental conditions and also to the particular selection criteria. This type of sensitivity will be an important issue in the episode of possible low-mass e^+e^- states discussed below, in which the observed effects seemed to be very sensitive to the experimental conditions.

Robustness may also be provided by a sequence of experiments, rather than by variations in a single experiment. Thus, in the case of Millikan's measurement of the charge on the electron the variations in the selection criteria he applied to both his data and to his analysis procedures were provided by subsequent experiments. These experiments used different experimental techniques and had different backgrounds and selection criteria. The fact that Millikan's and the subsequent experiments all gave the same value for the charge of the electron argued that the result was not an artifact produced by the cuts. It is highly improbable, if the result is an artifact, that the same result would be obtained under very different circumstances. The replicability, or apparent replicability, of results also played an important role in the episode concerning possible low-mass e^+e^- states.

Sometimes one may be able to use a surrogate signal to demonstrate that the cuts do not mask the presence of an effect.[5] This was the case in both the search for gravity waves and for a 17-keV neutrino. Conversely, one may cast doubt on a result by showing that it could be produced by the application of cuts (as in the case of the just mentioned low-mass electron-positron states). This may be done either by the analysis of actual data or by computer simulation. This will be illustrated several times in the following discussion, and will be crucial in the discussion of low-mass electron-positron states.

One may also be able to argue that the application of cuts, while reducing background, could not produce the effect observed. In the case of the measurement of the K^+_{e2} branching ratio there was no possible way in which the application of cuts to the range of particles, to their decay time, or to their track-matching could produce positrons that would mimic those expected from K^+_{e2} decay.

This study examines five historical cases which involve arguments concerning the proper existence of an effect. The longest of these examines in detail the history of low-mass electron-positron states, from their initial report in the early 1980s until the present. To introduce the reader to the issues I will begin with four less complex examples taken from the history of modern physics. The first was an experiment designed to measure the K^+_{e2} branching ratio. The branching ratio is quite small (10^{-5}) and the desired events were masked by large numbers of events due to other, more prevalent, decay modes. Cuts were applied to preferentially reduce this background while preserving a large, and

---

[5] This is the issue of calibration. Some critics have questioned the use of calibration to validate a result. For a fuller discussion see Franklin (1997).

known, fraction of the K^+_{e2} events. I will show how the experimenters used the cuts to produce the experimental result and how they argued for its validity.

A somewhat more complex example is provided by Joseph Weber's claim that he had observed gravity waves, whereas six other experiments did not find his claimed effect. In this episode there were no arguments about what constituted good data. Both Weber and his critics used the same type of experimental apparatus, namely a large-mass gravity-wave antenna known as a Weber bar. The question was whether the data were being analyzed correctly. Weber and his critics used different data-analysis algorithms. The linear algorithm used by Weber's critics was sensitive to changes in either the amplitude or the phase of the signal. The non-linear algorithm preferred by Weber was sensitive only to changes in amplitude. There was a suggestion that Weber based his choice of algorithm on the fact that it gave a larger gravity-wave signal in his experiment. A second question concerned the pulse-height threshold which was used to determine whether a signal was in fact present. Weber's use of varying thresholds raised the issue of whether he was tuning his threshold cut to maximize, or even to create, evidence for his positive signal. I will discuss how these issues were decided and the discord resolved. This episode will also emphasize and illustrate the importance of paying close attention to analysis procedures in order to understand arguments concerning the validity of an experimental result.

The third historical episode concerns Millikan's famed measurement of *e*, the charge on the electron. Examination of Millikan's notebooks has shown not only that Millikan excluded data, but that he also engaged in selective calculational procedures. We also know that Millikan had a clear expectation of the value of *e*, based on his earlier work. One might ask whether he used selectivity to obtain the answer he expected. I will also discuss the effect of what we might call Millikan's "cosmetic surgery" on his experimental result. Finally, in a fourth case, I will look at the role played by analysis cuts in the episode that decided against the existence of the 17-keV neutrino, a proposed new particle.

II. Measurement of the K^+_{e2} branching ratio[6]

Perhaps the simplest and most straightforward strategy used to argue for the correctness of a result when selection criteria are used in its production is to vary the values of those cuts. If the result remains constant under that variation then this argues that it is not an artifact If it is a real effect then reasonable changes in the cuts used should not affect the result. If the result changes when the values are changed, then this suggests that the result is an artifact. This strategy is clearly illustrated in an experiment designed to measure the K^+_{e2} branching ratio, the fraction of all K^+ mesons that decayed into a positron and an electron neutrino ($K^+ \rightarrow e^+ + \nu_e$) (Bowen et al. 1967).

The motivation for this experiment was that it would be a stringent test, using strangeness-changing decays, of the then generally accepted V-A (Vector minus Axial vector) theory of weak interactions. In 1934 Enrico Fermi had proposed a theory of weak interactions that involved the four fermions: the neutron, the proton, the electron,

[6] For a more detailed discussion of this experiment see Franklin (1990, pp. 118–131).

and the neutrino. Pauli had previously shown that there were only five possible relativistically invariant mathematical forms of such an interaction: scalar (S), vector (V), pseudoscalar (P), tensor (T), and axial vector (A). (Combinations of these forms were also relativistically invariant). During the next twenty five years considerable effort was devoted to determining the correct mathematical form of the weak interaction. By 1957 it had been established that the correct form was V-A. At the time of our experiment the V-A theory had strong experimental support, although it had not been severely tested in strangeness-changing decays.[7]

The theoretical predictions of the K^+_{e2} branching ratio were explicit. If the interaction was pure axial-vector (A) the predicted ratio of K^+_{e2} to $K^+_{\mu 2}$ decays was 2.6×10^{-5}, corresponding to a branching ratio of 1.6×10^{-5}. Pure pseudoscalar (P) coupling, on the other hand, predicted a K^+_{e2} to $K^+_{\mu 2}$ ratio of 1.02. If even only a small amount of pseudoscalar interaction were present, along with the dominant axial vector interaction, the K^+_{e2} branching ratio would be much larger. For example, adding only one part in a thousand of pseudoscalar interaction to the axial vector interaction would increase the expected branching ratio by a factor of four. Thus, even a rough measurement of the K^+_{e2} branching ratio would be a stringent test for the presence of any pseudoscalar interaction in the decay, and of the V-A theory in general. The best previous measurement of the $K^+_{e2}/K^+_{\mu 2}$ ratio had set an upper limit of 2.6×10^{-3}, a factor of 100 larger than that predicted by V-A theory.

A. The experimental apparatus

The experimental apparatus is shown in Fig. 1. The incoming beam was positively charged, unseparated, and momentum selected. It consisted primarily of pions and protons and also contained small numbers of kaons, muons, and positrons. The desired kaons were separated from the more numerous pions and protons by range in matter[8] and by time of flight and were stopped in Counter C_3, the stopping region. (For details see Bowen et al. 1967).[9]

Decay particles which left the stopping region at about 90° traversed a set of six thin-plate optical spark chambers located in a magnetic field. Decay particles were detected by a coincidence telescope C_5C_6. To restrict the particles detected to kaon decays a 21 ns gate, triggered by a stopped kaon signal was used. (The mean life of the K^+ meson is

---

[7] Strangeness was a property of elementary particles. It was conserved in strong and electromagnetic interactions, but not in weak interactions. For details see Franklin (1986, Chapter 3).

[8] How far a charged particle will travel in matter before stopping depends on its velocity. For a momentum selected beam kaons, which are lighter than protons and heaviers than pions, will have a higher velocity than protons and a lower velocity than pions. Thus, the kaons will travel through more material than protons and have a shorter range than pions. Enough copper was placed in the beam in front of the kaon stopping region to remove all of the protons, whereas the pions passed through the stopping region and were vetoed.

[9] Muons, which have the very long lifetime of 2×10^{-6} s, compared with the K^+ lifetime, and positrons, which are stable, would not produce any background in this experiment.

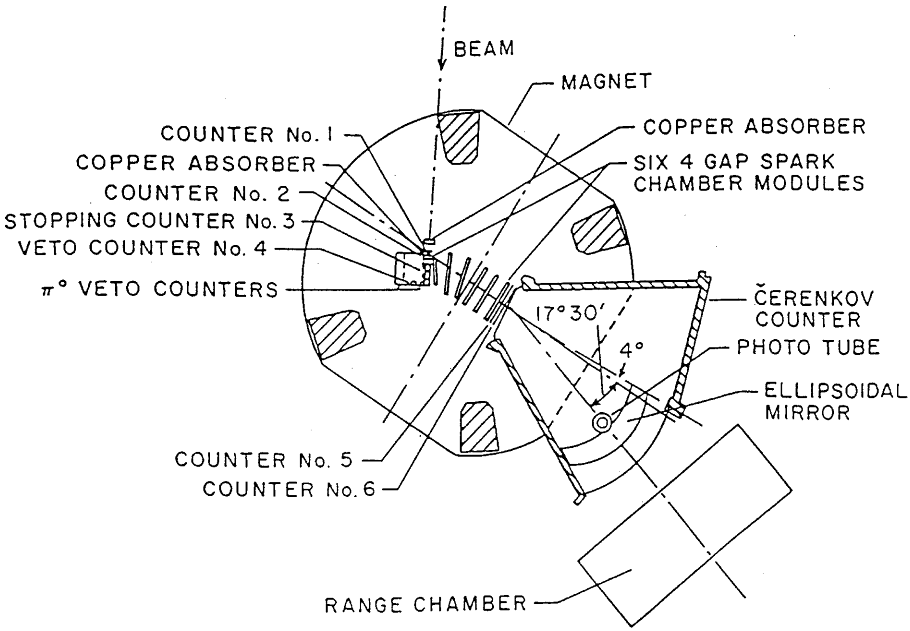

Fig. 1. Experimental apparatus for the measurement of the K^+_{e2} branching ratio. From Bowen et al. (1967)

12.4 ns). This avoided any effects from other synchrotron bunches (the internal beam of the Princeton-Pennsylvania accelerator used in this experiment had bunches separated by 34 ns). Pulses from counters C_3 and C_5 were displayed as oscilloscope traces and photographed, so that the time between them, which measured the time interval between the K^+ stop and its decay, could be measured for each event. If the decay particles were indeed due to kaon decay then the distribution of the time between the stopping kaon signal and the decay signal should match the kaon lifetime. It did. A small prompt peak was observed at short decay times due to kaon decays in flight. These were completely eliminated during the analysis procedure by a requirement that the decay time of the event be greater than 2.75 ns after the stop of the K^+ meson.

This apparatus incorporated both a Cerenkov counter, to identify positrons, as well as a range chamber to help eliminate background from other decay modes. The Cerenkov counter had a measured efficiency of greater than 99% for positrons and had a measured efficiency of 0.38% for other particles of comparable momentum. The thick-plate range chamber was placed behind the Cerenkov counter which permitted measurement of the position of particles emerging from the counter as well as a measurement of their total range.

B. Calculation of background

Because the expected K^+_{e2} branching ratio was so small, 1.6×10^{-5}, the experimenters needed an accurate calculation of expected background that might mask or mimic K^+_{e2} events to determine whether or not the experiment was feasible. If the background was

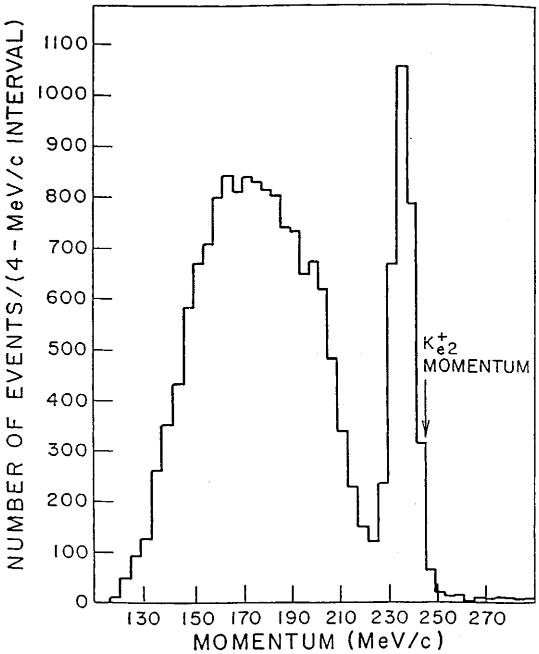

Fig. 2. Momentum distribution for all K^+ decay events obtained with the Cerenkov counter in the triggering logic. From Bowen et al. (1967)

too large the measurement could not be successfully carried out. This calculation also illustrated the need for selection criteria in the analysis of the data.

The positron from K^+_{e2} decay has a momentum of 246.9 MeV/c in the kaon center of mass system. This is higher than the momentum of any other direct product of K^+ decay. The closest competitor is the muon from $K^+_{\mu 2}$ decay, which has a momentum of 235.6 MeV/c. The principal sources of high-momentum positrons that might mimic K^+_{e2} decay are

1. $K^+ \rightarrow e^+ + \pi^o + \nu_e$, K^+_{e3} decay, with a maximum positron momentum of 228 MeV/c and a branching ratio of approximately 5%.
2. $K^+ \rightarrow \mu^+ + \nu_\mu$, followed by $\mu^+ \rightarrow e^+ + \nu_e + \nu_\mu$, with a maximum momentum of 246.9 MeV/c, the same as that for K^+_{e2} decay, and a branching ratio of approximately 1.2×10^{-4} per foot of muon path. This decay rate per foot was considerably larger than the total expected K^+_{e2} decay rate. If this source of background could not be eliminated then the experiment could not be done.

Using the momentum resolution measured in this experiment of 1.9%,[10] and the known K^+_{e3} decay rate and momentum spectrum (both of which had been measured in

[10] The momentum resolution was measured using the muon peak from $K^+_{\mu 2}$ decays in coincidence with noise pulses in the Cerenkov counter shown in Fig. 2.

previous group experiments) one could calculate that the number of K^+_{e3} events expected in the K^+_{e2} decay region, 242–252 MeV/c, was less than 5% of the expected K^+_{e2} decay rate. If the K^+ decayed in flight then the momentum of the positron from K^+_{e3} decay could be higher than 228 MeV/c. This possible source of background was completely eliminated when "prompt" events were removed from the sample as discussed later.

The background due to K^+ decay into a muon, followed by muon decay into a positron, was also calculated. This involved a detailed computation which included the decay rate, the momentum and angular distribution of the decay positrons relative to the muons, the extrapolation of the decay particle trajectory from the momentum chambers into the range chamber, and the momentum and angular resolution in the thin-plate momentum chambers. This last factor, which was quite important, was known from previous experiments performed by the group. The group calculated that the number of decay positrons with laboratory angles of less than 10° with respect to the muon flight path and with momentum greater than 225 MeV/c was less than 5% of the expected K^+_{e2} decay rate. The limits on momentum and angle were well within their previously determined experimental resolution. The total background rate was calculated to be approximately 15% of the expected K^+_{e2} decay rate in the K^+_{e2} momentum region (242–252 MeV/c). This strongly suggested that the experiment was feasible.[11]

C. Cuts

Figure 2 shows the momentum distribution of 16,965 events obtained with the Cerenkov counter in the triggering logic.[12] These events satisfied the following criteria: 1) the track came from the stopping region, C_3; 2) the track passed through both the front and rear windows of the Cerenkov counter; and 3) a track traversing at least three plates was seen in the range chamber. This is the haystack from which the needle of a few K^+_{e2} events was to be found. The momentum for K^+_{e2} decay is shown. It is clear that if the K^+_{e2} events are present they are rather well hidden.

The large peak at 236 MeV/c and the smaller peak at 205 MeV/c are due to accidental coincidences between accelerator produced background in the Cerenkov counter and muons and pions from $K^+_{\mu 2}$ and $K^+_{\pi 2}$ decay, respectively. If the K^+_{e2} events were to be found then the background due to $K^+_{\mu 2}$ events had to be reduced. Because the momentum of these muons is known, this peak was used both to calibrate the momentum scale and to measure the momentum resolution.

The experimenters applied a set of criteria to eliminate unwanted background events while preserving a reasonable and known fraction of the K^+_{e2} events. The first criterion applied was that of range, the path length in the range chamber before the particle stopped or produced an interaction. Muons lose energy only by ionization loss and thus have

[11] There were other sources of background due to experimental effects which were not *a priori* calculable. These are discussed below.

[12] These events are not raw data. The data for the experiment consisted of spark-chamber and oscilloscope photographs along with scaler readings. The sparks were fitted to a trajectory, which combined with the known magnetic field allow a determination of the decay particle's momentum. These momenta are plotted in Fig. 2.

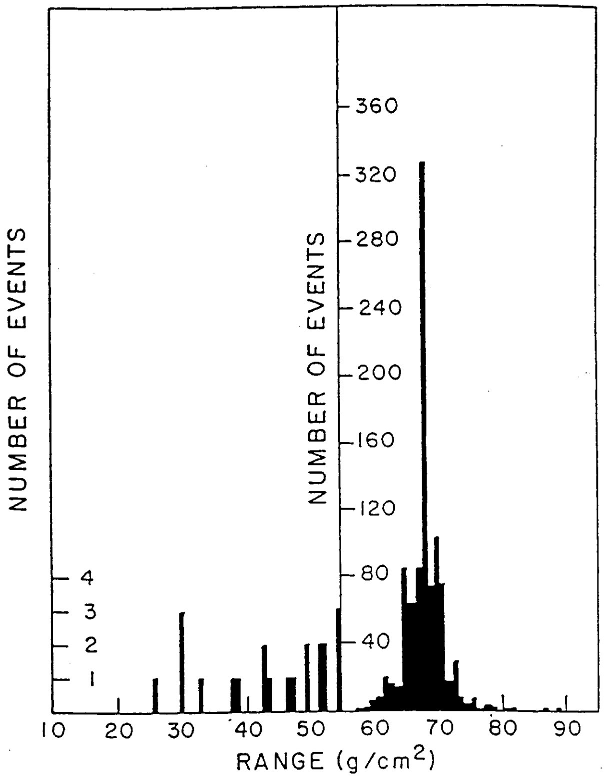

Fig. 3. Range spectrum for muons from $K^+_{\mu 2}$ decay. From Bowen et al. (1967)

a well-defined range in matter. The muons from $K^+_{\mu 2}$ had a mean measured range of 67 g/cm^2, with a straggle of about 4 g/cm^2 (Fig. 3). The 1% of such events with a range less than 45 g/cm^2 is too large to be accounted for by range straggling, and was due to the occasional failure of the range chamber to operate properly. These apparatus failures gave rise to a background of about 15% of the expected K^+_{e2} rate. The experimenters measured the range distribution for positrons with momenta between 212 MeV/c and 227 MeV/c (Fig. 4). These positrons differ from K^+_{e2} positrons by only 10% in momentum, and were expected to behave quite similarly. Positrons do not have a well-defined range because they lose energy by several different processes, some of which involve large energy loss, and the distribution of ranges is approximately constant from about 15 g/cm^2 to 70 g/cm^2. The percentage of positrons with range less than a given value is shown in Table 1. If events are required to have a range less than that of the muon from $K^+_{\mu 2}$ decay, this will serve to minimize the background due to those events, while preserving a large, and known, fraction of the high energy positrons. A selection cut was made at 45 g/cm^2. The limits on this cut were varied within reasonable limits ($\pm$5 g/cm^2) and it was found that the final result was robust against such changes. The effect of applying this selection criterion to the data is shown in Fig. 5. The haystack had gotten smaller.

As discussed earlier, a major source of background was decay of the kaon into a muon, followed by the decay of the muon into a positron. Most of these positrons are emitted at

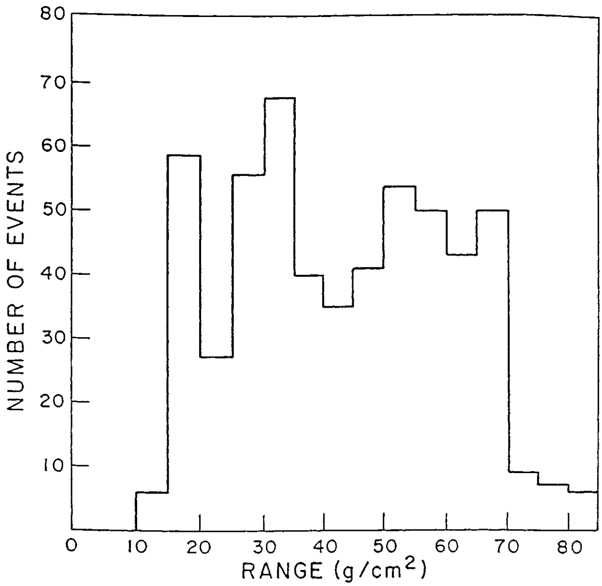

Fig. 4. Range spectrum for positrons from K^+_{e3} decay with momentum between 212 and 227 MeV/c. From Bowen et al. (1967)

Table 1. Range distribution of positrons from K^+_{e3} decay in the momentum interval 212-227 Me V/c. From (Bowen et al. 1967)

Range (g/cm^2)	% of events with Smaller range
40	45.6 ± 2.1
45	51.9 ± 2.1
50	55.9 ± 2.1
55	68.8 ± 1.9

large angles to the muon path. If the decay occurred in the momentum chambers, it would have been detected by a kink in the track (see discussion in note 12). Decays occurring between the end of the momentum chambers and the end of the Cerenkov counter, a very long distance, could not be seen. Because of the usually large decay angle such decays could be detected by comparing the measured position of the particle when it entered the range chamber with the position predicted by extrapolating the momentum chamber track. If a decay had occurred, then the difference between the two positions would be large. Even for decay angles as small as 5° the extrapolated momentum chamber track will not match the position of the range chamber track. Therefore, a cut on the difference between the measured and extrapolated positions, D_x and D_y, the x and y differences, respectively, could be used to eliminate muon decays in flight. If decays with decay angles greater than 5° were eliminated then the background due to muon decays in flight would be reduced to 5% of the expected K^+_{e2} rate. The experimental distributions for D_x and D_y are shown in Figs. 6 and 7 for positrons resulting from K^+_{e3} decay (these positrons do not decay). The width of these distributions and the accuracy of the comparison was

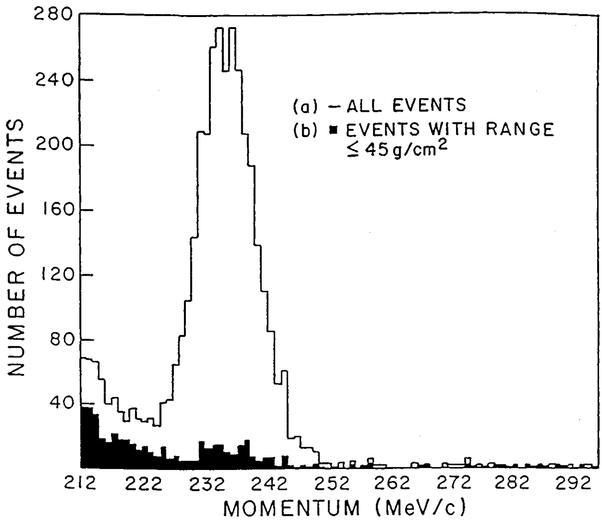

Fig. 5. Momentum spectrum of particles with momentum greater than 212 MeV/c: (**a**) all events (**b**) events with range $\leq$ 45 g/cm^2. From Bowen et al. (1967)

limited by multiple scattering in the momentum chambers and by the uncertainty in extrapolating the particle trajectory through the fringing field of the magnet. The full width at half maximum of the distributions is 16 cm for D_x and 13 cm for D_y. Fiducial areas −6 cm to +10 cm for D_x and −7 to +6 cm for D_y were chosen, as indicated in Figs. 6 and 7. Table 2 shows the variation in the number of accepted particles in the momentum regions of interest as the fiducial areas were varied. The ratios are constant, within the calculated statistical uncertainty, showing that the branching ratio was robust under these changes in the cuts. Such a track-match ing criterion was applied. In addition, because these decays occurred beyond the momentum chambers they would have a measured momentum equal to that of muons from $K^+_{\mu 2}$ decay, 236 MeV/c. These events would be further reduced by momentum cuts in the final analysis of the data.[13] The effects of this track-matching cut for events with range $\leq$45 g/cm^2 and momentum $\geq$212 MeV/c are shown in Fig. 8.[14] The selection criteria served to preferentially reduce the events in the $K^+_{\mu 2}$ region relative to the events in the K^+_{e2} region.

[13] The accepted momentum region for K^+_{e2} decays was 242–252 MeV/c, centered on the K^+_{e2} decay momentum of 246.9 MeV/c.

[14] The track-matching criterion was actually somewhat more complex. The widths of the distributions shown in Figures 6 and 7 are approximately twice that expected for multiple scattering and is due to errors in extrapolating the long particle path through the fringe field of the magnet. This accounts for the greater width of the horizontal (D_x) distribution. The center of the distributions also varied slightly with the momentum of the particles and the fiducial areas were adjusted accordingly. To guard against muon decays in flight and large multiple scattering in the momentum chamber, both of which would result in an incorrect value for the momentum of the particle, the experimenters required that the measured track give a good fit to the polynomial used to approximate the trajectory in the momentum chamber and also that the track have a relatively

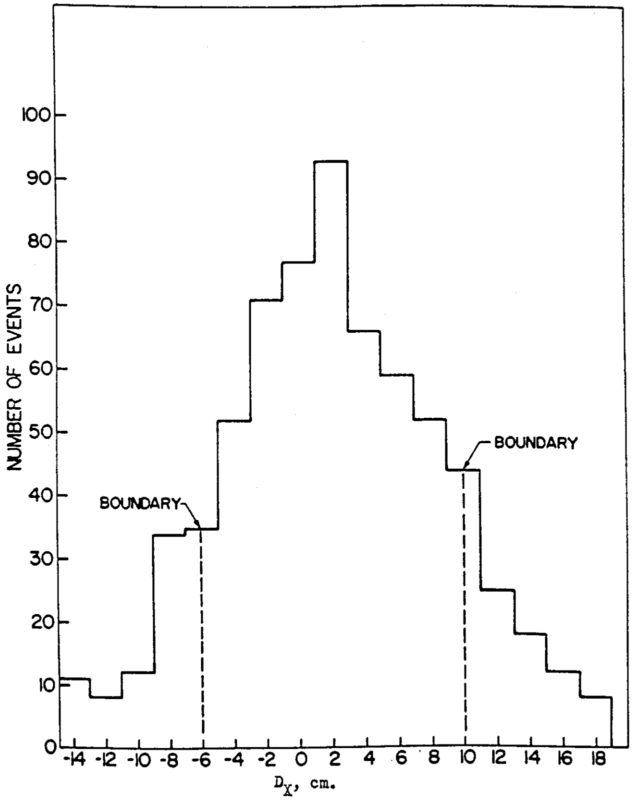

Fig. 6. Experimental distribution for D_X, the difference in x-position between the track extrapolated from the momentum chambers and the measured track in the range chamber, for positrons from K^+_{e3} decay with momenta between 207 and 227 MeV/c. These positrons do not decay in flight. From Bowen et al. (1967)

There was one further major source of background. This was due to decays in flight of the K^+ meson. If the kaon decayed in flight then the momentum of the decay particle could be increased, leading to possible simulation of K^+_{e2} decays. Examination of the distribution of time intervals between the stopped kaon and the decay positron revealed the presence of a small peak due to such decays in flight. The peak had a base width of 2 ns. A cut was made removing all events with a time interval of less than 2.75 ns. This eliminated all of the decays in flight. The effect of this selection criterion is shown in Fig. 9. This cut preferentially reduced the number of events in the K^+_{e2} region, indicating that decays in flight were indeed a source of simulated K^+_{e2} events. This cut was not varied because it was intended solely to eliminate decays in flight. It was clear from the

---

constant momentum along its entire path. This last criterion eliminated very few events, none of which were in the K^+_{e2} momentum region.

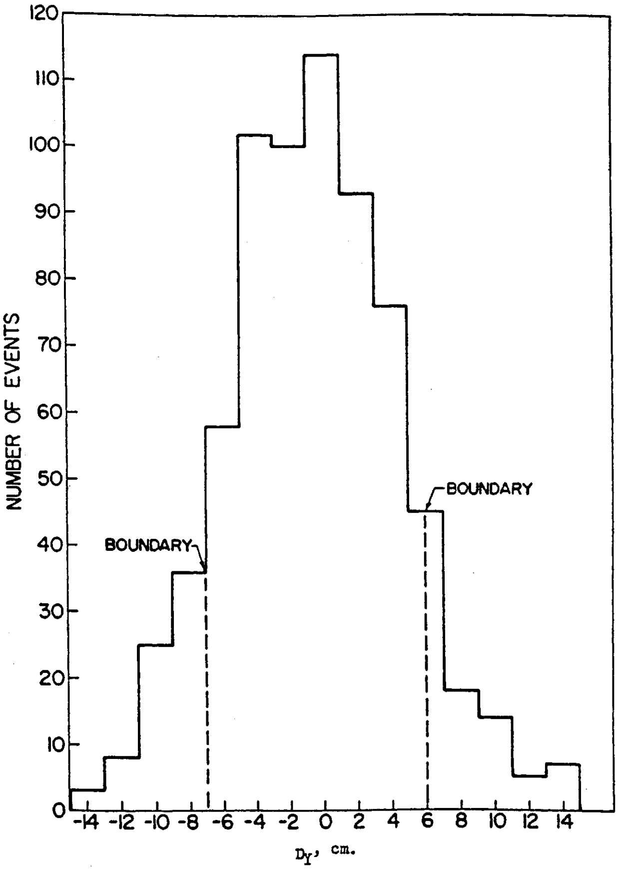

Fig. 7. Experimental distribution for D_y From Bowen et al. (1967)

decay-time distribution, which showed a small prompt peak that was 2 ns wide, that the cut at 2.75 ns was sufficient. In addition, it affected both the number of K^+_{e2} events as well as the events used to normalize the branching ratio equally. Changing the cut would not affect the branching ratio. The final number of K^+_{e2} candidates are the events in the momentum region 242–252 MeV/c in Fig. 9.[15]

The number of events in the K^+_{e2} region had to be corrected for various experimental effects to determine the final number of K^+_{e2} events. These included: 1) Flat background, estimated from the momentum region above 252 MeV/c, where no events due to real K^+ decays were expected (or even possible); 2) Events due to $K^+_{\mu 2}$ decay, which spilled over into the K^+_{e2} region because of the finite momentum resolution of the apparatus; and 3) K^+_{e2} events lost because of the finite momentum region selected for the decay and the

[15] This momentum restriction is also a selection cut. It was chosen so that the lower limit of the selected region was one standard deviation from both the K^+_{e2} and the $K^+_{\mu 2}$ momenta, 247 MeV/c and 236 MeV/c, respectively.

Table 2. Effect of the track-matching criterion for events with range less than 45 g/cm^2. The number of events in the momentum regions of interest does not change significantly for small variations of the fiducial area. From (Bowen et al. 1967)

Momentum Region (MeV/c)	None	Accepted Intervals	Fiducial Area: Accepted Intervals increased by 2 cm	Fiducial Area: Accepted Intervals decreased by 2 cm
$P \leq 212$	525	214	240	184
K^+_{e3} (212–228)	297	161	177	143
$K^+_{\mu 2}$ (231–241)	134	28	35	20
K^+_{e2} (242–252)	33	13	13	11
$P \geq 252$	36	5	6	4

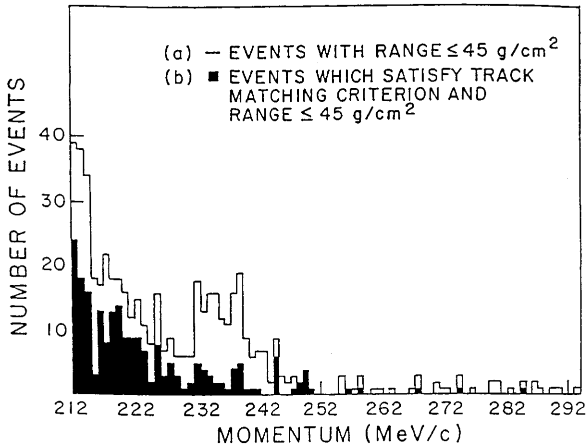

Fig. 8. Momentum spectrum of particles with momentum greater than 212 MeV/c and range $\leq$ 45 g/cm^2: (**a**) all events (**b**) events satisfying track-matching criterion. From Bowen et al. (1967)

finite momentum resolution. A final total of $6^{+5.2}_{-3.7}$ events were attributed to K^+_{e2} decay after these corrections.[16]

The branching ratio, the rate compared to all K^+ decays, was calculated by normalizing the K^+_{e2} events to known K^+ decay rates by two different methods. The first used the upper end of the K^+_{e3} spectrum, the region from 212 MeV/c–228 MeV/c in Fig. 9, which had been subjected to the same selection criteria as the K^+_{e2} events. To estimate

[16] The aysmmetrical uncertainty in the number of events is caused by the use of Poisson statistics, which is preferred for small numbers of events. The more usual, symmetric uncertainty (equal to $\sqrt{n}$ events) applies only to Gaussian statistics and large numbers of events.

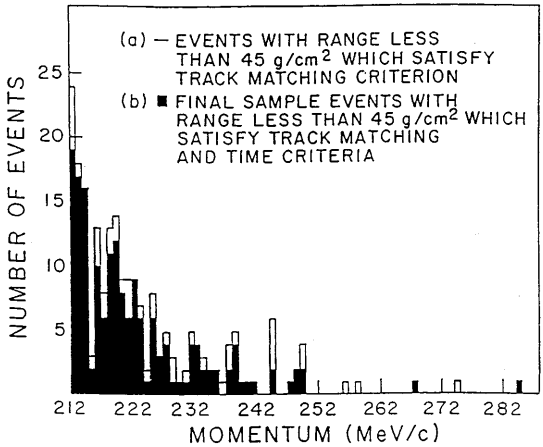

Fig. 9. Momentum spectrum of particles with momentum greater than 212 MeV/c, range $\leq$ 45 g/cm^2, and satisfying the track matching criterion: (**a**) all events (**b**) events with K^+ decay time $\geq$ 2.75 ns. From Bowen et al. (1967)

the total number of K^+_{e3} events the experimenters needed to know the shape of the K^+_{e3} decay spectrum. This had, in fact, been measured by the group in previous experiments. The second method used the total sample of 16,965 K^+ decays given in Fig. 2. (Note that these events had not had the selection criteria applied to them). The results for the branching ratio, using the two different methods, were R = $(2.0^{+1.8}_{-1.2}$ $\times$ 10^{-5} and R = $(2.2^{+1.9}_{-1.4})$ $\times$ 10^{-5}, respectively. The two different methods, which had very different selection criteria, agreed and the final result given was their average, R = $(2.1^{+1.8}_{-1.3})$ $\times$ 10^{-5} in agreement with the theoretical prediction of 1.6×10^{-5}. The fact that the branching ratio obtained was robust under these two very different normalization methods, which had very different dependences on the selection criteria, and also under reasonable changes in the selection criteria argued that the cuts did not affect the final result and therefore for the correctness of that result.

III. Early attempts to detect gravity waves[17]

Beginning in the late 1960s attempts were made to detect gravitational radiation (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

[17] For more details of this episode see Franklin (1994).

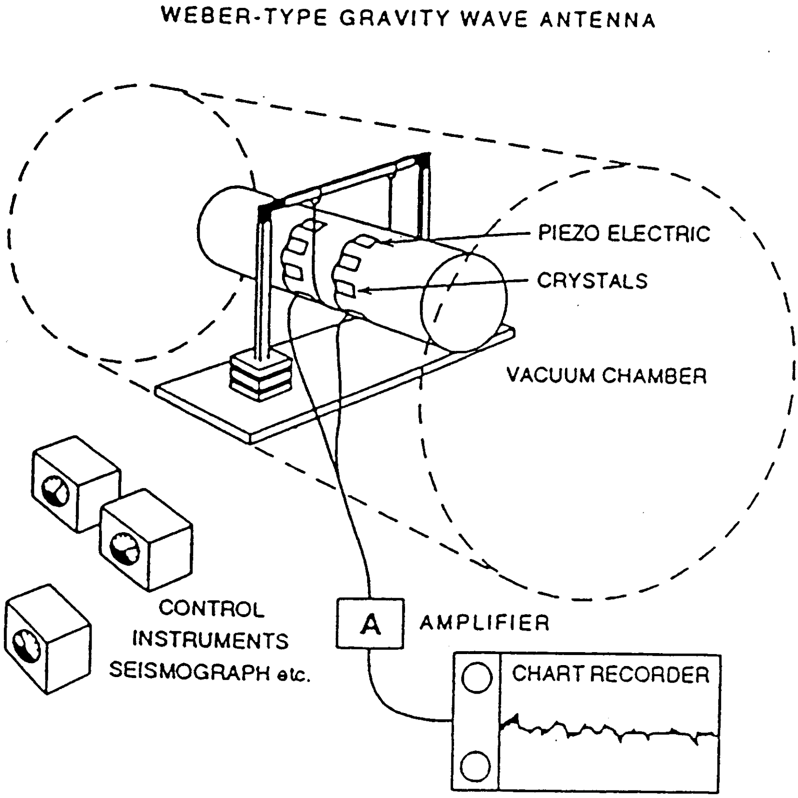

Fig. 10. A Weber-type gravity wave detector. From Collins (1985)

electromagnetic force, a large mass must be accelerated to produce a detectable gravity wave signal.[18] 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,[19] or antenna, which was supposed to oscillate when struck by gravitational radiation (Fig. 10). The oscillation was to be detected by observing the 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 even of those that are to be expected

[18] The ratio of the gravitational force between the electron and the proton in the hydrogen atom compared to the electrical force between them is 4.38×10^{-40}, a small number indeed.

[19] This device is often referred to as a Weber bar.

above the threshold from thermal noise).[20] 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 et al. 1973). He claimed that above-threshold peaks had been observed simultaneously in two detectors separated by one thousand 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, 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.

The reasons offered by different scientists for their rejection of Weber's claims were varied, and not all of the scientists engaged in the pursuit agreed about their relative significance. During the period 1972–1975 it was discovered that Weber had made several serious errors in his analysis. His computer program for analyzing the data contained an error, and his statistical analysis of residual peaks and background was questioned and thought to be inadequate. Weber's claim to have found coincidences between his detector and another, distant detector was rejected because the tapes used to provide the coincidences were actually recorded more than four hours apart. Weber had found a positive result where even he would not expect one. Others cited the failure of Weber's signal to noise ratio to improve, despite his "improvements" to his apparatus. In addition, the sidereal correlation previously observed disappeared when more data was taken, suggesting that it was a statistical fluctuation in the data. (See discussion below in Sect. VI). Critics also argued that Weber's apparatus, as described in his published work, could not produce the signal he reported. Perhaps most important were the uniformly negative results obtained by six other groups.[21]

Two of the reasons for the rejection of Weber's result by the physics community involve questions concerning selection criteria.[22] The first of these – the issue of calibration together with Weber's analysis procedure – is not so much a selection criterion applied to the data, but rather a choice of the analysis procedure used. The problem of

---

[20] 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.

[21] Note here the repetition of experiments of an important physical quantity. This will also be important in the next section.

[22] In this discussion I have relied, primarily, on a panel discussion on gravitational waves that took place at the Seventh International Conference on General Relativity and Gravitation (GR7), TelAviv University, June 23–28, 1974. The panel included Weber and three of his critics, Tyson, Kafka, and Drever, and included not only papers presented by the four scientists, but also included discussion, criticism, and questions. It includes almost all of the important and relevant arguments concerning the discordant results. The proceedings were published as Shaviv and Rosen (1975). Unless otherwise indicated all quotation in this section are from Shaviv and Rosen (1975). I shall give the author and the page numbers in the text.

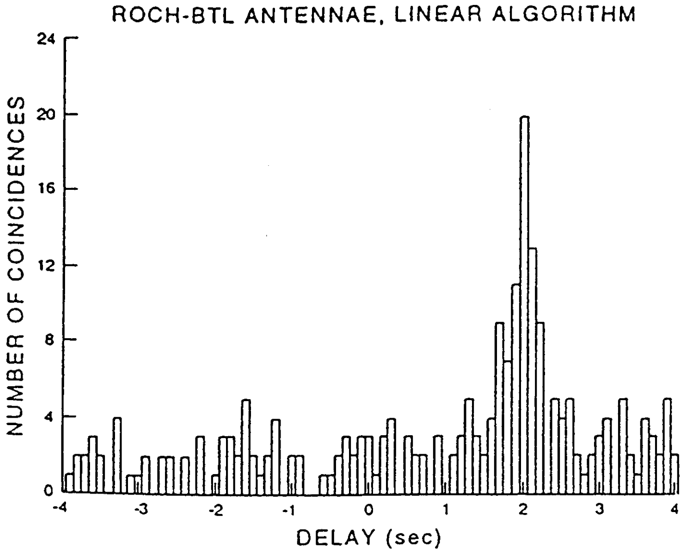

Fig. 11. A plot showing the calibration pulses for the Rochester-Bell Laboratory collaboration. The peak due to the calibration pulses is clearly seen. From Shaviv and Rosen (1975)

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. This difficulty informs the discussion of which was the best analysis procedure to use.

The non-linear, 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. Weber admitted, however, that the linear algorithm, preferred by his critics, was more efficient, by a factor of twenty, at detecting calibration pulses. 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. Similar results on the superiority of the linear algorithm for detecting calibration pulses were reported by both Kafka (pp. 258–9) and Tyson (pp. 281–2). Tyson's results for calibration pulse detection are shown for the linear algorithm in Fig. 11, and for the non-linear algorithm in Fig. 12. There is a clear peak for the linear algorithm, whereas no such peak is apparent for the non-linear procedure. (The calibration pulses were inserted periodically during data taking runs. The peak was displaced by two seconds by the insertion of a time delay, so that the calibration pulses would not mask any possible real signal, which was expected at zero time delay).

Nevertheless, Weber preferred the non-linear algorithm. His reason for this was that this procedure gave a more significant signal than did the linear one. This is illustrated in

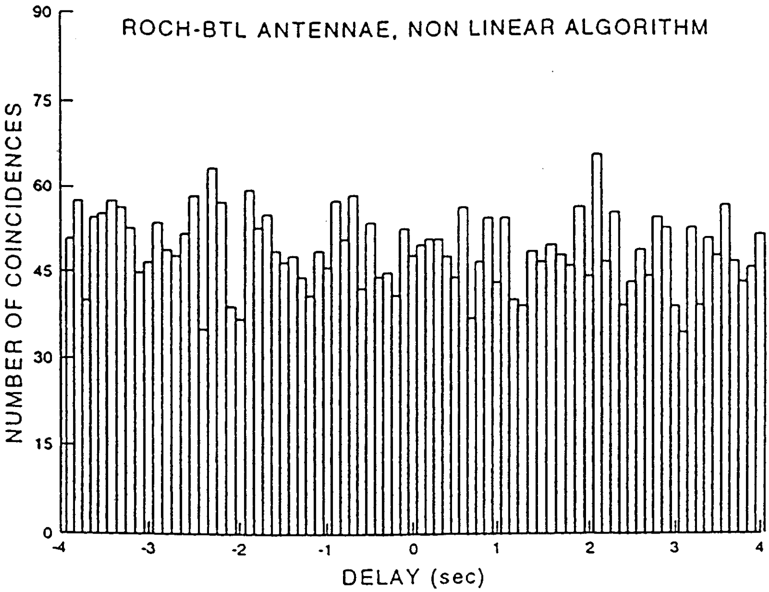

Fig. 12. A time-delay plot for the Rochester-Bell Laboratory collaboration, using the non-linear algorithm. No sign of any zero-delay peak is seen. From Shaviv and Rosen (1975)

Fig. 13, in which the data analyzed with the non-linear algorithm is presented in (a) and for the linear procedure in (b). "Clearly these results are inconsistent with the generally accepted idea that $\dot{x}^2 + \dot{y}^2$ [the linear algorithm] should be the better algorithm" (Weber, pp. 251–2). Weber was, in fact, using the positive result to decide which was the better analysis procedure. He was tuning his analysis procedure to maximize his result.

His critics, however, analyzed their own data using both algorithms. If it was the case that, unlike the calibration pulses where the linear algorithm was superior, using the linear algorithm either masked or failed to detect a real signal, then using the non-linear algorithm on their data should produce a clear signal. None appeared. Typical results are shown in Figs. 12 and 14. Figure 12, which is Tyson's data analyzed with the non-linear algorithm, not only shows no calibration peak, but it does not show a signal peak at zero time delay. It is quite similar to the data analyzed with the linear algorithm shown in Fig. 14. (Note that for this data run no calibration pulses were inserted). Kafka also reported the same result: no difference in signal between the linear and the non-linear analysis.

Weber had an answer. He suggested that although the linear algorithm was better for detecting calibration pulses, which were short, the real signal of gravitational waves was a longer pulse than most investigators thought. He argued that the non-linear algorithm that he used was better at detecting these longer pulses. Still, if the signal was longer, one would have expected it to show up when the critics's data was processed with the non-linear algorithm. It didn't. (See Figs. 12 and 14). The critics' results were robust under changes in the analysis procedure.

Drever also reported that he had looked at the sensitivity of his apparatus with arbitrary waveforms and pulse lengths. Although he found a reduced sensitivity for

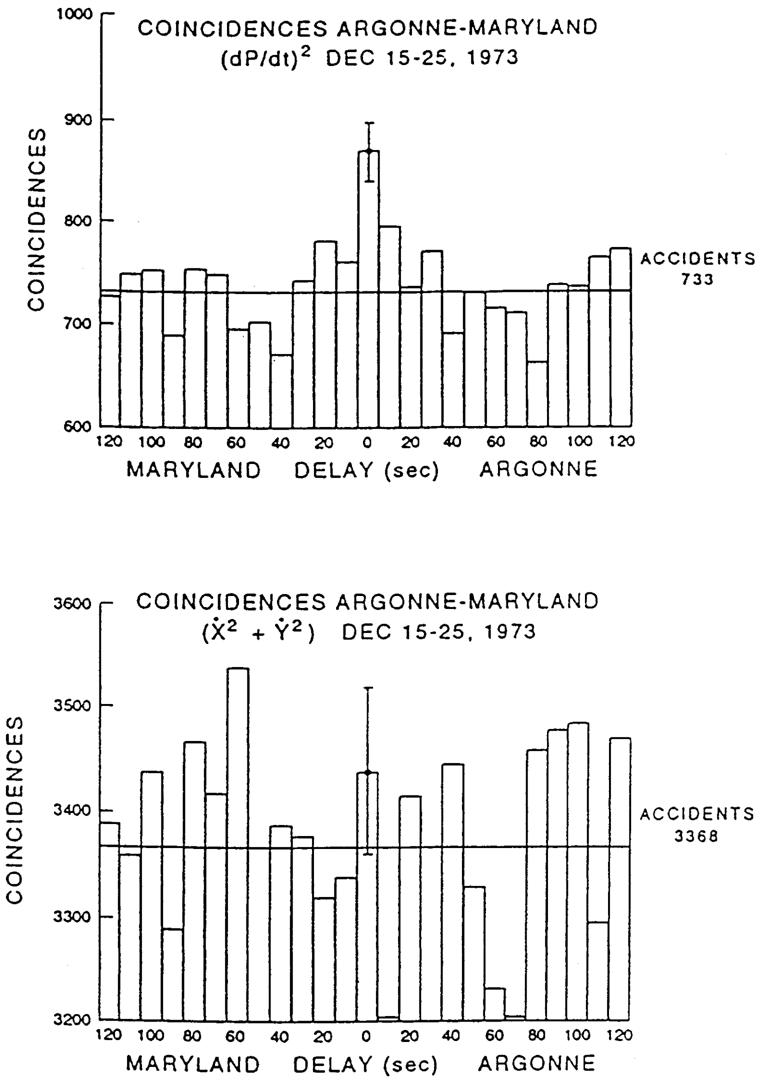

Fig. 13. Weber's time delay-data for the Maryland-Argonne collaboration for the period Dec. 15–25, 1973. The top graph uses the non-linear algorithm, whereas the bottom uses the linear algorithm. The zero-delay peak is seen only with the non-linear algorithm.
From Shaviv and Rosen (1975)

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.[23]

---

[23] 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

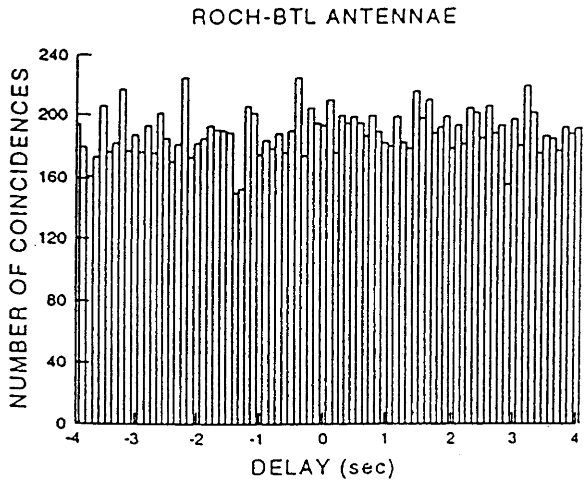

Fig. 14. A time-delay plot for the Rochester-Bell Laboratory collaboration, using the linear algorithm. No sign of a zero-delay peak is seen. From Shaviv and Rosen (1975)

How then did Weber obtain his positive result when his critics, using his own analysis program, could not? It was suggested that Weber had varied his threshold cut, to maximize his signal, whereas his critics used a constant threshold. Was Weber tuning the threshold cut to create a result?

> Tyson characterized the difference between Weber's methods and those of his critics.

---

other 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, pp. 287–8).

> I should point out that there is a very important difference in essence in the way in which many of us approach this subject and the way Weber approaches it. We have taken the attitude that, since these are integrating calorimeter type experiments which are not too sensitive to the nature of pulses put in, we simply maximize the sensitivity and use the algorithms which we found maximized the signal to noise ratio, as I showed you. Whereas Weber's approach is, he says, as follows. He really does not know what is happening, and *therefore he or his programmer is twisting all the adjustments in the experiment more or less continuously, at every instant in time locally maximizing the excess at zero time delay*. I want to point out that there is a potentially serious possibility for error in this approach. No longer can you just speak about Poisson statistics. *You are biasing yourself to zero time delay, by continuously modifying the experiment on as short a time scale as possible (about four days), to maximize the number of events detected at zero time delay*. We are taking the opposite approach, which is to calibrate the antennas with all possible known sources of excitation, see what the result is, and maximize our probability of detection. Then we go through all of the data with that one algorithm and integrate all of them. Weber made the following comment before and I quote out of context: "Results pile up". I agree with Joe (Weber). But I think you have to analyze all of the data with one well-understood algorithm (Tyson, p. 293, emphasis added).

A similar criticism was offered by Garwin, who also presented evidence from a computer simulation to demonstrate that a selection procedure such as Weber's could indeed produce his positive result.

> Second, in view of the fact that Weber at CCR-5 [a conference on General Relativity held in Cambridge][24] explained that when the Maryland group failed to find a positive coincidence excess "we try harder," and since in any case there has clearly been selection by the Maryland group (with the publication of data showing positive coincidence excesses but with no publication of data that does not show such excesses),[25] James L. Levine has considered an extreme example of such selections. In Fig. 15 is shown the combined histogram of "coincidences" between two independent streams of random computer-generated data. This "delay histogram" was obtained by partitioning the data into 40 segments. For each segment, "single events" were defined in each "channel" by assuming one of three thresholds a, b, or c. That combination of thresholds was chosen for each segment which gave the maximum "zero delay coincidence" rate for that segment, The result was 40 segments selected from one of nine "experiments." *The 40 segments are summarized in Fig. 15, which shows a "six-standard-deviation"zero-delay excess*(Garwin 1974, pp. 9–10, emphasis added).

[24] I have been unable to find a published proceedings of this conference. Richard Garwin (private communication) has informed me that these proceedings were never published.

[25] As Weber answered, the Maryland group had presented data showing no positive coincidence excess at GR7. Garwin was not, however, at that meeting, and the proceedings were not published until after Garwin's 1974 letter appeared.

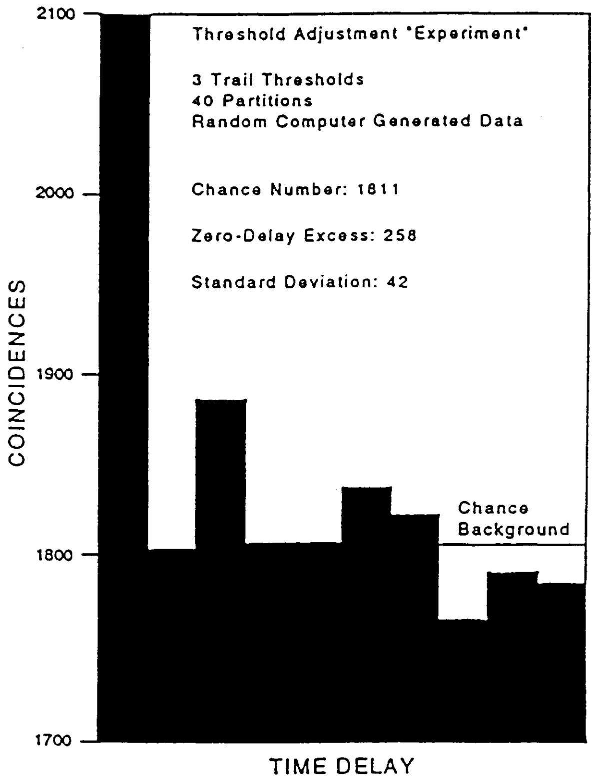

Fig. 15. The result of selecting thresholds that maximized the zero-delay signal, for Levine's computer simulation. From Garwin (1974)

Weber denied the charges.

It is not true that we turn our knobs continuously. I have been full time at the University of California at Irvine for the last six months, and have not been turning the knobs by remote control from California [Weber's group and one of his antennas was located at the University of Maryland]. In fact, the parameters have not been changed for almost a year. What we do is write the two algorithms on a tape continuously. The computer varies the thresholds to get a computer printout which is for 31 different thresholds. The data shown are not the results of looking over a lot of possibilities and selecting the most attractive ones. We obtain a result that is more than three standard deviations for an extended period for a wide range of thresholds. I think it is very important to take the point of view that the histogram itself is the final judge of what the sensitivity is (Weber, pp. 293–4).

Weber did not, however, specify his method of data selection for his histogram. In particular, he did not state that all of the results presented in a particular histogram had the same threshold.

Interestingly, Weber cited evidence provided by Kafka as supporting a positive gravity wave result. Kafka did not agree. This was because the evidence resulted from performing an analysis using different data segments and different thresholds. Only one showed a positive result, indicating, in fact, that such selectivity could produce a positive result. Kafka's results are shown in Fig. 16. Note that the positive effect is seen in only the bottom graph. "The very last picture (Fig. 16) is the one in which Joe Weber thinks we have discovered something, too. This is for 16 days out of 150. There is a 3.6 σ [standard deviation] peak at zero time delay, but you must not be too impressed by that. It is one out of 13 pieces for which the evaluation was done, and I looked at least at 7 pairs of thresholds. Taking into account selection we can estimate the probability to find such a peak accidentally to be of the order of 1%"(Kafka, p. 265).

In this episode it was suggested that Weber's positive result was caused, in part, by tuning both his analysis procedure and his threshold cut to produce that result. Weber's critics dealt with this problem by analyzing their data using both their own linear analysis algorithm and Weber's preferred non-linear algorithm and obtained negative results with both procedures. The results were robust. They also showed how one might produce a positive result by tuning the threshold cut. This was shown by Kafka, using his actual data, and by Levine and Garwin, whose Monte Carlo simulation was able to obtain a positive result from random data by manipulating the threshold cut. These latter arguments did not conclusively demonstrate that Weber's results were incorrect, but they strongly suggested that they had credibility problems. These problems combined with Weber's calibration failure and the other difficulties discussed earlier convinced the physics community that Weber had not observed gravitational radiation. There were, in addition, the negative results from six other experiments.

IV. Millikan's oil-drop experiments

Millikan's oil-drop experiments are justly regarded as a major contribution to twentieth-century physics (Millikan 1911, 1913). They established the quantization of electric charge, the existence of a fundamental unit of charge, and also measured that unit of charge precisely. Earlier determinations of the charge of the electron had not established whether there was a fundamental unit of electricity.[26] This was because previous experiments, which used a cloud of charged water droplets and observed the motion of the cloud under the influence of both gravity and an electric field and under gravity alone, measured the total charge of the cloud, and could not therefore demonstrate that the value obtained wasn't a statistical average. Millikan was able to perform all of his measurements on a single oil drop and avoid that difficulty.

A. Millikan's method

Let us briefly examine how Millikan demonstrated the existence of a fundamental unit of electrical charge and measured its value. Millikan allowed a single oil drop to fall a known distance in air. He did not measure the time of fall from rest, but allowed the

[26] For a discussion of some of these early experiments see Millikan (1917, Chapter III).

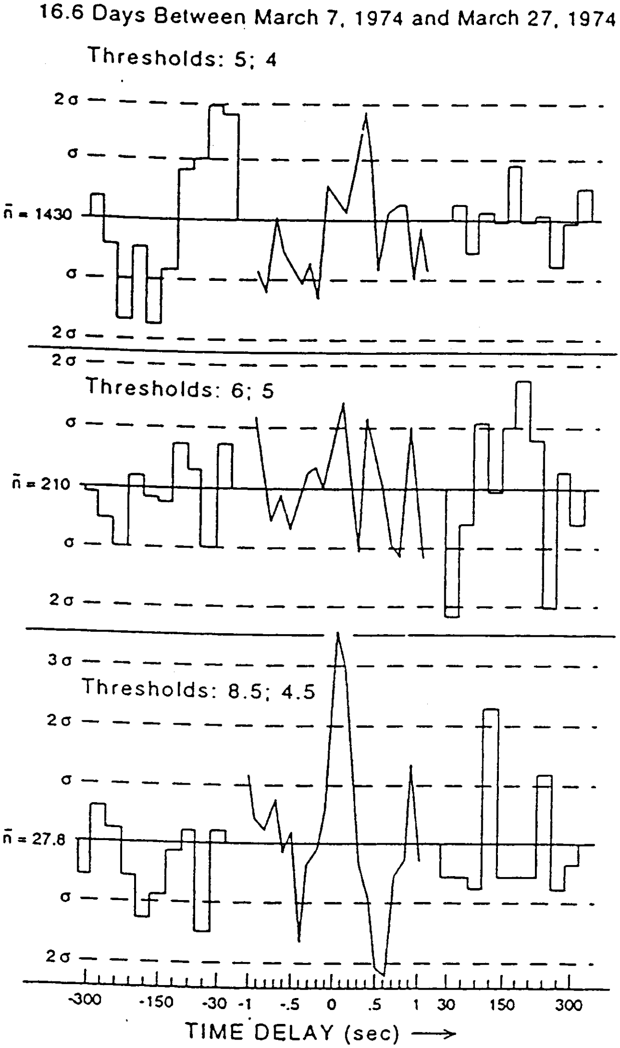

Fig. 16. Kafka's results using varying thresholds. A clear peak is seen at zero-delay. From Shaviv and Rosen (1975)

drop to fall freely for a short distance before it passed a crosshair, which was the start of the time measurement. Because of air resistance the drop was then traveling at a constant, terminal velocity. After the drop passed a second crosshair, which determined the time of fall at constant speed for the known distance between the crosshairs, an electric field was turned on. The charged oil drop then traveled upward at a different constant speed and the time to ascend the same distance was measured. These two time measurements allowed the determination of both the mass of the drop and its total charge.

The equation of motion of an oil drop falling under an upwards electric field F is

$$\dot{m}\ddot{x} = mg - K\dot{x} - e_n F,$$

where e_n is the drop's charge and *m* its mass compensated for the buoyant force of air. According to Stokes' law, which holds for a continuous retarding medium, $K = 6\pi a\mu$, where *a* is the drop's radius and μ the air's viscosity; to take into account the particulate character of air Millikan replaced *K* by *K/(1 + b/pa)*, where *p* is the air pressure and *b* a parameter determined from all of the experimental data. Because all measurements were made at terminal velocities $\ddot{x}$ vanished, (recalling that ν_g is opposite in direction to ν_f):

$$e_n = mg/F[(\nu_f + \nu_g)/\nu_g];$$

$$e_n = ne = (9\pi d\sqrt{(}2/g))/F[\mu^3/(\sigma - \rho)(1 + b/pa)^3]^{1/2}[\sqrt{\nu}_g(1/t_g + 1/t_f)]$$

where e_n, the total charge on the drop is assumed to be an integral multiple of a unit *e*, *d* is the distance traveled, either up or down, *g* is the acceleration of gravity.[27]

Not only did the charge on the oil drop sometimes change spontaneously due to absorption of charge from the air or by ionization, but Millikan induced such changes with either a radioactive source or an x-ray source. One can calculate the change in charge using successive times of ascent (before and after the change):

$$\Delta e_n = (\Delta n)e = (9\pi d\sqrt{(}2/g))/F[\mu^3/(\sigma - \rho)(1 + b/pa)^3]^{1/2}[\sqrt{\nu}_g(1/t'_f - 1/t_f)]$$

where t'_f and t_f are the successive times with the field on. If the total charge on the drop and the change in charge are both multiples of some fundamental unit of charge then $1/n(1/t_g + 1/t_f)$ should equal $(1/\Delta n)(1/t'_f - 1/t_f)$. Millikan could easily estimate $(1/t'_f - 1/t_f)$ because Δ_n was usually a very small number, often equal to one, for the smallest change in the charge of the drop. "Since *n*' [our Δn] is always a small number and in some of the changes always had the value 1 or 2 its determination for any change is obviously never a matter of the slightest uncertainty. On the other hand *n* is often a large number, but with the aid of the known values of *n'* it can always be found with absolute certainty as long as it does not exceed say 100 or 150" (Millikan 1913, pp. 123–24).

Sample data sheets from Millikan's experiments are shown in Figs. 17 and 18. (These sheets are from Millikans notebooks of 1911 and 1912. The results of that experiment were published in 1913). The columns labeled G and F are the measurements of $\mathrm{t_g}$ and $\mathrm{t_f}$, respectively. The average value of $\mathrm{t_g}$ and its reciprocal are given at the bottom of column G. To the right of column F are calculations of $1/\mathrm{t_f}$ and of $[1/\Delta \mathrm{n}(1/\mathrm{t'_f} - 1/\mathrm{t_f})]$. Further to the right is the calculation of $[1/\mathrm{n}(1/\mathrm{t_f} + 1/\mathrm{t_g})]$. The top of the page gives the date, the number and time of the observation, the temperature θ, the pressure *p*, and the voltage readings, which include the actual reading plus a correction, and the time at which the voltage was read. The data combined with the physical dimensions of the apparatus, the density of clock oil and of air, the viscosity of air, and the value of *g* are all that is required to calculate *e*.

[27] The subscripts indicate terminal velocities without (ν_g) and with (ν_f) the field, respectively. Now *m* can be replaced by *a* using $m = 4/3\pi a^3(\sigma - \rho)$, σ and ρ being the densities of oil and air, respectively; *a* can be done away with in favor of μ using Stokes' law; and the ratios of distance *d*, to times of fall and rise, t_g and t_f,can be substituted for the velocities. Millikan did not make all of these substitutions. He left a factor of ν_g in his final formula, presumably for ease of calculation.

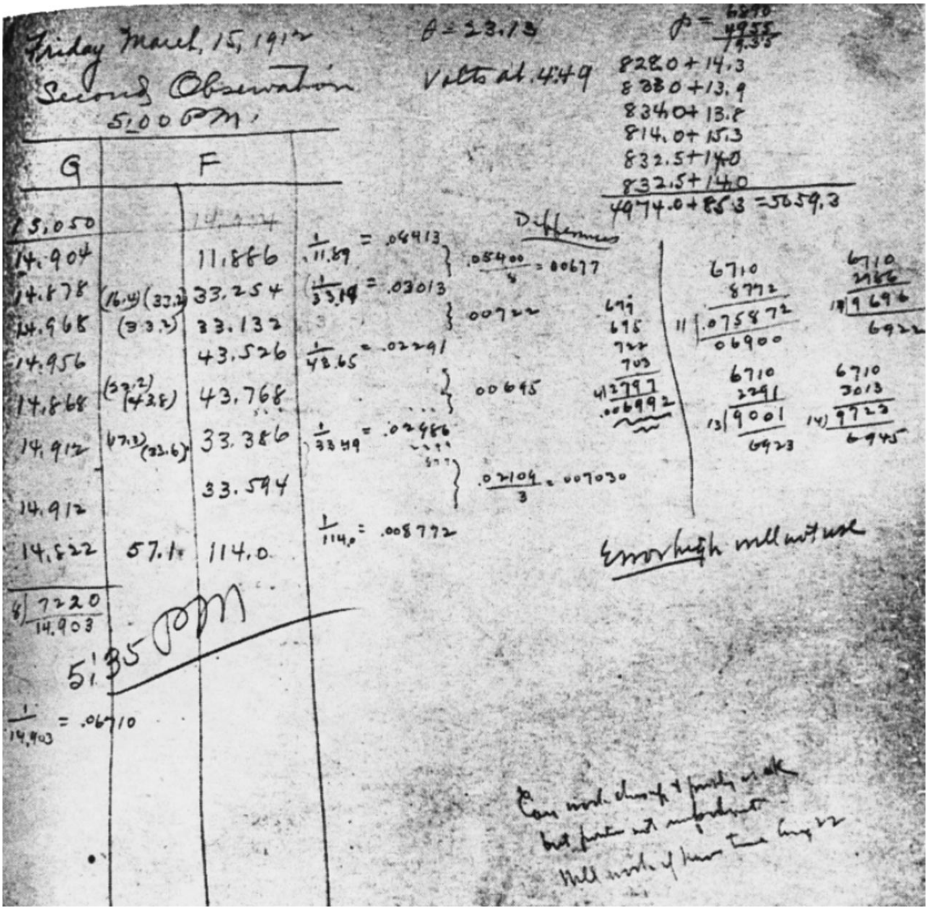

Fig. 17. Millikan's data sheet for 15 March 1912 (second observation). Notice "*Error high* will not use," at the center-right. Courtesy California Institute of Technology Archives

B. Millikan's results

Millikan could determine *e* from both the total charge of the drop and from the changes in the charge. Not only did these values agree but the average value obtained from different drops were also the same. Millikan remarked, "The total number of changes which we have observed would be between one and two thousand, and *in not one single instance has there been any change which did not represent the advent upon the drop of one definite invariable quantity of electricity or a very small multiple of that quantity*" (Millikan 1911, p. 360). For Millikan, and for most of the physics community, these results established the quantization of charge. The value that Millikan found in 1911 for the fundamental unit of charge, the charge on the electron, was 4.891×10^{-10}

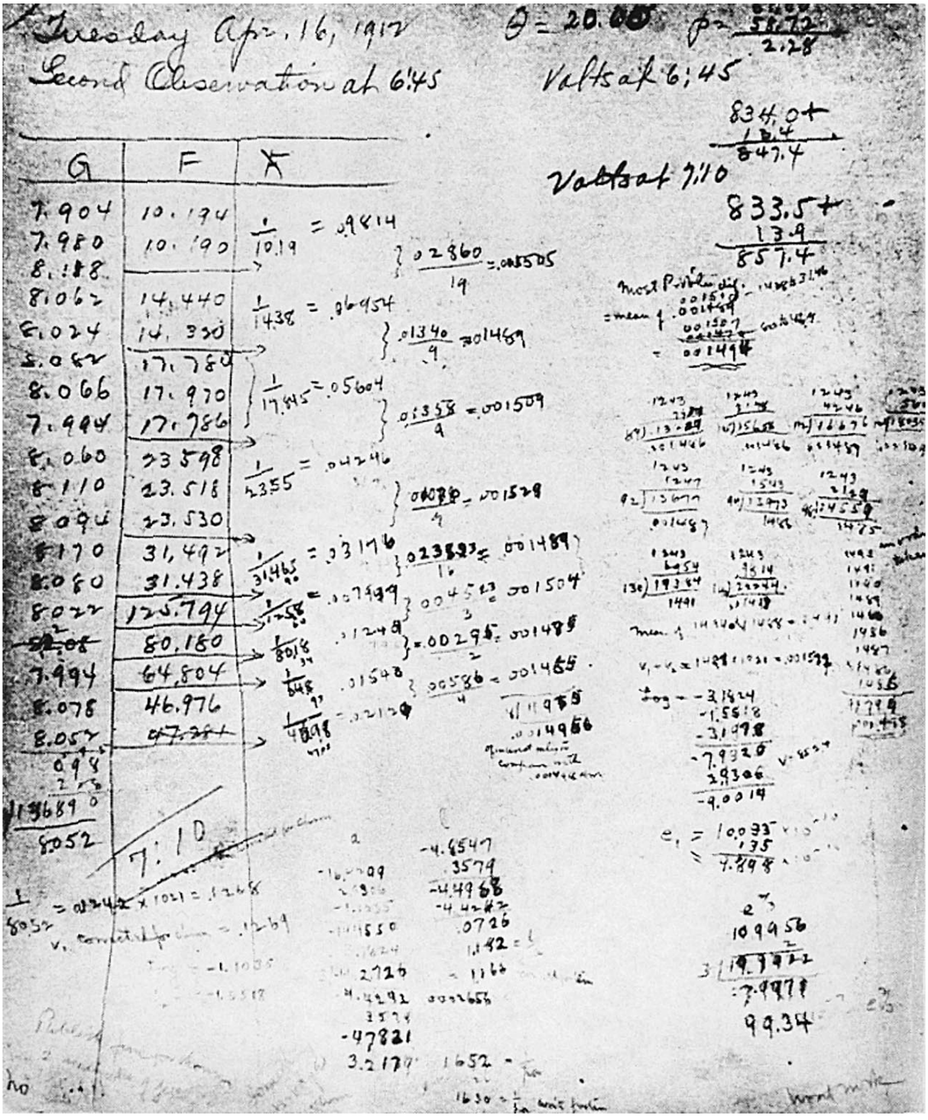

Fig. 18. Millikan's data sheet for 16 April 1912 (second observation). Notice "Won't work" in lower right-hand corner. Courtesy California Institute of Technology Archives

esu. (Millikan did not give a numerical uncertainty, but estimated the uncertainty as approximately 0.2 percent).[28]

---

[28] This is a very small uncertainty. Millikan estimated this using the statistical uncertainty in his final value for e. He did not include any uncertainty caused by systematic effects.

Following the completion of his 1911 paper, Millikan continued his oil drop measurements. His intent was to improve both the accuracy and the precision of the measurement of e. He made improvements in his optical system and determined a better value for the viscosity of air. In addition, he took far more data in this second experiment. Millikan's new measurement gave a value of $e = (4.774 \pm 0.009) \times 10^{-10}$ esu.[29] This value differs considerably from his 1911 value of 4.891×10^{-10} esu. Millikan stated that, "The difference between these numbers and those originally found by the oil-drop method, $e = 4.891$, was due to the fact that this much more elaborate and prolonged study had the effect of changing everyone of the three factors η [the viscosity of air], A [related to the correction parameter b in Stokes' Law], and d [the distance between the crosshairs], in such a way as to lower e and to raise N [Avogadro's number]. The chief change, however, has been the elimination of faults of the original optical system" (Millikan 1913, pp. 140–41).

In producing his new value of e Millikan engaged in selectivity in both the data he used and in his analysis procedure. In presenting his results in 1913 Millikan stated that the 58 drops under discussion had provided his entire set of data. "*It is to be remarked, too, that this is not a selected group of drops but represents all of the drops experimented upon during 60 consecutive days,* during which time the apparatus was taken down several times and set up anew" (Millikan 1913, p. 138). This is not correct. An examination of Millikans notebooks for this period shows that Millikan took data for this measurement from 28 October 1911 to 16 April 1912.[30] My own count of the number of drops experimented on during this period is 175. Even if one were willing to count only those observations made after 13 February 1912, the date of the first observation Millikan published, there are 49 excluded drops: that is, of 107 drops experimented on between 13 February and 16 April, Millikan published only 58. We might suspect, and worry, that Millikan selectively analyzed his data to support his preconceptions.[31]

Millikan's selectivity included the exclusion of all of the data of some single drops, exclusion of some of the data within the data set for a single drop, and a choice of methods of calculation. In discussing this selectivity we should, however, remember that Millikan had far more data than he needed to improve the measured value of e by approximately a factor of ten. He used only published drops – 23 out of a total of 58 – which had a Stokes' Law correction of less than six percent to calculate his final value of e. This was to guard against any effect of an error in that correction.[32]

---

[29] Millikan's value for e differs from the modern value $e = (4.8032068 \pm 0.0000015) \times 10^{-10}$ esu. This difference is due, in large part, to a difference between the modern value for the viscosity of air and the one that Millikan used.

[30] My work here is based on Millikan's notebooks at the California Institute of Technology. For details of my recalculation of Millikan's data see Franklin (1981).

[31] Daniel Siegel raises the same question. "Millikan was in this sense choosing data according to his presuppositions, and then using those data to support his presuppositions" (Siegel 1979, p. 476). As discussed in Franklin (1981), and below, I don't agree with Siegel's statement.

[32] Recall that Millikan modified Stokes' Law substituting $K/(1 + b/pa)$ for K to take into account the particulate nature of the air. The parameter b was empirically determined from the entire data set and has an uncertainty. In addition, this was a first order approximation to Stokes' Law. Other terms in the approximation may have been important. This was the case for the 12

In experiments before 13 February 1912, Millikan had labored to make his apparatus work properly. He was particularly worried about convection currents inside the device that could change the path of the oil drop. He made several tests on slow drops, for which convection effects would be most apparent. Millikan's comments on these tests are quite illuminating. On 19 December 1911 he remarked, "This work on a very slow drop was done to see whether there were appreciable convection currents. The results indicate that there were. Must look more carefully henceforth to tem[perature] of room".[33] On 20 December: "Conditions today were particularly good and results should be more than usually reliable. We kept tem very constant with fan, a precaution not heretofore taken in room 12 but found yesterday to be *quite* essential". On 9 February 1912 he disregarded his first drop because of uncertainty caused by convection; after the third drop he wrote, "This is good for so little a one but on these very small ones I must avoid convection still better". No further convection tests are recorded. By 13 February it seems that the device was working to Millikan's satisfaction because he eventually published data from the very first drop recorded on that day. The data from 68 drops taken before 13 February were omitted from publication because Millikan was not convinced that his experimental apparatus had been working properly. It was not producing "good" data.[34]

After this date, we must assume that the apparatus was working properly unless we are explicitly told otherwise. There are 107 drops in question, of which 58 were published. Millikan made no calculation of e on 22 of the 49 unpublished drops. The most plausible explanation for why Millikan did not do so is that when he performed his final calculations in August 1912 they seemed superfluous: he did not need more drops for the determination of e when he already had so many.

The 27 events that Millikan did not publish and for which he calculated a value of e are more worrisome. Millikan knew the results he was excluding. Twelve of these were excluded from the set of published drops because they seemed to require a second-order correction to Stokes' law. These were very small drops, for which the value of Millikan's correction to Stokes' law, b/pa, was larger than one. This made Millikan's use of a perturbation series expansion of Stokes' law for those drops very questionable. There is no easy way to calculate the correction for such drops so Millikan, having so much data, decided to exclude them.[35] Of the remaining 15 calculated events Millikan excluded two because the apparatus was not working properly, five because there was insufficient data to make a reliable determination of e, two for no apparent reason. We are left with six drops. One is quite anomalous and is discussed in note 36.[36] In the other

---

drops, discussed below, that Millikan did not publish because they seemed to require a second-order correction to Stokes' Law.

[33] The quotations are from Millikan's notebooks.

[34] Interestingly, the value for e that one could determine from these 68 excluded events was $e = (4.75 \pm 0.01) \times 10^{-10}$ esu. This was, in fact, more precise than any other measurement available at the time. The data was, however, untrustworthy.

[35] I attempted, without success, to calculate a second-order correction to Stokes' law for these 12 drops. I found no consistent way to do so.

[36] The second drop of 16 April 1912 is quite anomalous. (The data sheet for this drop is shown in Figure 18). It is also quite worrisome because it is among Millikan's most consistent measurements. Not only are the two methods of calculating einternally consistent, but they agree

five cases Millikan not only calculated a value for *e* but compared it with an expected value. The four earliest events have values of *pa* that would place them in the group Millikan used to determine *e*. His only evident reason for rejecting these five events is that their values did not agree with his expectations. Including these events among the 23 that Millikan used to determine *e* would not change the value of *e*, but would increase the statistical error of the measurement.

In addition to excluding these five drops from publication, Millikan's cosmetic surgery touched 30 of the 58 published events. As shown in Figs. 17 and 18, Millikan made many measurements of the time of fall under gravity and of the time of ascent with the electric field on for each drop. In the data set for each of these 30 drops Millikan excluded one or more (usually less than three) of these measurements. This group of 30 drops included several of the 23 drops used by Millikan in his final determination of *e* as well some of the 35 published drops that were not used. My recalculation of these events, using all of the data for each drop, gives results little different from Millikan's. Millikan's exclusion of these measurements was not based on the value of *e* he obtained for the drops, because, in general, Millikan did not include these measurements in his calculations, and therefore did not know their effect.

As discussed earlier, there are two ways to calculate *e* from the oil-drop data. The first uses the total charge on the drop, whereas the second uses the changes in charge. Millikan claimed that he had used the first method exclusively because the large number of measurements of t_g provided a more accurate determination of *e* . In at least 19 of the 58 published events,[37] however, he used either the average value of the two methods, some combination of the two that is not a strict average, or the second method alone (Fig. 17 shows the data sheet for an event that Millikan excluded because he thought the difference in the value of *e* obtained by the two different methods was too large). In general, the effects are small and the result of his tinkering, once again, is to reduce the statistical error very slightly rather than to change the mean value of *e.*

The effect of all of Millikan's selectivity is shown in Table 3. This includes Millikan's results along with my own recalculation of Millikan's data. As we can see the results of his selectivity are quite small.

---

with each other very well. Millikan liked it: "Publish. Fine for showing two methods of getting v." My own calculation of e for this event gives a value $e = 2.810 \times 10^{-10}$ esu, or approximately 0.6*e*. Millikan knew this. Note the comment, "Wont work" in the lower right-hand corner. There were no obvious experimental difficulties that could explain the anomaly. Millikan remarked, "Something wrong with therm[ometer]," but there is no temperature effect that could by any stretch of the imagination explain a discrepancy of this magnitude. Millikan may have excluded this event to avoid giving Ehrenhaft ammunition in the controversy over the quantization of charge. In retrospect Millikan was correct in excluding this drop. In later work William Fairbank Jr. and I found that Millikan's apparatus gave unreliable charge measurements when the charge on the drop exceeded a value of about 30e. This drop had a charge of greater than 50e, and the data was quite unreliable (Fairbank Jr. and Franklin 1982).

[37] This group of 19 drops included some of those used in Millikan's final calculation of *e* as well as some of those omitted from the calculation.

Table 3. Comparison of Millikans and Franklin's values of *e*

	values of $e \times 10^{10}$ esu			
Published drops	e(RM)	e(AF)	σ(RM)[a]	σ(AF)
First 23[b]	4.778[c]	4.773	±0.002	±0.004
All 58	4.780	4.777	±0.002	±0.003
Almost all drops[d]	4.781	4.780	±0.003	±0.003

[a] Statistical error in the mean

[b] These are the events that Millikan used to determine *e*.

[c] Although Millikan used a value $\mu = 0.001825$ for the viscosity of air in almost all of his calculations, in reporting his final value for *e* he used $\mu = 0.001824$. This accounts for the change from 4.774 to 4.778 in Millikans final value. In order to make the most accurate comparison I used $\mu = 0.001825$ in all of my recalculations.

[d] This includes the 58 published drops, 25 unpublished drops measured after 13 February 1912, and some small corrections. For details see Franklin (1981).

B. Discussion

What can we conclude from this episode? Millikan intended to establish the quantization of charge and to measure the fundamental unit more accurately and precisely than had been done previously. It's apparent that he certainly succeeded: there is no reason to disagree with his assessment that, in 1913, there was "no determination of *e*... by any other method which does not involve an uncertainty at least 16 times as great as that represented in these measurements." His apparently arbitrary exclusion of 5 drops for which he had calculated *e*, a possible worry, had, as we have seen, an utterly negligible effect on his final result. Because Millikan knew the value of *e* obtained from the events he was excluding, he also knew that the effect of the exclusions, and of his selective calculations, on his final result was small.

Nevertheless there is strong evidence that Millikan tuned his cuts and his analysis procedure to obtain the result he wanted. Several of these cuts seem quite legitimate: the exclusion of the early drops because he was not sure that the experimental apparatus was working properly; the exclusion of some data within the set for a drop; and the exclusion of later events because he simply did not need them for his calculations. The exclusion of drops for which he calculated a value of *e* and could thus select the value he wanted as well as his choice of calculational method are not justified. The fact of Millikan's tuning was unavailable to the physics community. Unlike the episodes of the measurement of the K^{+}_{e2} branching ratio and of Weber's attempts to detect gravity waves in which the cuts were publicly accessible,[38] Millikan's questionable selectivity remained private. In fact, his statements in his 1913 paper that the drops were not selected and that he used

[38] Although the fact that Weber used a threshold cut was publicly known, the value of that cut, and whether he used a single threshold value, was not known. This lessened the credibility of his result.

only one method of calculation seem to have been designed to conceal that selectivity. What then are the safeguards against procedures such as Millikan's, which, in less sure hands, could easily have unfortunate results? In this case the answer is replication. The value of *e* was an important physical quantity. It was used in the calculation of many important physical constants, i.e. Avogadro's number, the Rydberg constant etc. There were many repetitions of Millikan's measurement.[39] Had Millikan's selectivity grossly affected his measured value of *e*, there would certainly have been a discrepancy with later measurements.

V. The disappearing particle: The case of the 17-keV neutrino[40]

Radioactivity, the spontaneous decay of an atomic nucleus, produces alpha particles (positively charged helium nuclei), or beta particles (electrons), or gamma rays (high energy electromagnetic radiation). Experimental work on the energy of the electrons emitted in β decay began in the early twentieth century, and the continuous energy spectrum observed posed a problem. If β decay were a two-body process (for example, neutron $\rightarrow$ proton + electron) then applying the laws of conservation of energy and of conservation of momentum requires that the energy of the electron have a unique value, not a continuous spectrum.[41] Physicists at first speculated that the observed continuous energy spectrum might be due to energy lost by electrons in escaping the substance, with different electrons losing different amounts of energy, thus accounting for the energy spectrum. Careful experiments showed that this was not the case, which also raised doubts concerning the applicability both conservation laws to this situation. In the early 1930s Pauli suggested that a low-mass neutral particle was also emitted in beta decay.[42] This solved the problem of the continuous energy spectrum because in a three-body decay (for example, neutron $\rightarrow$ proton + electron + neutrino) the energy of the electron was no longer required to be unique. The electron could have a continuous energy spectrum and the conservation laws would be saved.[43] Fermi named this new particle the neutrino,

---

[39] It is rare that an important physical quantity is measured only once. Recall the numerous attempts to replicate Weber's experiment on gravity waves. Other examples abound. (See, for example, Franklin 1986, Ch. 1–3, and the section below on electron-positron states). The only instance I can think of in which an important physical quantity was measured only once was in the SLAC E122 experiment. This experiment measured the asymmetry in the scattering of polarized electrons from deuterons, an important prediction of the Weinberg-Slam unified theory of electroweak interactions. Because of its expense and complexity, this experiment was done only once. For that reason the arguments for the validity of its results were extensive. (See Franklin 1990, Ch. 8).

[40] For details of this episode see Franklin (1995).

[41] The conservation of momentum also requires that the electron and proton have equal and opposite momenta, for a neutron decay at rest. They will be emitted in opposite directions.

[42] Pauli's suggestion was first made in a December 4, 1930 letter to the radioactive group at a regional meeting in Tubingen.

[43] With a three-body decay the electron and the proton also didn't have to come off back to back. This observation was not made until the late 1930s. Assigning the neutrino a spin, with intrinsic angular momentum, of 1/2 h, also preserved the law of conservation of angular momentum.

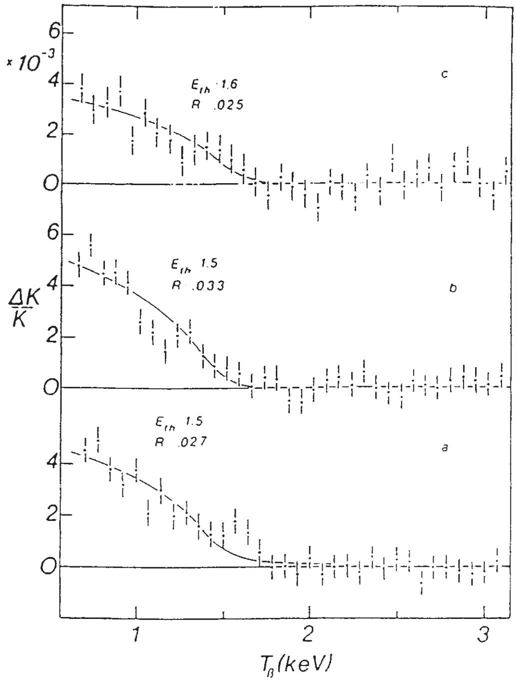

Fig. 19. The data of three runs presented as $\Delta K/K$ (the fractional change in the Kurie plot) as a function of the kinetic energy of the β particles. E_{th} is the threshold energy, the difference between the endpoint energy and the mass of the heavy neutrino. A kink is clearly seen at $E_{th} = 1.5$ 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)

little neutral one, because its mass was very close to zero.[44] The neutrino was directly observed in 1956 by Reines and Cowan, although the empirical success of Fermi's 1934 theory of β decay, as well as its role in the preservation of the conservation laws had already persuaded physicists that it existed. Our story begins with later work on this rather elusive particle, with the claim that there existed a heavy neutrino, in addition to the ordinary one.

The 17-keV (or heavy) neutrino was "discovered" by Simpson in 1985 (Simpson 1985a). He had searched for a heavy neutrino by looking for a kink in the decay energy

---

[44] For a long time physicists believed that the mass of the neutrino was zero. Physicists also currently believe that there are actually three kinds of neutrino, each associated with one elementary particle. These are the electron, muon, and tau neutrinos, along with their respective antiparticles. The particle we are discussing here is the electron antineutrino. The question of whether the mass of these neutrinos is zero is currently under investigation.

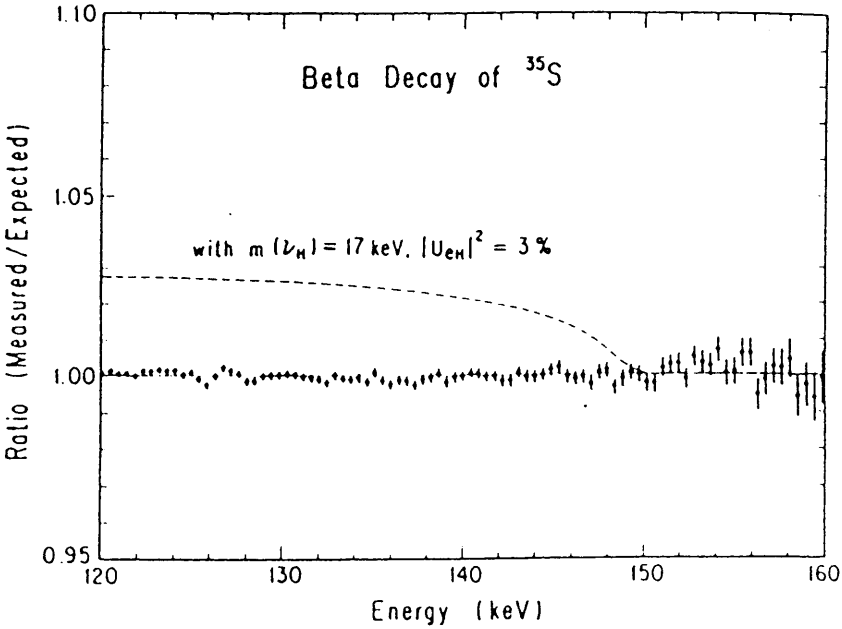

Fig. 20. The ratio of the measured ^{35}S beta-ray spectrum to the theoretical spectrum. A three percent mixing of a 17-keV neutrino should distort the spectrum as indicated by the dashed curve. From Ohi et al. (1985)

spectrum, or in the Kurie plot,[45] 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 $\Delta K/K$ is approximately $R[1 - M_2^2/(Q - E)^2]^{1/2}$, 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 original experimental result for the decay energy spectrum of tritium is shown in Fig. 19. A kink, a marked change in slope of the $\Delta K/K$ graph, is clearly seen at T_β (the electron energy) equal to 1.5 keV, corresponding to a 17 keV neutrino. (The maximum decay energy for tritium is 18.6 keV. If there were no effect due to the presence of a heavy neutrino this graph would be a horizontal straight line.) "In summary, the β spectrum of tritium recorded in the present experiment is consistent with the emission of a heavy neutrino of mass about 17.1 keV and a mixing probability [the fraction of heavy neutrinos] of about 3%"(Simpson 1985, p. 1893).[46]

---

[45] In a normal beta-decay spectrum the quantity $K = (N(E)/[f(Z,E)\,(E^2 - 1)^{1/2}\,E])^{1/2}$ 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.

[46] Later work, including some by Simpson, reduced the size of the positive effect to approximately one percent.

Within a year there were five attempted replications of Simpson's experiment. These initial replications all gave negative results. A typical result, that of Ohi and collaborators (Ohi et al. 1985), is shown in Fig. 20 (The dotted line is the effect expected if there were a 17-keV neutrino with a 3% mixing probability). Suggestions were made that attempted to explain Simpson's result using accepted physics, without the need for a heavy neutrino. Subsequent positive results by Simpson and others led to further investigation. Several of these later experiments found evidence supporting his claim, whereas others found no evidence for such a particle. Theorists took two broad approaches. One group rejected the heavy neutrino altogether. Among them, some attributed the positive observations to either systematic instrumental effects or to faulty analysis; others thought that the observations required a deeper explanation, and they provided one, but without introducing a new particle. A second group accepted the positive observation. They also divided into two. Some in this group presumed the existence of the particle and sought to trace out its implications. Others proposed an altogether new theory, one that incorporated the particle at a fundamental level.

The question of the existence of such a heavy neutrino remained unanswered for several years. Recently, doubt has been cast on the two most convincing positive experimental results, and errors found in those experiments. In addition, two recent, extremely sensitive experiments have found no evidence for the 17-keV neutrino. The present consensus is that it does not exist.

A. Selectivity and the 17-keV neutrino

One of the issues involved in the resolution of the discordant results in this episode was that of the choice of the energy range used to fit the decay energy spectrum, so that the experimental and theoretical spectra could be compared. Simpson argued that because 45% of the effect expected occurred within 2 keV of the neutrino threshold a narrow energy range around that threshold should be used. "...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). Hetherington and others (1987) suggested caution. "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% [much lower than the 3% effect originally found by Simpson]. 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

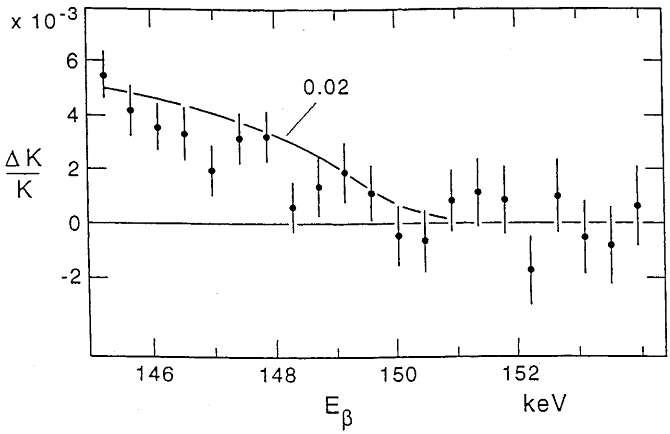

Fig. 21. ΔK/K for the ^{35}S spectra of Ohi and others (1985) as recalculated by Simpson. From Simpson (1986a)

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).[47]

This issue was dramatically demonstrated by Simpson's reanalysis (1986a) of the data of Ohi et al. (1985). Recall from Figure 20 that Ohi's result showed no evidence for a 17-keV neutrino. Simpson's reanalysis of that same data shows clear positive evidence (Fig. 21). How can the same data provide both positive and negative evidence for the same effect? The answer is that the analysis procedures were quite different. Ohi and collaborators had used a wide energy range for their analysis.[48] As Morrison later showed, the positive effect found by Simpson was due to his use of a narrow energy range for his reanalysis of Ohi's data. "The question then is, How could the apparently negative evidence of Fig. 20 become the positive evidence of Fig. 21? The explanation is given in Fig. 22, 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

[47] Hime, one of Simpson's collaborators, agreed. "The difficulty remains, however, that an analysis using such a narrow region could mistake statistical fluctuations as a physical effect. The claim of positive effects in these cases [by Simpson] should be taken lightly without a more rigorous treatment of the data" (Hime 1992, p. 1303).

[48] There was yet another problem with the analysis of Ohi et al. As Borge et al. noted, "We feel, *in complete agreement with the opinions expressed by J.J. Simpson...that the limits on c_2 derived in (the experiments of Ohi et al. (1985) and of (Datar et al. 1985)) are misleading as the parameters were not fitted again under the assumption of a heavy neutrino; instead the contribution from this was simply added.*(Borge et al. 1986, pp. 593–594, emphasis added").

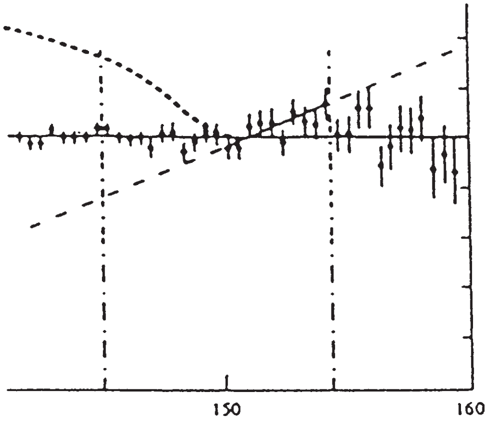

Fig. 22. Morrison's reanalysis of Simpson's reanalysis of Ohi's result. From Morrison (1992)

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.[Simpson's straight-line fit to the data just above 150 keV and its extrapolation is the line going from lower left to upper right. Comparing the data points to this line generates the positive effect seen in Fig. 21. The dotted curve above the data is the effect expected for a heavy neutrino.] 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). The caution urged by both Hetherington and collaborators and by Hime was justified.[49]

How was this issue dealt with? Several later experiments used *both* a narrow and a wide energy range in the analysis of their data (Hetherington et al. 1987; Kawakami et al. 1992; Radcliffe et al. 1992; Ohshima et al. 1993a; Ohshima 1993b). For example, Hetherington et al. concluded that their results from both the wide- and narrow-energy range analyses agreed and that "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" (Hetherington et al. 1987, p. 1510).

Ultimately, the decision that the 17-keV neutrino did not exist was based on finding errors in the two most persuasive positive results and by the overwhelming negative evidence provided by experiments which avoided explicitly the data analysis issues posed by narrow vs. wide energy range. The first of these experiments was that of a Tokyo group (Kawakami et al. 1992; Ohshima et al. 1993a; Ohshima 1993b). These experimenters noted some of the problems that plague experiments using wide energy regions and decided therefore to concentrate, "on performing a measurement of high statistical

[49] One might speculate that had Simpson varied the endpoints of the energy range he used in his reanalysis he would have avoided his difficulty. As Morrison clearly showed, Simpson's reanalysis of Ohi's data lacked robustness.

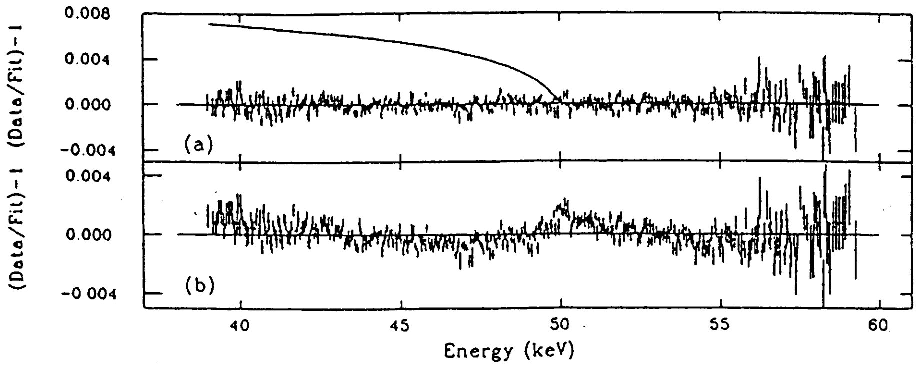

Fig. 23. Deviations from the best global fit with $|U|^2$ free (**a**) and fixed to 1% (**b**). The curve in (**a**) indicates the size of a 1% mixing effect of the 17-keV neutrino. From Ohshima (1992)

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 are shown in Fig. 23, 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) $\pm$ 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 and was certainly far lower than the approximately one percent effect found be Simpson and others. "The result clearly excludes neutrinos with $| U |^2 \geq$ 0.1% for the mass range 11 to 24 keV" (Ohshima 1993b, p. 1128).

Although the experiment's narrow energy range was designed to minimize the dependence of the result on the shape correction,[50] 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 answered essentially all of his criticism of spectrometer experiments convincingly. "Thus, I conclude that this experiment could not possibly have missed the

[50] Magnetic spectrometer experiments required a shape-correction factor, usually of the form $(1 + \alpha E)$, in order to fit their spectra. This was an important issue in the resolution of the discord. For details see Franklin (1995).

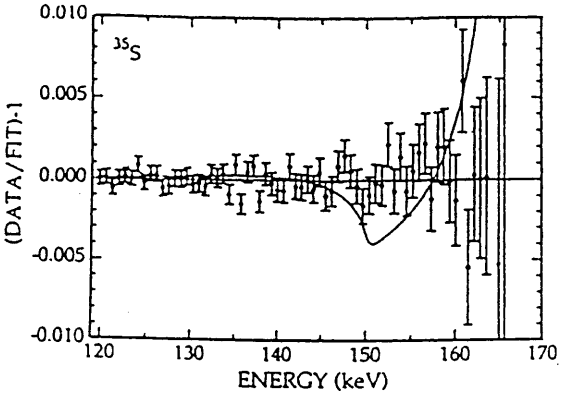

Fig. 24. Residuals from a fit to the pile-up corrected ^{35}S data assuming no massive neutrino; the reduced χ^2 for the fit is 0.88. The solid curve represents the residuals expected for decay with a 17-keV neutrino and $\sin^2\theta = 0.85\%$; the reduced χ^2 of the data is 2.82.
From Mortara et al. (1993)

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).[51]

The second experiment that provided decisive evidence against the existence of the 17-keV neutrino was that of a group at the Argonne National Laboratory (Mortara et al. 1993). Not only did this experiment find no evidence for a 17-keV neutrino (Fig. 24),[52] but the experimenters demonstrated that they would have observed such evidence had it been present.[53]

> 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 ($E_0 - m_e$ 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. 25. 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 $(1.4 \pm 0.1)\%$.

[51] Bonvicini had reanalyzed many of the early experiments. He showed, using Monte Carlo techniques, that the shape-correction factor needed in the magnetic spectrometer experiments, combined with the limited statistics of those experiments, could mask or mimic the presence of a heavy neutrino. His analysis showed that the negative evidence provided by the early replications of Simpson's experiment was not as strong as had been originally claimed. He did, however, conclude that the experiment of Hetherington et al. (1987) was sufficient to rule out a 3% effect (Bonvicini 1993). Bonvicini's work also influenced the design of later experiments.

[52] The upper limit found by the Argonne group was $\sin^2\theta = -0.0004 \pm 0.0008$(statistical) $\pm$ 0.0008 (systematic). This was also far lower than one percent.

[53] This is an example, albeit a complex one, of the calibration of an experimental apparatus. For details see (Franklin 1997).

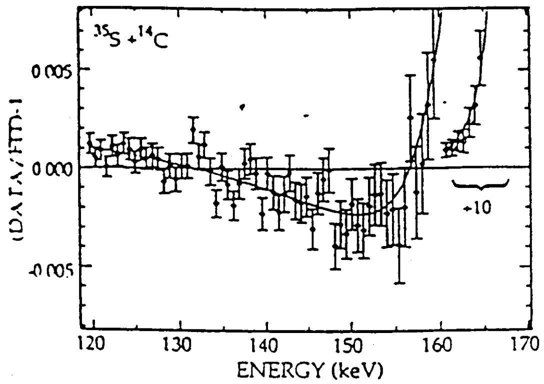

Fig. 25. Residuals from fitting the beta spectrum of a mixed source of ^{14}C and ^{35}S with a pure ^{35}S shape; the reduced χ^2 of the data is 3.59. The solid curve indicates residuals expected from the known ^{14}C contamination. The best fit yields a mixing of $(1.4 \pm 0.1)\%$ and reduced χ^2 of 1.06. From Mortara et al. (1993)

> 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. (Mortara et al. 1993, p. 396).

The fact that an effect of both the right size and shape was observed when the spectra of ^{35}S and ^{14}C were mixed demonstrated that the analysis procedure used did not mask a real effect. The failure to observe any effect in the ^{35}S spectrum alone showed that the analysis procedure did not create an artifact that would mimic the effect of a heavy neutrino.

B. Discussion

The case of the 17-keV neutrino is yet another illustration of selectivity applied to analysis procedures. The problem of the appropriate analysis procedure, in particular the energy range to be used in the analysis, was resolved, in part, by several experiments in which both procedures were used, and in which the results were robust under changes in the procedures. The two decisive, negative results, those of the Tokyo and Argonne groups, both demonstrated, albeit in different ways, that the choice of analysis procedure was not a relevant issue in assessing the validity of their results. The Tokyo group not only used both procedures, but had such high statistics that a kink in the spectrum of the size found by Simpson could not have been missed in their analysis. The Argonne group demonstrated that their experiment (apparatus and analysis procedure) would have detected a 17-keV neutrino at the one percent level had one been present

VI. Are there really low-mass electron-positron states?

There has been recent controversy concerning the possible existence of low-mass electron-positron states, or particles, produced in high-energy, heavy ion-atom collisions. Briefly, the history is as follows. Early experiments found evidence that positrons with discrete energies, or positron lines, were produced in such collisions. Later experiments also detected an electron produced in coincidence with the positron and found peaks in the sum-energy spectrum ($E_{e+} + E_{e-}$), suggestive evidence for low-mass electron-positron states. The fact that the effects appeared for various pairs of projectile and target nuclei, as well as in three different experimental apparatuses, gave the results credibility, although there were some problems concerning the reproducibility of the results. This credibility led others to further investigate the phenomena. These later experiments produced conflicting claims. Several groups claimed to find evidence for such states, whereas others found no evidence for such effects. Part of the difficulty in attempting to resolve this issue was that the results were not reproducible. Even when later experiments were done under very similar conditions to the original experiments the effects did not always appear, and when they did the states sometimes had different energies from those observed initially. As Tom Cowan, one of the experimenters, remarked, "You had to search for these things. [And] you did not see all the lines, all together, always" (quoted in Taubes 1997, p. 149). The failure to reproduce the effects cast doubt on the original results themselves.

A recent analysis by Ganz and the EPOS II (Electron POsitron Solenoidal spectrometer) group (Ganz et al. 1996) has suggested that the observed effects might be artifacts created by the selection criteria used to produce the experimental result.[54] Applying certain selection cuts to one half of his data and tuning the cuts to produce a maximal effect, Ganz found rather strong evidence (a five standard deviation effect)[55] for a low-mass state. Applying identical cuts to the other half of his data showed no such evidence.[56] This failure cast doubt on the result and also on the statistical procedures used. The originally observed effect was five standard deviations above background. If the effect really is statistical it is very improbable that it would disappear in the analysis of the other half of the data. The implication is that the effect is an artifact created by the cuts. Although most physicists working in the field believe that the proposed low-mass electron-positron states do not exist, not everyone agrees (see discussion below and Taubes 1997).

This episode nicely raises the question of whether particular experimental cuts or selection criteria can create an effect that is not really present.[57] I will examine how the physicists involved dealt with this issue and with the associated problem of discordant

[54] The EPOS I group was one of those that reported the original effect. The I and II refer to different versions of the experimental apparatus. The membership of the group also changed.

[55] The probability of a five standard-deviation effect is 5.7×10^{-7}.

[56] Ganz divided his data set into two subsets by using a random-number generator on an event by event basis. This guarded against any systematic effects that varied with time.

[57] By suitable cuts on the data one can argue that all small odd numbers are primes. Consider 1, 3, 5, 7, 9, 11, 13. If we exclude 9 as an experimental error, our hypothesis is supported by the data.

experimental results. The situation is made more difficult by the fact that the heavy ion-atom systems under investigation contain a large number of particles with many possible interactions. The fact that one might study these reactions as a function of no fewer than 14 different variables further increases the difficulty of the experiment. As P. Kienle, one of the original experimenters, stated, "A theoretician dreams of somebody who would measure the differential cross section

$$d^n\sigma\,(Z_{p'}, T_{p'}, \theta_{p'}, Q, \Delta Z, \Delta A, T_{e+}, \theta_{e+}, T_{e-}, \ldots\text{etc.})$$

This seems to be nearly impossible, but all quantities may be of some relevance" (Kienle 1983, p. 293). ($Z_{p'}$ is the charge on the projectile after the interaction, $T_{p'}$ is its kinetic energy, θ_{e+} is the angle at which the positron is produced, etc.)

On the theoretical side the situation is made difficult by the fact that the heavy ion-atom systems being studied are quite complex and because the positrons may be created by both electromagnetic processes in the strong electric fields produced in heavy ion-atom collisions, and also by nuclear interactions. "One can only admire the enormous ingenuity and effort that has gone into understanding the physics processes of colliding, say, uranium on uranium. This forms a composite system of 576 nucleons and 184 electrons (and innumerably more particles on the quark level). The physical processes are mind-boggling to untangle: bremsstrahlumg, δ-electrons [electrons produced in the collisions], Coulomb excitation, internal conversion; direct-, spontaneous-, and induced emission of positrons, to say nothing of nuclear interactions...." (Biedenharn 1983, p. 841).

A. Positron line spectra

The history of possible low-mass electron-positron states began with the demonstration that positrons were indeed produced in high-energy, heavy ion-atom collisions (Backe et al. 1978; Kozhuharov et al. 1979). The motivation for these experiments was the theoretical speculation that positrons would be produced in the strong electric fields produced in such collisions. The positrons were expected to be produced in supercritical systems, those for which the binding energy of the lowest energy state was greater than $2m_ec^2$, where m_e is the mass of the electron. This would be true for heavy ion-atom systems which had a total nuclear charge, Z_u, greater than 173. These early experiments, all performed at the heavy-ion accelerator at the Gesellschaft für Schwerionforschung (GSI), Darmstadt, Germany, showed a surprising enhancement of positron production at energies less than 400 keV. They were followed by more detailed experiments which produced unexpected, on the basis of accepted theory, results.[58] These early experimental results illustrate quite clearly the problems of reproducibility and of the sensitivity of experimental results to both the selection criteria used to produce them and to the experimental conditions. These would be continuing problems in the subsequent history.

The three early experiments used two different methods for high-efficiency positron detection. The first method, used by both the EPOS I group and by the TORI group, used

---

[58] These new results were presented by three different groups at the NATO Advanced Study Institute on Quantum Electrodynamics of Strong Fields held on 15–26 June, 1981 in Lahnstein, Germany (Greiner 1983).

solenoid transport systems in which the positrons are guided by a magnetic field to a silicon [(Si(Li)] detector placed a large distance from the target.[59] The positron spectra were measured with the Si(Li) detector, in coincidence with one or two annihilation gamma rays[60] detected by a ring of sodium iodide (NaI) scintillators placed around the Si(Li) detector. The background due to δ-rays, electrons produced in the collisions, which is orders of magnitude larger than the positron signal, was eliminated by spiral baffles in the straight solenoid, or by using a segment of the toroidal solenoid which causes the positrons and electrons to drift in opposite directions perpendicular to the toroidal plane so that they can either be removed by a diaphragm, or recorded with a detector. The heavy ions produced were detected in position-sensitive avalanche counters or in the particle counters. Sodium iodide counters were used to detect the γ rays resulting from nuclear transitions The second method used a large solid angle "orange-type" β-ray spectrometer. In this spectrometer a toroidal magnetic field sector was used to focus the positrons on the detector, which consisted of a plastic scintillator surrounded by a position-sensitive proportional counter.

In the subsequent history no questions were raised concerning the adequacy of the different detectors. Everyone agreed that they were good positron detectors and that they had been carefully calibrated.[61]

One of the intriguing new results, found by the EPOS I group, was this: "the most striking result for the ^{238}U + ^{238}U-measurement at E(^{238}U) = 5.9 MeV/u" (Bokemeyer et al. 1983, p. 287) is shown in Fig. 27. The lower graph shows the positron energy spectrum for the angular region $11^\circ \leq \mid \Delta\theta \mid = \mid \theta_1 - \theta 2 \mid \leq 19^\circ$ and $89.2^\circ \leq \Sigma\theta = \theta_1 + \theta_2 \leq 89.8^\circ$, where θ_1 and θ_2 are the scattering angles of the projectile ion and the target atom. Two peak-like structures are seen at positron energies of approximately 320 keV and 590 keV. The peaks do not appear in the upper graph, which is the positron energy spectrum for $25^\circ \leq \mid \Delta\theta \mid \leq 35^\circ$ under otherwise identical conditions. The EPOS I experimenters were aware of the problems of both reproducibility and selectivity. In discussing the significance of their results they stated, "However, before drawing any far

---

59 The momentum of a charged particle determines its trajectory in a magnetic field.

60 The Si(Li) detector is sensitive to the positron which then annihilates with an electron in the detector, producing two gamma rays, which are then detected by the NaI counters.

61 A typical calibration procedure is that of the APEX group.

> The ability of APEX to observe positron-electron pair events in the presence of an intense background of electrons can be tested with an 90Y source. ^{90}Y decays largely by e^-emission to the ground state of ^{90}Zr but has a weak (0.011%) branch to the 1761 keV 0^+ state which subsequently decays by pair emission to the ground state. The sum energy of the pair is 739 Kev. Figure [26] shows the sum-energy spectrum of positron-electron pairs from such a measurement. The sum-energy peak is observed with a width of approximately 27 keV. The background below the peak results from out-scattered positrons and electrons which do not deposit their full energy in the silicon. These data were taken with a 30 μCi source which gives 10^6 electrons per sec, comparable to the number produced during in-beam measurements. (Ahmad et al. 1995, p. 251c).

Some commentators on science, notably Harry Collins (1985) have questioned the efficacy of calibration. I have argued elsewhere that they are wrong (Franklin 1997).

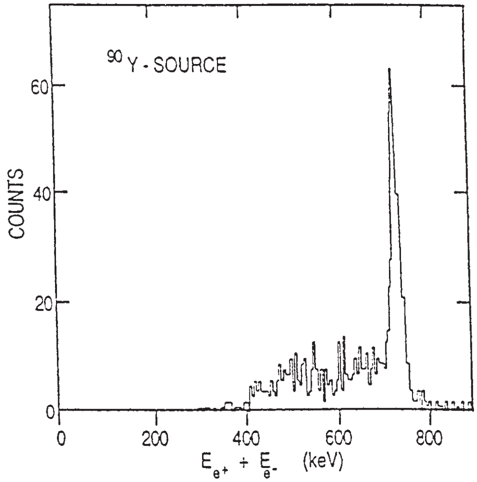

Fig. 26. Sum-energy spectrum for positron-electron pairs from ^{90}Y decay. The pairs should have a sum energy of 739 keV

reaching conclusions from the experimental data presented, we feel that the following questions should be solved. First of all one must show by further analysis that the structures are not produced by some yet unknown background effects *associated with one of the event selecting criteria. Secondly, the reproducibility of the effect has to be shown"*(Bokemeyer et al. 1983, p. 290, emphasis added).

The problem of reproducibility mentioned was not merely methodological and abstract. It appeared quite dramatically in the results presented by the ORANGE group (Kienle 1983). Figure 28 shows peak-like structures appearing at positron energies of 370 keV, and possibly at 720 keV and 950 keV. The structures appear most strongly only for certain scattering angles, an effect similar to that of the EPOS I group (Bokemeyer et al. 1983). The energy of the lowest peak differs from that of EPOS I by approximately 50 keV and the 590 keV peak seen by the latter is not visible at all. The peaks at 720 keV and 950 keV observed by the ORANGE group were not seen by EPOS I. Nor were they seen when the ORANGE group repeated the experiments. "Figure 29 shows the results of some of these more recent experiments. The peaked structure appears to be a much less prominent effect and the mean energy of the peak has shifted to a lower energy of $\sim$310 KeV [in better agreement with that of EPOS I]. At a bombarding energy of 6.0 MeV/amu, the peak structure is not evident within the statistical accuracy of the data. If not statistical in origin, the non-reproducibility of the data is a troubling aspect which is not well understood" (Greenberg 1983, p. 881).

An excess of positrons produced at low energy, reported by the TORI group (Backe et al. 1983a) at the same conference, suffered from the same problem. "However, it is again disconcerting to find that the inconsistency observed for some of the measurements reported by Kienle is repeated here. A subsequent experiment was not able to reproduce the deviations from theory found in the initial data on the U + U system. This begins to suggest that controlling the bombarding conditions carefully may be a crucial ingredient

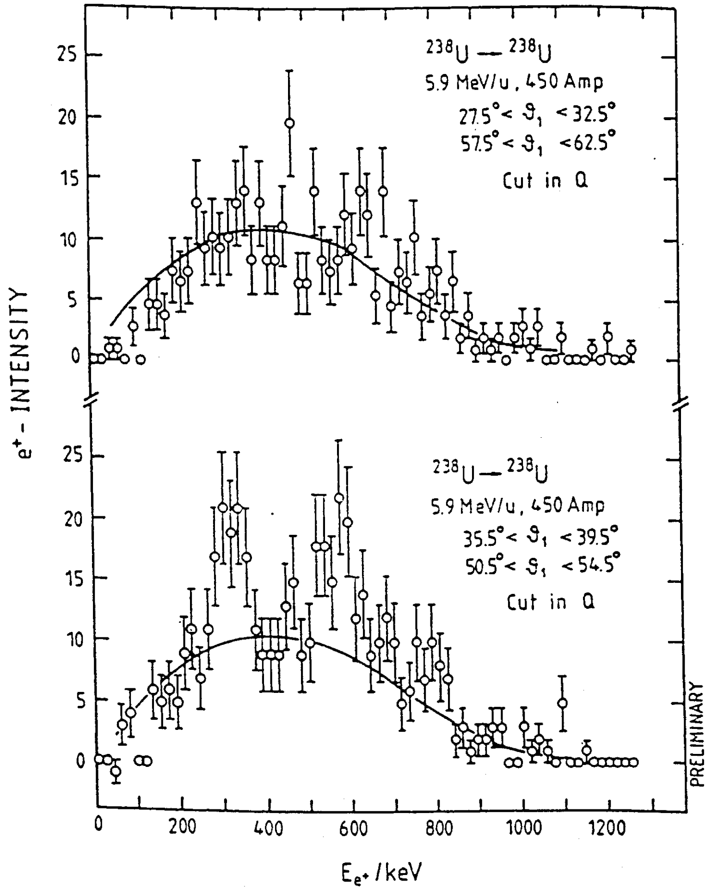

Fig. 27. Two selected e^+-energy spectra observed in $^{238}U + ^{238}U$ collisions at 5.9 MeV/u. The solid lines represent the theoretical spectra normalized to the upper spectrum. Notice that the energy peaks appear only in the lower graph. From Bokemeyer (1983)

in studying these effects. As I noted already, the underlying reason for this sensitivity is presently not understood, unfortunately" (Greenberg 1983, p. 883). The question of the sensitivity of the observed results to the experimental conditions and what constitutes an adequate replication had appeared.

Greenberg's summary of the experimental situation at the time of the 1981 conference was, despite the problems he had noted, quite positive. "Thus, these last experiments, like the others we have discussed, suggest very convincingly that there are excess positrons above the dynamically induced background, but they go even further in pointing out that the additional positrons are associated with selected kinematic conditions possibly reflecting a focussed nuclear reaction" (Greenberg 1983, p. 886). Unfortunately, theory did not provide either suggestions or justification for what those conditions and associated selection criteria should be, further complicating the problem.

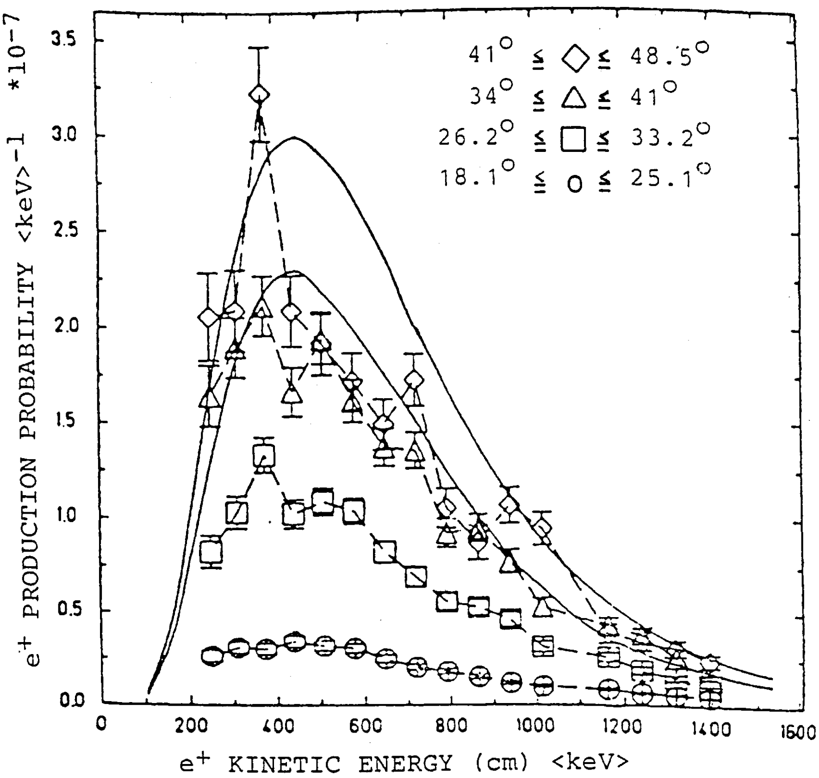

Fig. 28. Positron production probability as a function of positron kinetic energy in the center-of-mass system for (U + U) collisions at 5.9 MeV/u bombarding energy in coincidence with ions scattered into the designated angular regions. From Kienle (1983)

Experimental work continued, but the uncertainty remained. For example, the EPOS I group reported a 316-keV positron line produced in U + Cm collisions (Greenberg and Greiner 1982; Schweppe et al. 1983). "In summary, a well defined, narrow positron peak has been detected in the U + Cm system at an energy commensurate with a supercritically bound state for the combined system. Internal pair conversion of nuclear transitions does not seem to yield a plausible explanation of the peak structure. A consistent interpretation of the peak, however, can be provided by spontaneous positron emission enhanced by the formation of a metastable, giant dinuclear system" (Schweppe et al.1983, p. 2264). Unfortunately, the group using the TORI spectrometer concluded that for the same reaction "no statistically significant structures have been observed in this experiment" (Backe et al. 1983b, p. 1840).

The existence of a 300-keV peak was further supported by the ORANGE group (Clemente et al. 1984). Their results, however, seemed to be sensitive to both the bombarding energy and to the scattering angle, as shown in Figure 30. The position and strength of the observed peaks are sensitive to both the bombarding energy and to the heavy-ion scattering angle. Further work by both EPOS I and ORANGE groups achieved,

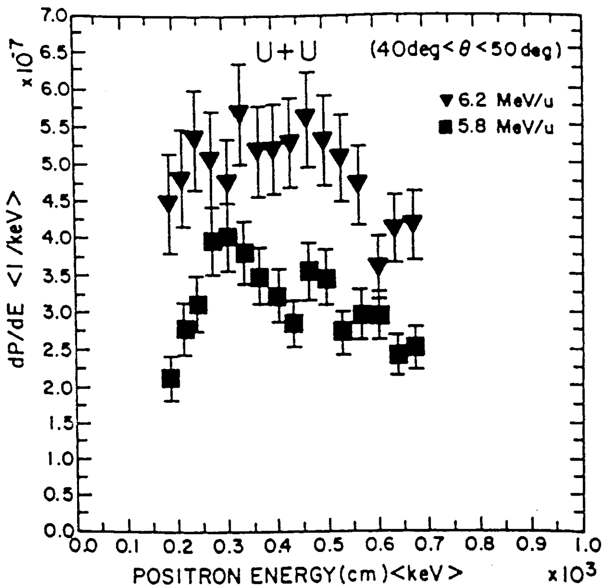

Fig. 29. Positron production probability as a function of positron kinetic energy for (U+U) collisions in coincidence with ions scattered into designated angular region. From Greenberg (1983)

at least, internal consistency. The EPOS I group observed peaks in the Th + Th, Th + U, U + U, Th + Cm, and U + Cm systems (Cowan et al. 1985). Their average value for the peak energy was 336 keV. Similar consistency was found by the ORANGE group for both the U + U and the U + Th systems. Their peak energy, however, was approximately 280 keV. Under virtually identical conditions different results had been obtained. The ORANGE group concluded, "Recent observations [a reference to the EPOS I result] show that the near constancy of the line energy, observed already in our previous work [see above][62] persists even up to a combined charge $Z_u = 188$ and down to $Z_u = 180$ (*although systematic differences in the absolute values, found by the two groups are not yet cleared up*" (Tsertos et al. 1985, p. 276, emphasis added). There was another difference between the results reported by the two groups. EPOS I had found a constant cross-section as a function of Z_u in the range 180–188, whereas ORANGE found a factor of two difference between U + U (184) and U + Th (182).

Both groups agreed that neither nuclear transitions nor spontaneous positron emission could be the cause of the positron lines. The latter required a very sharp dependence of the peak energy on the total charge of the system ($\sim Z_u^{20}$) which was obviously not present in the data. For the U + Th and U + Cm systems, with total charges of 180 and 188, respectively, a Z_u^{20} dependence, predicted that the energy of the peak would change

[62] The experimenters are concentrating on the consistency of their earlier result, whereas I noted some of the differences. Of course, I have the advantage of hindsight.

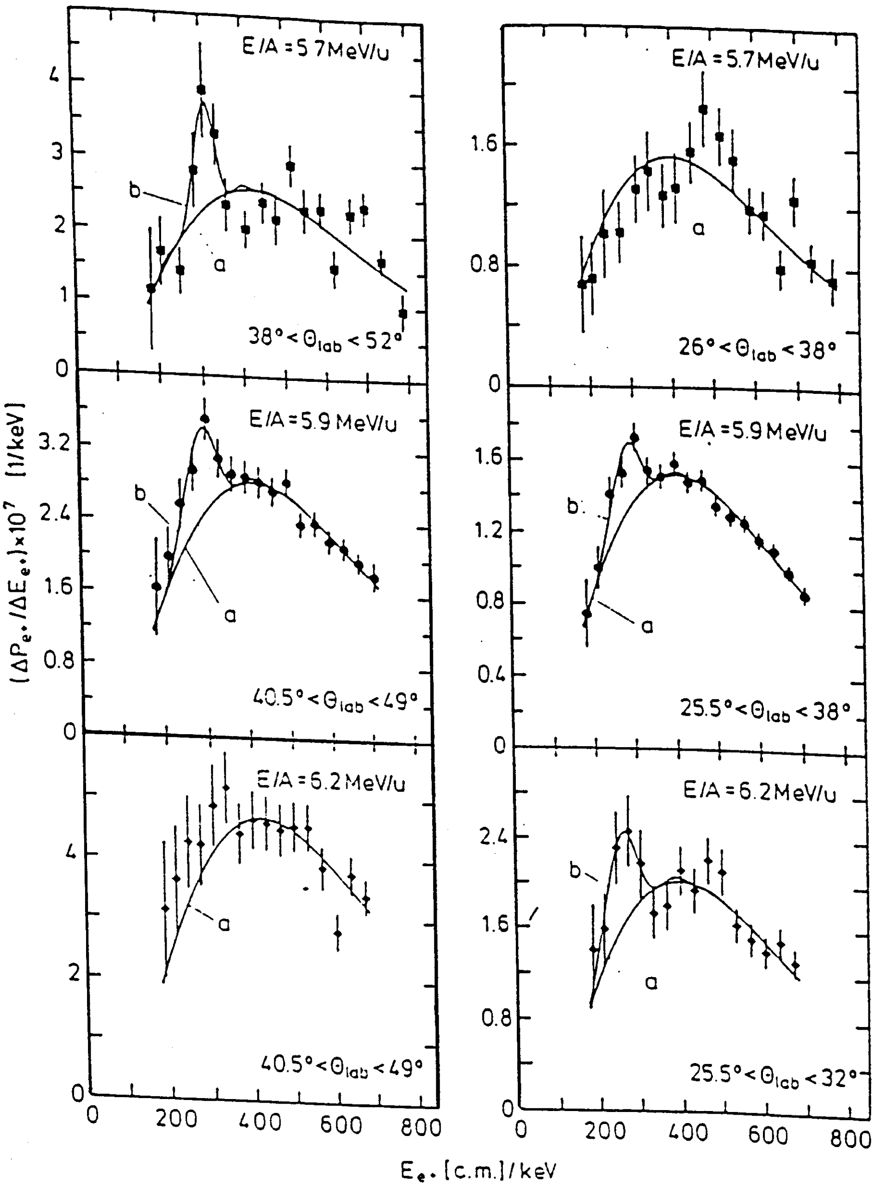

Fig. 30. Positron creation probability as a function of positron energy for various bombarding energies and scattering angles. Notice that the peaks do not appear in all of the graphs, not do they appear at the same energy in each of the graphs. From Clemente et al. (1984)

by a factor of 2.4. The values of the peak energy, although different from one another, were constant as a function of the total charge, Z_u.

One could save spontaneous positron emission as a cause of the observed positron lines, but only by invoking changes in both the charge configuration and in the ionization states for the compound system that were regarded as physically unrealistic and implausible. “Without independent evidence to support these unusual conditions, this explanation for the peaks does not presently appear very plausible” (Cowan et al. 1986, p. 444). Both groups suggested that the constant energy peak might have a common, and different, source. “An obvious speculation is that the source of the monoenergetic positrons is the two-body decay of a previously undetected particle...” (Cowan et al. 1985, p. 2764).

Theoretical physicists also investigated such a possibility. Schäfer et al. concluded, “We have studied the hypothesis that the positron lines observed in heavy-ion collisions are caused by the decay of a neutral massive boson. Avoiding the need to postulate the existence of a whole family of new particles with adjusted masses and coupling constants, most possible cases were found to lead to conflict with established high-precision data of atomic physics. We were left with one viable alternative, namely the creation of a pseudoscalar particle from the nuclear current” (Schafer et al. 1985, p. L73). This particle would have a mass of 1.68 MeV, assuming the energy of the positron peak was 330 keV, and a lifetime in the range 5×10^{-13} s to 10^{-11} s. Balantekin et al. reached a similar conclusion. Their neutral particle would have a mass of 1.6 MeV and a somewhat shorter lifetime of approximately 1.3×10^{-13} s (Balantekin et al. 1985). Chodos and Wijewardhana reached a slightly different conclusion. They remarked that the neutral particle explanation for the positron peak required an appreciable probability that the particle was produced at rest in the center of mass system of the two ions, otherwise Doppler broadening would obliterate the peak. [63] “We shall find, in the simplest version of our model, that the production amplitude is actually suppressed for small values of the c.m. momentum k of the ϕ [the neutral particle], and we therefore cannot reproduce the observed peak. Another version of the model incorporates resonant behavior in the ion-ion system. Although we still find suppression at small k, we note that an alternative explanation of the peak is possible in which the ϕ is produced with a sharp nonzero value of k and then decays essentially at threshold into e^+e^-” (Chodos and Wijewardhana 1986, p. 302).

Everyone, both experimentalists and theorists alike, agreed that “a clear signal for a neutral particle could be provided by the detection of a monoenergetic electron in coincidence with the peak positrons” (Cowan et al. 1985, p. 1764).

B. Electron-positron states

The first report of a possible low-mass, neutral particle was by the EPOS I group (Cowan et al. 1986). This paper illustrated the dependence of the observed effects on selection criteria or cuts, discussed and illustrated earlier. In this experiment the EPOS I spectrometer had been modified so that both electrons and positrons could be observed in coincidence and their respective energies measured. Striking effects appeared, but only under certain circumstances. Figure 31(a) shows the intensity distribution for all

---

[63] All of the other speculations agreed with this conclusion. There is negligible difference between the center of mass system and the laboratory system because their relative velocity is only $1/20\ c$, where c is the speed of light.

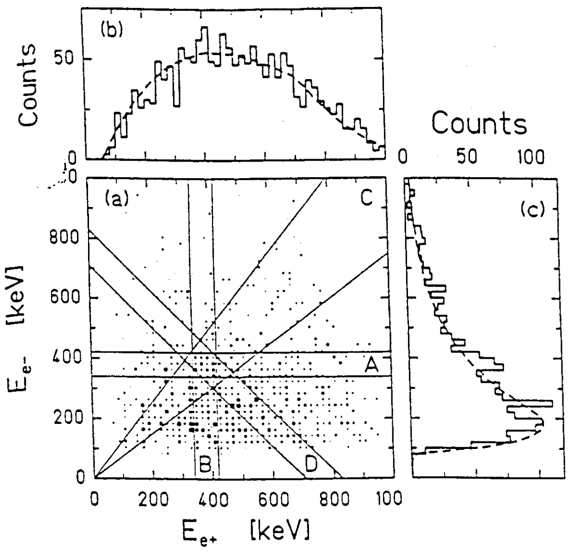

Fig. 31. Scatter plot of positron and electron energies along with the projections. The projections along the energy axes are shown along with the wedge cut (C), $E_{e+} \approx E_{e-}$, and the cuts for constant electron, positron, and sum-energy (A, B, D). From Cowan et al. (1986)

coincidence events as a function of the kinetic energies of the positron and of the electron, E_{e+} and E_{e-}, respectively, for U + Th collisions at 5.83 MeV/u. Figure 31(b) and 31 (c) show the projections of the distribution on the E_{e+} and E_{e-} axes, respectively. No structure is apparent. The curves shown are the result of a Monte Carlo calculation which included both nuclear and atomic processes. It is a good fit to the spectra.

When cuts were made on the data, however, structures did appear. Figure 32(a) shows the positron energy spectrum obtained when the energy of the coincident electron was restricted to the range $340 < E_{e-} < 420$ keV. Figure 32(b) shows the complementary distribution for electron energy when the positron energy was restricted to the same region. Both graphs show peaks at approximately 380 keV, with widths of approximately 80 keV, indicating that a significant fraction of the coincident events had electrons and positrons with the same energy. "The correlated yields of positron and electron peak events above the continuous background are 26.7 ± 7.7 and 31.7 ± 7.6 counts, respectively. The possibility that only a statistical fluctuation of the background produces the excess structure in the electron spectrum [Fig. 32(b)] at the specified energy of the positron peak is precluded at a statistical confidence level corresponding to 6σ [standard deviations]" (Cowan et al. 1986, p. 446).[64] Equal energies are what one would expect if both particles resulted from the decay of a very slowly moving particle.

[64] The question of how one should properly estimate such statistical confidence levels when cuts are applied will be an issue in the subsequent history. The probability of a six standard-deviation statistical effect is 2.0×10^{-9}.

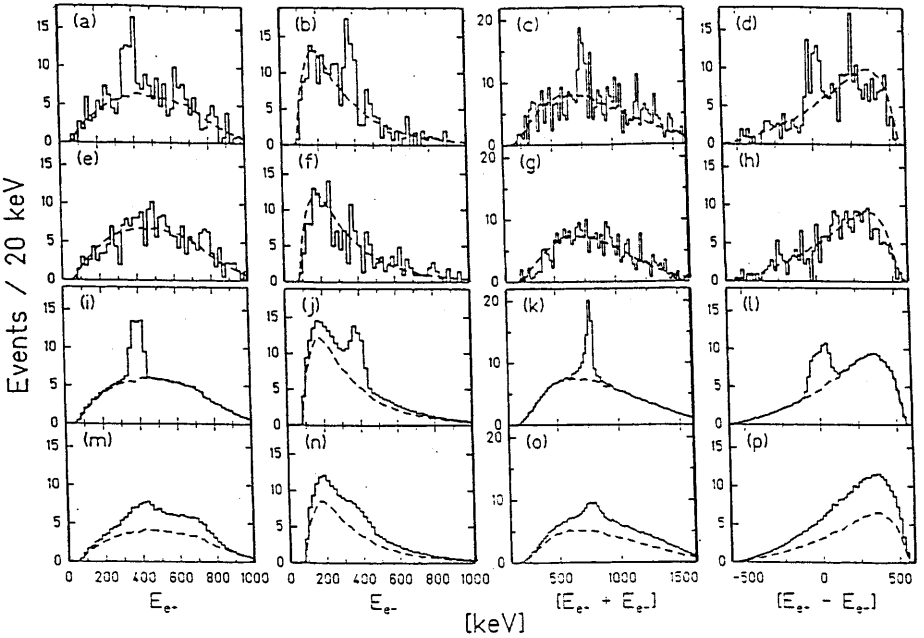

Fig. 32. Projections of the measured intensity distribution onto the E_{e+}, E_{e-}, $E_{e+} + E_{e-}$, and $E_{e+} - E_{e-}$ axes. The columns correspond to the gates A–D of Fig. 40. (i)–(l) and (m)–(p) are Monte Carlo calculations for two-body decay of a neutral particle and for internal pair conversion, respectively

The careful reader will note that this peak, although similar to those previously reported, is, in fact, different from them. EPOS I had previously found a peak at approximately 336 keV (Cowan et al. 1985) and the ORANGE group had found a peak at approximately 280 keV (Tsertos et al. 1985). The new peak was at 380 keV.

Figure 32 (c) shows the number of events as a function of the sum of the electron and positron energies, $E_{e-} + E_{e+}$. The experimenters required that $E_{e-} \approx E_{e+}$ which, when combined with the kinematic broadening of the energies, resulted in Cut C (the wedge cut, Fig. 31(a)). “The resulting sum-energy spectrum contains a narrow peak, at a mean energy of 760 ± 20 keV, with 35.3 ± 9.4 events in excess of the fitted continuous background” (Cowan et al. 1986, p. 446). Figure 32(d) is the intensity of events as a function of $(E_{e+} - E_{e-})$, the energy difference, requiring that $(E_{e+} + E_{e-})$ be constant (Cut D, Fig. 31(a)). A peak is also seen at $(E_{e+} - E_{e-}) \approx 0$.

None of these structures appeared in any of the projections for the energy regions adjacent to cuts A–D, on either side, (Fig. 32(e)–32(h)). They were quite sensitive to the energy cut. The authors were able to fit a Monte Carlo calculation to the observed peaks by assuming that neutral particle with mass 1.8 MeV was produced in the collisions (Fig. 32(i)–32(l)). Nuclear interactions such as internal pair conversion, the conversion of an emitted γ-ray into an electron-positron pair, which would also produce coincident electrons and positrons, were eliminated by Monte Carlo calculations (Fig. 32(m)–

32(p)). The EPOS I group had found that a significant fraction of the coincident electron-positron pairs had approximately the same energy and were given off at large angles ($\sim 180^\circ$) to one another.[65] This was compatible with the view that they were the result of the decay of a low-mass particle, produced at very low velocity in the center-of-mass system. They concluded that, "Features associated with the electron-positron decay of a slowly moving neutral particle appear to be reflected in the observations involving electron-positron coincidences" (Cowan et al. 1986, p. 447).

The EPOS I group presented new results in a paper presented at the NATO Advanced Study Institute on *Physics of Strong Fields*, Maratea, Italy, 1 -14 June 1986 (Cowan et al. 1987).[66] This paper included a history of the positron lines to that time, details of the modified EPOS spectrometer, the results on the low-mass electron positron states previously presented by the group (Cowan et al. 1986), and a detailed discussion of new results. We have discussed previously how the experimental results seemed to be sensitive to both experimental conditions and to selection cuts. This was made even more apparent in this paper.

The most striking new result was evidence for two additional low-mass electron-positron states at sum energies ($E_{e+} + E_{e-}$), of 620 keV and 810 keV, respectively. As was the case with the previously reported 760 keV state, the effects did not appear in all of the data. Observing them required selectivity. Figure 33 shows the new results. The original figure caption read, "Results of a preliminary analysis of U + Th collisions near 5.87 MeV/u (Feb 1986). ($E_{e+} + E_{e-}$) and ($E_{e+} - E_{e-}$) projections for two subsets of data gated on beam energy, heavy-ion scattering angle and e^+ or e^- TOF [time of flight] *chosen to enhance the prominent lines at* $\sim$*810 keV and* $\sim$*620 KeV, respectively*" (Cowan et al. 1987, p. 117, emphasis added). The data in (a) and (b) (the 810 keV state) were obtained with beam energy 5.87–5.90 MeV/u, and the TOF difference between the heavy-ion signal and the electron and positron signals set for "prompt" events. ("prompt" events are expected from the decay of a single particle). The data in (c) and (d) (the 620 keV state) had beam energy 5.85–5.90 MeV/u and with the electron TOF set for prompt electrons with the positron TOF delayed by an average of 3 ns. The sensitivity to cuts is clear. Only one state appears in each set of graphs.[67] They were sensitive to both the bombarding energy and to the time of flight.

The sensitivity of the observed effect to the TOF cut had been found empirically. The TOF cut was applied because another result had been seen earlier at one time delay, but not at others. No theory, at the time, predicted any sensitivity to the time of flight of the electrons and positrons. There were other such effects. The targets used tended to deteriorate with increased exposure to the high-energy heavy-ion beam. Some effects were seen only in data taken with a "fresh" target. Figure 34b shows a positron line observed only with data taken in the first hour of heavy-ion irradiation of the target. Figure 34a, the total data sample for the run, shows no effect.

---

[65] For a discussion of these issues see Cowan et al. (1987).

[66] The full proceedings of this institute appeared in Greiner (1987).

[67] The signal could also be enhanced relative to the background by making cuts on the heavy-ion scattering angles. See Figs. 3 and Fig. 32 in (Cowan et al. 1987) and the discussion on pp. 185–186.

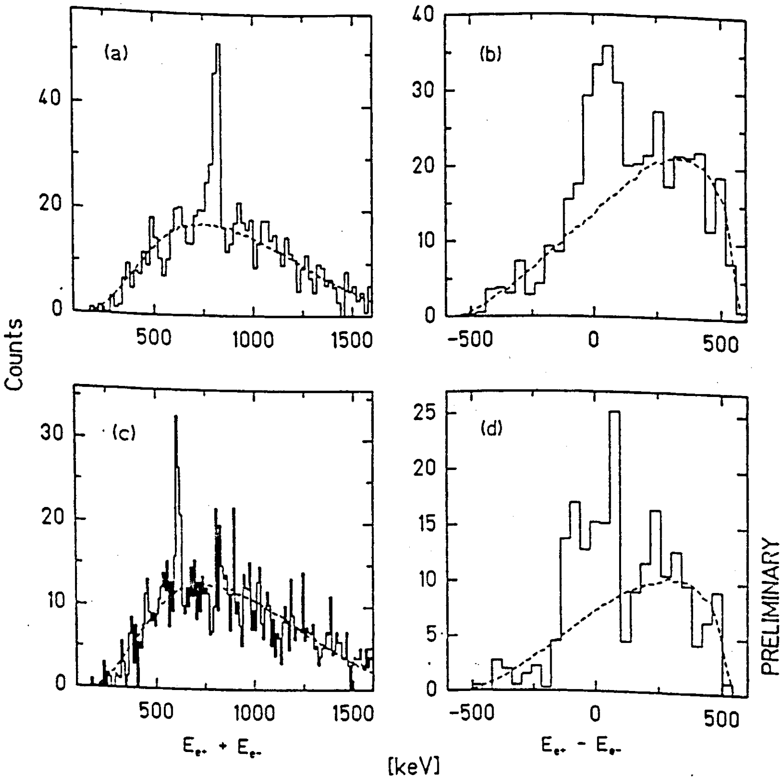

Fig. 33. Sum- and difference-energy spectra For U + Th collisons, with cuts chosen to enhance the lines at ~810 keV and ~620 keV, respectively (see discussion in text). From Cowan et al. (1987)

This illustrates and emphasizes the problem of cuts. In studying a rare, and previously unobserved phenomenon, one may very well have to apply cuts in order to see any effect at all. One might ask whether the observed effect is real, is an artifact of the cuts, or whether the experimenter is tuning the cuts to create the effect. Is the sensitivity of the result to beam energy due to a real resonance phenomenon, as some physicists suggested, or is it only a statistical effect enhanced by the cut on beam energy? One might ask similar questions for the time-of-flight and target exposure cuts.

Although the experimental results were compatible with the production of a new particle, no theory or model could actually explain the results. Reinhardt summarized the situation as follows

> After two years of speculation that the positron peaks and the $e^+ - e^-$ coincidence event discovered at GSI might be caused by elementary particle decay, no definite conclusion has been reached. In fact, theoretical and experimental aspects of this fascinating hypothesis have – to a large extent – developed in opposite directions. On the side of theory the hypothesis has lost a good deal of attractiveness because every single model which was quantitatively analyzed has

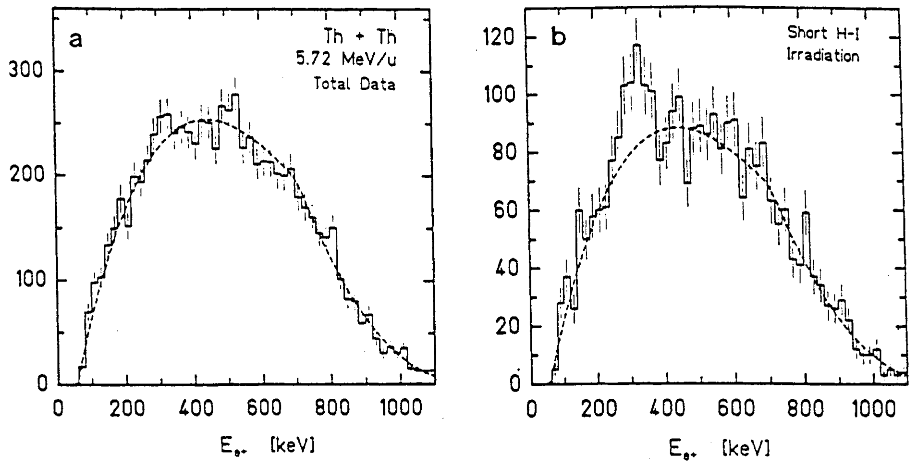

Fig. 34. (**a**) Total positron energy distribution for Th + Th collisions at 5.75 MeV/u and 5.70 MeV/u. (**b**) Subset of the data selecting approximately the first hour of irradiation. From Cowan et al. (1987)

turned out to be untenable. Either there was a clear contradiction to well-founded experimental data from atomic or nuclear physics, or the models fell far short of describing the GSI events. On the side of the GSI experiments, on the other hand, all evidence gathered during the past year is compatible with the hypothesis of particle decay, albeit of several distinct particle states, and many pieces of data clearly seem to suggest a two-body decay. It is true that all recent attempts to search for new light particles in nuclear or high-energy physics processes have been futile [see discussion above], and a first search for 2γ-coincidences in U + Th collisions has been unsuccessful. (Reinhardt et al. 1987, pp. 344–345).[68]

Despite the theoretical difficulties, the experimental results on both the positron lines and on the possible low-mass electron-positron state were sufficiently credible that other experimenters searched for similar effects in other interactions in which such effects might be expected to appear. All but one of these searches was negative. No trace of a new particle was seen in radiative upsilon decay (Albrecht et al. 1986; Bowcock et al. 1986; Mageras et al. 1986); electron bremsstrahlung (Davier, Jeanjean and Nguyen-Ngoc 1986; Konaka et al. 1986); nuclear decays (Baba et al. 1986; DeBoer et al. 1986; Hallin et al. 1986; Savage et al. 1986); muon or pion decay (Bryman and Clifford 1986; Eichler et al. 1986); or hadronic showers (Brown et al. 1986). One of the negative results, searching for a possible low-mass particle in U + Th collisions (Meyerhof et al. 1986)

[68] Not all such attempts were unsuccessful. A few months later Wong and Becker (1986), using strong, short-range magnetic forces, found an e^+ e^-resonance with a mass of 1.579 MeV and a lifetime of 4.2×10^{-19} s, resulting in decay positrons and electrons with kinetic energies of 279 keV. This was similar to the "observed" particle, which had a mass of 1.6–1.8 MeV, a lifetime greater then 10^{-21} s, and decay positrons and electrons with kinetic energies of 380 keV.

was the first experiment of this type performed at a heavy-ion accelerator other than GSI. This experiment was done at the SUPERHILAC at Berkeley.

The only positive result reported in these searches was that of a suggestive peak in the sum-energy spectrum for electrons and positrons produced in e^+ + Thorium collisions (Erb, Lee and Milner 1986). The experimenters were quite positive about the previously reported positron lines. "The peaks observed in the energy spectra from heavy-ion collisions raise a variety of important questions. Established beyond any doubt by a series of heavy-ion experiments their properties are puzzling and their existence appears to be unexplainable in terms of conventional or atomic phenomena" (p. 52). They concluded that "The width and energy of the peak structure in our data is similar to that found in the heavy-ion data [their sum-energy peak was at 670 keV, in contrast to the peaks found at 620 keV, 760 keV, and 810 KeV by the EPOS group], leading to the speculation that both may reflect a common underlying process" (p. 56).

The subsequent history of this particular effect is a microcosm of the entire episode. Attempted replications of the experiment gave positive, negative, and inconclusive results. There were, in addition, problems of interpretation. Peckhaus and collaborators, for example, also searched for a peak structure in e^+ + Th collisions and found no effect (Peckhaus et al. 1987). They noted, however, that the identification of the peak at 340 keV, the energy reported by Erb et al. (1986) was complicated by the Compton edge in the scattering of annihilation γ rays. Bargholtz and collaborators found a small, 2.5 standard-deviation, effect at the same energy, but remarked that the effects of multiple scattering and energy loss were large in their experimental apparatus and concluded that, "The evidence for this peak is not considered conclusive" (Bargholtz et al. 1987, p. L265). Wang et al. noted a problem with the original effect reported by Erb and collaborators. "Since the positron endpoint energy of the source is only 1.9 MeV, a resonance with a rest mass of 1.8 MeV cannot be produced with free electrons; however, it is kinematically possible that the bound electrons in the Th atom could participate. This leaves a substantial kinematic puzzle: the fact that the electron and positron peak energies are the same would require a resonant state that is nearly at rest in the laboratory. Except for multiple scattering, the decay of such an object would not allow the electron-positron pair to be emitted at 90° [the angle between the electron and positron spectrometers in the Erb experiment]" (Wang et al. 1987, p. 2136). (Particles resulting from the two-body decay at rest of a single particle should be emitted back to back, or at 180° relative to one another). Wang and collaborators also looked for e^+e^- coincidences in the scattering of positrons from Th, but in their apparatus the positron and electron spectrometers were back to back (180°), The configuration also minimized the background due to ordinary electron-positron scattering. They found no effect, but they did observe a Compton edge due to the scattering of annihilation γ rays.

On the other hand, Sakai et al. (1988) investigated positron scattering from Th, U, and Ta. They found evidence for a very narrow peak at ~330 keV with a width of 3.7 keV, but only for the Th and U targets. No effect was seen with the Ta target. There was no explanation for this unusual and unexpected difference. Tsertos pointed out, however, that the narrow peak could not be due to the decay of a particle produced by Bhabha scattering of positrons from bound electrons (Tsertos 1989). In this case the center-of-mass motion would broaden the peak to 60 keV. Sakai responded that Tsertos was correct, but that their result indicated that the decaying particle was produced at rest in the laboratory. He also reported that new data showed the existence of an energy peak at 409 keV, corresponding the sum-energy peak at ∼810 keV reported by the ORANGE group (Sakai 1989). To say the least, the situation was unclear.

The uncertainty concerning both the effects and their explanation was increased when ORANGE group continued its work on the single positron lines, extending it to subcritical systems, Pb + Pb, U + Ta, and U + Au, in which spontaneous positron creation was not expected. Nevertheless they found two lines, at energies of ∼258 keV and ∼340 keV, respectively. The peak energy of these lines was, once again, independent of Z_u, the total charge of the system. They found, however, that the cross-section for positron line production had a strong Z_u dependence, $\propto Z_u^{22\pm2}$. This was in disagreement with the results previously reported by the EPOS I group that both the production of positrons and the angular distribution was independent of the total charge of the system. There was, in addition, a factor of 10 difference in the production rates.

> It is worth mentioning that this strong rise of the line cross section with Z_u, found previously in the systems U + The and U + U and established with this work within a relative large Z_u-range is in disagreement with the results of the EPOS-collaboration, which favors a Z_u-independent line cross section of about 10 μb/sr for all systems studied between $180 \leq Z_u \leq 188$. *Especially our result for the 330 keV* e^+ *line in U + Ta (∼0.23 μb/sr) is difficult to reconcile with a ∼10 μb/sr cross section for an* e^+ *line at ∼375 keV in Th + Ta* ($Z_u = 163$) *collisions, as had been recently reported by the EPOS collaboration.*
> On the other hand, it should be emphasized again that the e^+-lines are observed by both groups under very similar kinematical conditions and, therefore, the assertion that the majority of these e^+-lines are not of the same nature, seems to be very improbable (Koenig et al. 1987, p. 141, emphasis added).

They regarded the qualitative similarities as more significant than the quantitative disagreement.

The ORANGE group also provided a summary of all the positron lines that had been observed up to that time (Table 4). They arranged these lines into three groups with mean energies of 255 ± 7 keV, 337 ± 6 keV, and 393 ± 5 keV, respectively. Note, however, that the measured line widths varied from 24–50 keV, 13− ≥ 100 keV, and 31–80 keV, for each of the lines. It is not at all clear that the energies and the widths of the lines are the same. One might also remark that the grouping of the measurements seems somewhat arbitrary. An equally good case could be made for a continuous spectrum or other groupings (Fig. 35).[69] The experimental situation with respect to these lines remained unclear.

The experimental uncertainty in the positron lines combined with the lack of any acceptable theoretical explanation of these lines or of the proposed low-mass electron-positron state and the failure to detect such a state in other systems, led theorists to

[69] A admittedly not-so-random survey of ten of my colleagues in the Physics Department indicated unanimous support for a random distribution of the positron line energies. They suggested no obvious groupings, but said that if forced to do so they would group the six lowest energy lines together, rather than the four used by the ORANGE group. They saw no evidence for any other groups.

Table 4. Positron lines observed From (Koenig et al. 1987)

E_{e+}^{cm} (keV)	FWHM (keV)	Collision System	Z_u	Bomb. Energy (MeV/u)	Apparatus	Reference
238 ± 10	50	U + U	184	5.6, 5.9	ORANGE	(Tsertos et al. 1987a)
250 ± 5	34	U + Ta	165	5.9	ORANGE	Present work
261 ± 4	26	Pb + Pb	164	5.7	ORANGE	Present Work
263 ± 5	24	U + Au	171	5.9	ORANGE	Present Work
Mean Value 255 ± 7						
277 ± 6[a]	65	U + Th	182	5.9	ORANGE	(Tsertos et al. 1985)
280 ± 6[a]	70	U + U	184	5.7, 6.2	ORANGE	(Tsertos et al. 1985)
310 ± 5	~ 25[b]	U + Th	182	5.85–5.90	EPOS, e^+e^-	(Cowan et al. 1986)
313 ± 10	~ 75	U + U	184	5.9	EPOS	(Cowan et al. 1985)
316 ± 10	~ 75	U + Cm	188	6.05	EPOS	(Cowan et al. 1985)
327 ± 10	~ 75	Th + Th	180	5.75	EPOS	(Cowan et al. 1985)
330 ± 4	~ 13	U + Ta	165	5.9	ORANGE	Present Work
334 ± 10	~ 75	U + Cm	188	6.07	EPOS	(Cowan et al. 1985)
337 ± 4	33	U + Au	171	5.9	ORANGE	Present Work
348 ± 10	~ 75	U + U	184	5.8	EPOS	(Cowan et al. 1985)
349 ± 10	≥ 100	Th + U	182	5.82	EPOS	(Cowan et al. 1985)
350 ± 5	39	Pb + Pb	164	5.7	ORANGE	Present Work
352 ± 4	34	U + U	184	5.6, 5.9	ORANGE	(Tsertos et al. 1987a)
354 ± 10	~ 75	Th + Cm	186	6.02	EPOS	(Cowan et al. 1985)
Mean Value 337 ± 6						
375 ± 10	~ 75	Th + Ta	163	5.78	EPOS	GSI Report
380 ± 5	~ 80[a]	U + Th	182	~ 5.86	EPOS, e^+e^-	(Cowan et al. 1986)
380 ± 10	≥ 80	U + U	184	5.9	TORI	GSI Report
405 ± 5	~ 40[b]	U + Th	182	5.87–5.90	EPOS, e^+e^-	EPOS Report
409 ± 5	31	U + U	184	5.6, 5.9	ORANGE	(Tsertos et al. 1987a)
Mean Value 396 ± 5						

[a] These values are not included in the determination of the mean values.
[b] Widths observed for the e^+e^--sum energy peaks

search for another interaction that might provide both an explanation of the effects and a possible experimental confirmation. The suggested interaction was Bhabha scattering, the scattering of electrons and positrons. Although some work on this problem at energies of a few MeV had already been done, and the predictions of quantum electrodynamics confirmed to an accuracy of approximately 10%, it had been a neglected area of study. The problems of the GSI results provided a motivation for more detailed study.

> In view of the present controversial interpretation of the GSI positron lines it is highly desirable to have an independent experimental confirmation of the existence and properties of the "X^0". The most unambiguous test could be expected from an experiment inverting the decay process, i.e. producing X^0 via

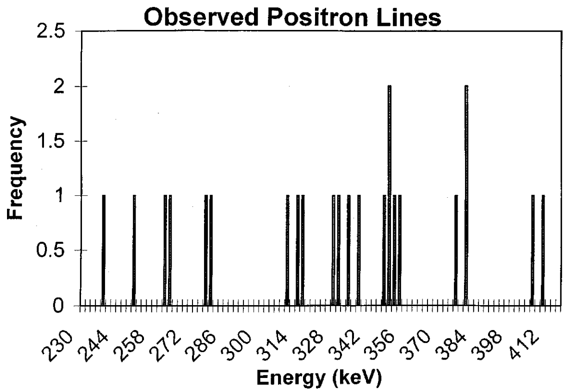

Fig. 35. Energy of the reported positron lines. Data from Koenig et al. (1987)

> $e^+ + e^- \rightarrow X^0$. *Thus narrow resonances should be observable in Bhabha scattering in the region* $E_{e+}^{kin} = 1.6 \ldots 2.3\ MeV \ldots$.
> The purpose of our paper is not to make predictions on the nature of the suggested resonances. Rather, we will collect some results on the strength and width of resonance in Bhabha scattering that can be expected quite independently of any specific model. (Reinhardt et al. 1987, p. 368, emphasis added).[70]

Experimenters were quick to investigate Bhabha scattering in the appropriate energy region. Once again, experiment gave conflicting answers. Six of the searches were negative. No resonant structures were seen in the results reported by (Mills Jr. and Levy 1987; Connell et al. 1988; Lorenz et al. 1988; Van Klinken et al. 1988; Tsertos et al. 1988a; Tsertos et al. 1988b).[71] Three experiments found positive, but small, effects. Maier and collaegues found a largest deviation from the fitted Bhabha spectrum of approximately 5% at an energy of 824 keV, in agreement with one of the lines found by EPOS I (Maier et al. 1987). This was confirmed in a later experiment by the same group which found an average enhancement of 1.6% at an energy of 810 keV (Maier et al. 1988). Von Wimmersperg and collaborators (Von Wimmersperg et al. 1987) found a deviation from the Bhabha fit of about 6% at an energy of 710 keV. They noted, however, that this result (and that of Maier et al.) was not easily reconcilable with the EPOS I observation of low-mass electron-positron states, which was that approximately 50% of the events produced in the energy region where the peak occurred were in the peak.

> The mass excess at 710 keV is compatible with the energy of the positron and electron lines observed in heavy-ion scattering within the energy resolution quoted,

[70] Reinhardt and collaborators also provided an explanation of the effect seen by Erb at al. (1987) in e^+ + Th scattering. "A first experiment claiming the observation of structures in e^+ + Th collisions was not sensitive to the kinematical conditions of resonance formation. It would have required the stopping of the X^0 in the target material to produce the line structures. Most probably, the observed effect was caused by background processes" (Reinhardt et al. 1987, p. 368).

[71] Connell et al. (1988) had investigated the reaction $e^+ + e^- \rightarrow 2\gamma$.

> but does not agree with Wong and Beckers calculation of 558 keV. Considering the prominence of the e^+e^- lines observed in heavy-ion scattering (of order 50% of the peak plus continuum sum), our observation of a 6% effect could only be considered as a possible source of these lines if a special mechanism favoring the production of such a resonance can be shown to exist in the environment of heavy-ion collisions. (Von Wimmersperg et al. 1987, p. 268).

1989 began with a flurry of new experimental results on the possible low-mass states. These were presented at the Moriond Workshop held 21-28 January 1989.[72] Once again, some results were supportive of the existence of such states, whereas others added to the confusion.

Kienle presented results obtained by the ORANGE group with a new experimental apparatus (Kienle 1989). (These results were published later as (Koenig et al. 1989)). The apparatus consisted of two orange-type beta-ray spectrometers placed back to back. One was set to detect electrons and the other positrons. In the total data set for electron-positron coincidences for U + U and U + Pb collisions at 5.9 MeV/u "no prominent structures were observed in these spectra" (Kienle 1989, p. 71). When the angle between the electron and the positron was restricted to $(180 \pm 20)^\circ$, and the difference in energy between the positron and the electron limited to -150 keV $\leq (E_{e+} - E_{e-}) \leq 0$ keV structures were observed in the sum energy spectra at energies of (540 ± 16), (640 ± 10), (716 ± 10), (809 ± 8), and (895 ± 10) keV, respectively (Fig. 36). These features appeared in both the U + U and the U + Pb collisions. No structures were seen for events with positive $(E_{e+} - E_{e-})$ energy differences. The experimenters concluded that,

> The present results from the sum energy coincidence spectra are consistent with the energies of e^+ lines which we presented before and which are shown in Fig. [37] for comparison. As can be seen from Fig. [36] the ensemble of lines, available in a wide range of collision systems suggests a family of resonances with invariant masses of ~1.54, ~1.66, ~1.72, ~1.83 and ~1.93 MeV/c^2 respectively. Two of these lines (1.66 and 1.83 MeV/c^2) may be identified with the sum peak at ~620 keV and ~810 keV found recently by the EPOS collaboration in U-Th collisions. *The line intensities in this work however are approximately by a factor of 10 lower compared to those of* (Cowan et al. 1987). *No indication of a sum line at 760 keV, as reported by the EPOS collaboration* (Cowan et al. 1986) *was found* (Kienle 1989, p. 74, emphasis added).

Once again, the observation of an effect was strongly dependent on the cuts applied. In this case there were plausible physical reasons underlying the cuts. In the new apparatus the electrons were detected in the forward direction relative to the beam, whereas the positrons were detected in the backward direction. This should result in an observed energy difference due to the Doppler shift caused by the center-of-mass motion of the particle produced.[73] The electron should have a higher energy that the positron. (Note,

[72] The Moriond Workshops provide a forum for speculative work in physics. Thus from 1987 to 1990 the Fifth Force, a proposed modification of Newton's law of gravity was extensively discussed at the workshops. For details see Franklin (1993)

[73] I am assuming that the center of mass of the prodcued particle is moving in the same direction as the overall center of mass of the heavy-ion system.

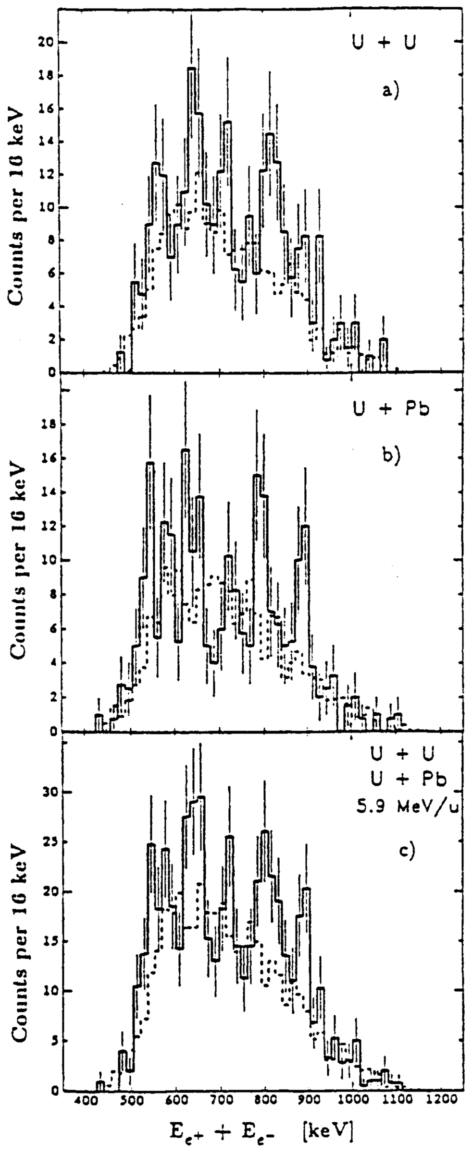

Fig. 36. Sum energy spectra from U + U and U + Pb collisions. From Kienle (1989)

however, that the original EPOS I result reported electrons and positrons with equal energies, the result one expected for the two-body decay of a very slowly moving object). In addition, the region of positive energy differences was where the background due to internal pair conversion, a nuclear process, was expected to be most pronounced. The angular restriction, $(180 \pm 20)^\circ$, was applied to select electrons and positrons resulting from the decay of a single particle. The decay particles should be emitted back to back or at 180° in both the center of mass, and in the laboratory because of the low center-of-mass velocity.

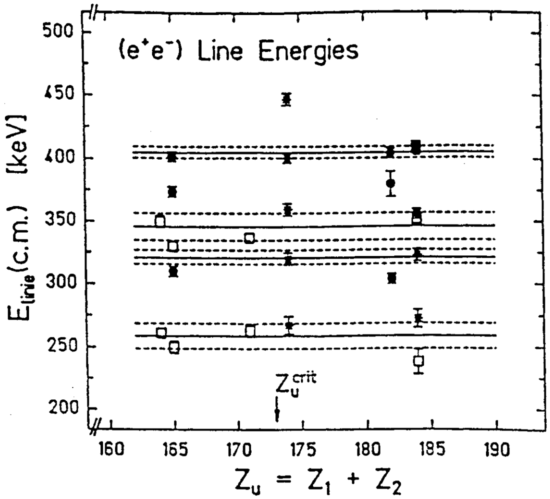

Fig. 37. Positron line energies as a function of the combined charge of the collision system. "For all systems studies, the data may suggest four common emission energies as indicated by the full lines" Kienle (1989)

The EPOS I group also presented new results at this workshop (Table 5). They found evidence in both U + Th and U + Ta collisions for resonances at 610, 750, and 810 keV. As before, as shown in Table 5, the observed effects were dependent on the experimental conditions, or the selection criteria. The resonances observed were quite sensitive to both the bombarding energy and to the time-of-flight difference between the electron and the positron. For the U +Th collisions, for example, the 610 keV resonance is barely visible in Fig. 38(a3), but quite visible in Fig. 38(b3). The difference between the two figures is that the bombarding energy for (a3) was 5.87–5.90 MeV/u, whereas for (b3) it was 5.86–5.90 MeV/u. Increasing the energy range accepted by only 0.01 MeV/u changed the character of the observed peak considerably. The 810 keV peak seen in (a1) (3σ level) was enhanced to a 6σ effect by restricting the time-of-flight difference to "prompt" events, that expected for the decay products of a single particle. The 610 keV peak, on the other hand, does not appear in the prompt events, but does appear in the

Table 5. Details of the sum-energy line family From (Bokemeyer et al. 1989)

Collision System	Beam Energy (MeV/u)	$E_{\Sigma ave}$ (keV)	ΔE_{Σ} (keV)	Counts
^{238}U + ^{232}Th	5.86–5.90	608 ± 8	23 ± 3	80 ± 13
	~5.83	760 ± 20	≤80	35 ± 9
	5.87–5.90	809 ± 8	42 ± 4	90 ± 14
^{238}U + ^{181}Ta	6.24–6.38	620 ± 8	20 ± 3	41 ± 9
	5.93–6.13	748 ± 8	26 ± 5	120 ± 19
	6.24–6.38	805 ± 8	20 ± 3	49 ± 12

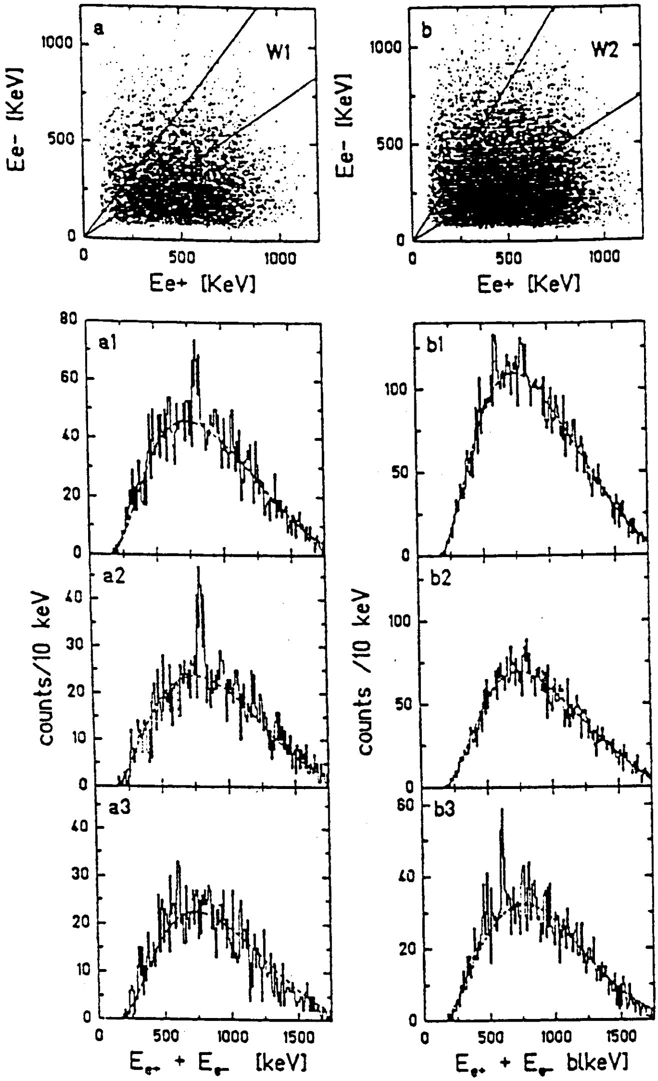

Fig. 38. Sum-energy spectra in ^{238}U + ^{232}Th collisions. Beam energy 5.87–5.90 MeV/u (left) and 5.86–5.90 MeV/u (right). The 810 keV state appears only for "prompt" events, whereas the 620 keV state appears only in "non-prompt" events. From Kienle (1989)

remaining non-prompt events. Similar effects of the cuts were seen in U + Ta collisions. In the U + Ta run an additional feature had been added to help decide if the electrons and positrons were due to the decay of a single particle. Counters were added to identify into which hemisphere (forward (F) or backward (B)) the decay particle was emitted. For a two-body decay of a single particle one expects the electrons and positrons to be emitted in different hemispheres (FB or BF). They found that only the 810 keV state was consistent with this condition. In fact, the 748 keV line seemed to be caused primarily by events in which both leptons were emitted into the forward hemisphere. They concluded that, "The presence of narrow electron-positron sum-energy lines in heavy-ion collisions seems to be undebatable. The persistency with which these lines occur in independent experiments–together with their statistical relevance reaching values up to 6σ–makes their interpretation in terms of statistical fluctuations rather unprobable, if not impossible. The absence of the 760-keV line observed in the first coincidence experiments in ^{238}U + ^{232}Th collisions in our more recent runs (presumably scanning the beam energy region of our previous experiment), however, is not easy to understand" (Bokemeyer et al. 1989, p. 88). Nevertheless, the EPOS results were inconsistent with both the new ORANGE results and with their own previous results.

Yet another experimental group reported results at the 1989 Moriond Workshop. The TORI group used a solenoidal spectrometer to detect both electrons and positrons produced in collisions U + Th at 5.85 MeV/u, almost identical conditions to those used by the EPOS I group in the experiment reporting their first positive result for a low-mass particle. The TORI apparatus could detect electrons in one of two conditions; the 0° condition ($0^\circ \leq \theta_{\mathrm{e+e-}} \leq 115^\circ$) and the 180° condition ($67^\circ \leq \theta_{\mathrm{e+e-}} \leq 180^\circ$). For the 180° condition, that expected for the decay of a single particle, and applying the "wedge" cut ($E_{\mathrm{e+}} - E_{\mathrm{e-}} \approx 0$), that used by EPOS I, they found no structure in their sum-energy spectrum. "A sum line with a strength as reported by the EPOS I group would appear with a yield $N_{\mathrm{e+e-}} \approx 67$ counts at the sum energy $E_{\mathrm{e+}} + E_{\mathrm{e-}} = (750 \pm 40)$ keV. This is obviously not observed" (Rhein et al. 1989, p.200). In contrast to the EPOS result, however, they did see structures in the 180° sum-energy spectra at 551 keV, 642 keV, and 749 keV, but only when the energy difference between the positron and the electron was not close to zero. In fact, the structures became more pronounced as the energy difference was increased.

Evidence against the existence of a low-mass particle in experiments on other interactions, was also presented at Moriond. Van Klinken summarized recent work on Bhabha scattering. He reported that five experiments – at Giessen, at Grenoble, at Munchen, at Groningen, and at Stuttgart– had all searched for a particle in the mass region around 1.8 MeV/c^2 and he concluded that "Searches have now been done with widely different and complementary approaches, leading to an overall conclusion that so far no resonances have been found within meaningful limits" (Van Klinken et al. 1989, p. 147). The Grenoble group presented their own negative result (Schreckenbach 1989), and negative searches in the reactions $e^+e^- \rightarrow n\gamma$ (Schreckenbach 1989) and in nuclear decay (Sona et al. 1989) were also reported.

The final EPOS I results were published in 1990 (Salabura et al. 1990).[74] They were essentially the same as those presented at the 1989 Moriond Workshop. The group concluded that

> Three very narrow e^+e^- sum-energy peaks around 610, 750, and 810 keV have been observed in U + Th as well as in U + Ta collisions at beam energies around the Coulomb barrier. As no processes involving conventional atomic and nuclear physics were found to describe their origin, the data were in particular confronted

[74] The UNILAC at GSI was shut down in 1989 for improvements

with the hypothesis that the lines are due to the two-body decay of neutral objects in an e^+e^- pair. Although the 810 keV sum-energy line observed in U + Th is consistent with the prompt two-body e^+e^- decay of a neutral object if created nearly at rest in the heavy-ion center-of-mass system, the other lines require at least a considerably more complicated scenario if they are to be explained in the context of a two-body decay (Salabura et al. 1990, p. 153).

Somewhat later the ORANGE group presented their last results before an upgrading of both their apparatus and of the accelerator (Koenig et al. 1993). They reported several lines in the energy range from ~550 keV to ~810 keV in the sum-energy spectra obtained from U + U, U + Pb, and U + Ta heavy-ion collisions. Not all of these lines were seen under all experimental conditions. As was the case for the EPOS I results, some of the lines were consistent with the two-body decay of a neutral object, whereas others were not.

Further evidence against the existence of such a neutral object was provided by continuing experiments on Bhabha scattering "Within statistical uncertainties (0.27%) no evidence has been observed for deviations from Bhabha scattering over the entire invariant-mass region 1560 keV/c^2 $< M_{xo} <$ 1860 keV/c^2 that can be associated with the e^+e^- sum-peak energies in the GSI heavy-ion experiments" (Wu et al. 1992, p. 1729). Another experiment using the same beam as Wu et al., but with an apparatus that eliminated all events due to Bhabha scattering, found no evidence for a low-mass particle (Henderson et al. 1992). Widmann et al. (1991) searched for such a particle in both the e^+e^- and the 2-γ ray final states and found nothing. Negative results were reported also by other groups (Judge et al. 1990; Tsertos et al. 1991).

Let us summarize the situation as of 1993. Discrete energy lines had been reported in the positron energy spectra obtained from heavy-ion collisions (Table 4). These lines had varying energy and did not appear in all of the systems studied, nor were they observed at all bombarding energies. They also seemed to be sensitive, at least on occasion, to other experimental conditions such as the scattering angle and to the exposure time of the target. There were also problems concerning the reproducibility of the observations under seemingly identical experimental conditions.

Similarly, peaks had been observed in the sum-energy spectra of electrons and positrons produced in such heavy-ion collisions. These results exhibited not only the same kinds of sensitivity as did the single positron lines (Table 4), but they also seemed sensitive to both the time of flight difference between the electron and the positron, to their energy difference, and to the angle between them. Some of these electron-positron peaks were consistent with the two-body decay of a neutral object, but others were not. No evidence of such a neutral object was seen in other interactions, particularly in the Bhabha scattering of electrons and positrons, in which one would also expect to observe effects if the sum-energy peaks were real.

Nevertheless the consensus seemed to be that something unusual and unexpected was being observed and that both the positron energy lines and the sum-energy peaks were real effects. Greiner and Reinhardt, two theoretical physicists who had worked extensively in the field, summarized the situation, both experimental and theoretical, as follows.

During the last decade the development of this field [spontaneous positron emission] was overshadowed by the spectacular narrow lines in the positron spectrum and later the monoenergetic electron-positron pairs discovered by the EPOS and ORANGE groups at GSI....

Table 6. Sum-energy lines observed in ORANGE and EPOS I experiments From (Bar et al. 1995)

Collision System	ORANGE	EPOS I	ORANGE	EPOS I	ORANGE	EPOS I	ORANGE	EPOS I
U + U	558 ± 8		630 ± 8				815 ± 8	
U + Th				608 ± 8		760 ± 20		809 ± 8
U + Pb	575 ± 6				787 ± 8	773 ± 10		
U + Ta			634 ± 5	625 ± 8		748 ± 8		
						740 ± 10		

This apparent universality of the positron lines has created much excitement and led to the belief that some fundamental new process had been discovered. A large variety of speculations, most of them based on very shaky ground, were put forward. The most natural explanation for a constant line energy and two-body decay characteristics would be the creation and subsequent decay of a new elementary particle, e.g. the axion. This, however, soon could be ruled out by various arguments, in particular by many control experiments (high-energy beam dump searches, pair production in nuclear transitions)....
However, a probably fatal blow was dealt at the hypothesis of a new particle, be it elementary or composite, by a set of experiments looking for resonances in electron-positron scattering in the mass region around 1.8 MeV. The outcome of these experiments (which are sensitive to resonances with a width down to the μeV level and have fully covered the relevant region of lifetimes) has been completely negative.

Thus one has to conclude that the GSI positron lines are only observable in experiments which involve heavy ions. Unfortunately the experimental results have changed considerably over time and it is difficult to decide which of the data are to be considered reliable (Greiner and Reinhardt 1995, pp. 217–218).[75]

Experimentalists were more positive. A 1995 paper published jointly by the ORANGE and EPOS II groups offered the following summary.

> Narrow sum-energy lines have been identified in $^{238}U+^{238}U$, $^{238}U+^{232}Th$, $^{238}U+^{208}Pb$, $^{238}U+^{206}Pb$, and $^{238}U+^{181}Ta$ collisions in EPOS and ORANGE experiments. The sum energies of the lines depend on the collision system and kinematical parameters. The lines group around $\sim$550, $\sim$620, $\sim$740, and $\sim$810 keV if the assumption of a common origin can be made. The observed cross sections vary from $d\sigma_{e+e-}/d\Omega^{CM}_{ion} \simeq 0.1$ μb/sr (815 keV, $^{238}U + ^{208}Pb$) to 3.6 μb/sr (748 keV, $^{238}U + ^{181}Ta$). The statistical significance of the lines reaches up to 6.5 σ (634 keV), but in some cases is limited. The experimental knowledge on production and decay channels is poor (Bar et al. 1995, p. 241, see Table 6).

[75] Greiner and Reinhardt noted that, "The situation may resemble another long-standing experimental puzzle at the beginning of the century when unaccountable narrow lines were observed in the radiation spectrum from the sun. Hypothetical new elements (nebulium and coronium) were invented to explain these lines. It took about three decades until it could be shown that they originate from transitions involving metastable states in highly ionized atoms in the sun's corona" (Greiner and Reinhardt 1995, p. 218)).

The APEX group (ATLAS Positron Experiment), a new player in the field, also agreed.

> Line structures have been reported in the energy spectra of positrons produced in collisions of very heavy ions. More remarkably, sharp lines have also been observed in the sum-energy spectra of coincident positrons and electrons. In the experiments, carried out at the GSI UNILAC accelerator, lines were reported for a number of scattering systems over a range of bombarding energies near the Coulomb barrier. The features of these observations suggest the possible formation and decay of some hitherto unknown light particle or composite object. (Ahmad et al. 1995b, p. 2658).

It is clear that, at the very least, the EPOS and ORANGE results were regarded as sufficiently credible to merit further investigation.

C. The search ends

Papers published in early 1995 by the EPOS II and ORANGE groups (Bar et al. 1995) and by the APEX collaboration (Ahmad et al. 1995a) marked the beginning of the final act of the drama and gave hints of what was to come. The joint EPOS II-ORANGE paper described the improvements that had been made in both experimental apparatuses. In particular, EPOS II was now able to detect both positrons and electrons in either side of their detector, increasing the yield of events. The groups reported results for an experimental run on the reaction $^{238}U + {}^{181}Ta$ at heavy-ion bombarding energies of 5.98–6.07 MeV/u. EPOS I had previously reported a sum-energy peak at $\sim$748 keV in this reaction (Table 5). The EPOS II group reported a similar peak in their new run. "A sum-energy line around 740 keV is identified in the first run, with an energy uncertainty of $\sim$10 keV (Fig. [39]). The line is poorly visible on the total spectrum of 45000 pair events which can be fully described otherwise by a MC-calculation [Monte Carlo] based on quasi-atomic and nuclear pair production, reproducing all global dependences established by the previous experiments" (Bar et al. 1995, p. 242). Making cuts on the

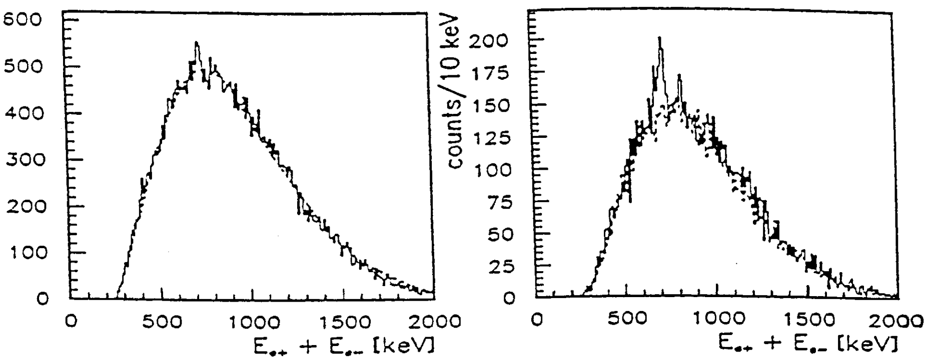

Fig. 39. Sum-energy spectra for U + Ta collisions at 5.98 and 6.07 MeV/u. Total spectrum (left) and selected for scattering angle (right). From Bar et al. (1995)

data requiring an interaction distance larger than 18.6 fm (this is related to the heavy-ion scattering angle) and a positive energy difference between the positron and the electron enhanced that peak to a 5.5 σ effect (Fig. 39, right-hand side). In a foreshadowing of later events, however, the peak did not appear in a subsequent run under seemingly identical circumstances. "In the second run with $\sim$3 times more total pair events the existence of a comparable line is not evident. Changes in the experimental set-up are presently investigated to clarify if these could influence the observation of the line. Nevertheless, this apparent inability to properly set the experimental conditions again supports the assumption that important parameters for the source of the lines still remain unidentified" (Bar et al. 1995, p. 242). (A similar effect is shown in Fig. 40. A peak that is visible in an early run is not seen in a later run which had more statistics and was taken under identical conditions and used identical cuts). The groups did not interpret their failure to reproduce the observed effect as casting doubt on the reality of that effect, or as suggestive evidence that the observed effect might be an artifact created by the cuts, but rather as demonstrating their lack of knowledge of all of the important experimental conditions and their inability to reproduce them.

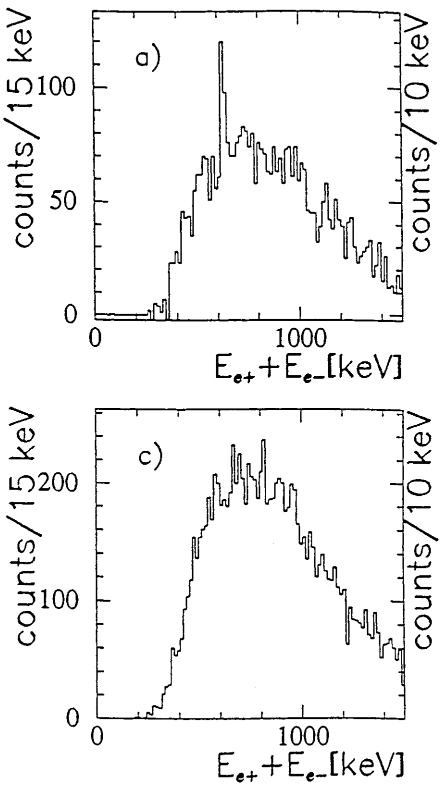

Fig. 40. Sum-energy spectra obtained by EPOS II. The peak that appears in the early data (top) disappeared in the larger total data sample. From Ganz (1995)

The APEX group, in a contiguous paper, reported preliminary results. They also examined the reaction $^{238}U + ^{181}Ta$ at energies comparable to those used by EPOS II. They found no peaks in either the positron energy spectrum or in the sum-energy spectrum. "The [positron energy] spectra are smooth within statistics. For those events with an electron detected in prompt time coincidence, positron-electron sum-energy spectra are shown.... They are also smooth showing no statistically significant deviations. A wide variety of analyses of these data have been carried out, selecting on heavy-ion scattering, lepton energies, etc. In all of these no statistically significant evidence has been found for sharp sum energy lines" (Ahmad et al. 1995a, p. 254).

The APEX group presented more detailed results later in 1995 (Ahmad et al. 1995b). They investigated the positron-electron sum-energy spectra in both the $^{238}U + ^{181}Ta$ system at bombarding energies of 5.95, 6.10, and 6.30 MeV/u and the $^{238}U + ^{232}Th$ system at 5.95 MeV/u. These were the systems and the energies at which both the ORANGE and the EPOS I groups had previously reported peaks.

> A variety of analyses have been carried out on the present data sets. These have not resulted in any statistically significant evidence for sharp sum-energy lines [Fig. 41]. The analyses ranged from systematic searches for peak structures by gating [selecting] on parameters such as heavy-ion scattering angle, positron and electron difference energy, positron-electron opening angle, and positron emission angle with respect to the beam direction to searches based on proposed physics scenarios and, to the extent possible, duplication of published analyses (Ahmad et al. 1995b, pp. 2659–2660).

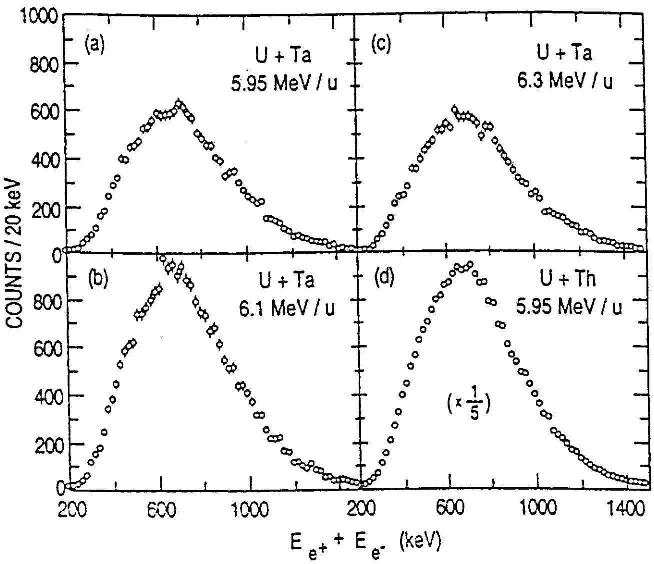

Fig. 41. Sum-energy spectra for positrons and electrons measured in coincidence with two quasielestically scattered ions. From Ahmad et al. (1995b)

The group applied selection criteria to attempt to observe the sum-energy peaks previously reported at 760 keV (Cowan et al. 1986) and at 809 keV (Salabura et al. 1990) in the $^{238}U + ^{232}Th$ system. These peaks had also been reported to be consistent with the two-body decay of a neutral object. Their results are shown in Fig. 42. No hint of any peak is seen. The events in the upper curve of Fig. 42a are those selected by the "Wedge Cut" (approximately equal electron and positron energies) previously used by EPOS I and ORANGE. The histogram is the spectrum of uncorrelated pairs generated by summing the energies of electrons and positrons from different events (event mixing). "The dashed peak, superimposed on the event-mixed spectrum, corresponds to the signal expected from the decay of an isolated neutral object of mass 1.8 MeV/c^2, produced with the cross section given in [Salabura et al. (1990)] ($d\sigma/d\Omega_{HI}$ ~5 μb/sr – the pair production cross section averaged over the heavy-ion detector acceptance)" (Ahmad et al. 1995b, p. 2660). Further cuts on the solenoid azimuthal angle and energy correlations expected for two-body decay are shown in the lower curve of Fig. 42a. Once again, the

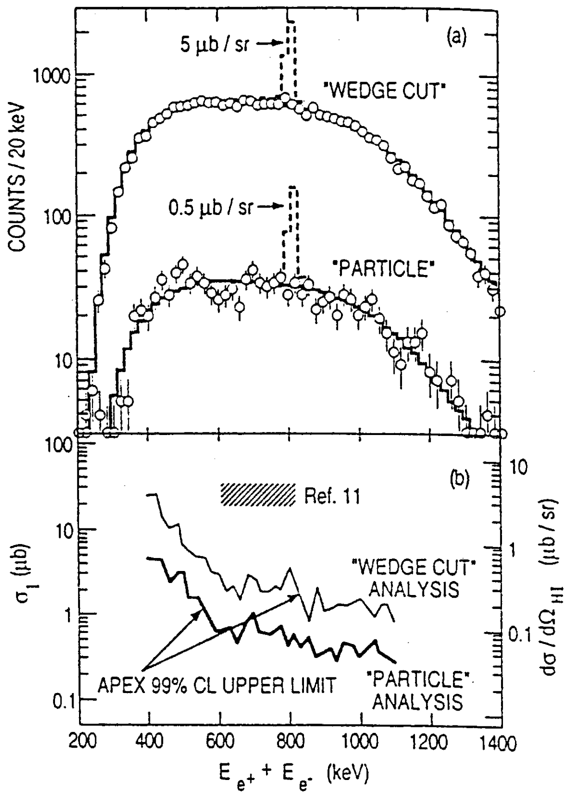

Fig. 42. Sum-energy spectra for ^{238}U + ^{232}Th at 5.95 MeV/u analyzed according to the expectations for the isotropic decay of a particle produced at rest in the center of mass. The "particle analysis" was for events with positrons and electrons with an opening angle of approximately 180°. From Ahmad et al. (1995b)

effect expected on the basis of the previous results was not seen. Nor was the 748 keV peak previously reported in the U + Ta system observed.

"The absence of the reported sum-energy lines in our data is puzzling. The origin of this apparent discrepancy may lie in so far unknown characteristics of the phenomenon. The overlap between the acceptance of APEX and that of the previous experiments is large. Nevertheless it is conceivable that the energy and angle correlations of the lepton pairs are such that they escape detection in our apparatus, although rather extreme situations are required for this to occur" (Ahmad et al. 1995b, 2661). The APEX group examined other possibilities for explaining the discrepancy and concluded, "Nevertheless, we believe that the results of the present experiment represent a real disagreement with the previous observation" (Ahmad et al. 1995b, p. 2661).

Strong evidence against the existence of the sum-energy peaks was also provided by the EPOS II group (Ganz et al. 1996). "The experiments with the new EPOS II spectrometer performed so far were devoted to the reinvestigation of the U + Ta and U + Th collision systems and the question of the reproducibility of the e^+e^- sum energy lines reported by EPOS I (Cowan et al. 1986; Salabura et al. 1990). Therefore special care was taken to subtend as closely as possible the beam energy range investigated by EPOS I" (Ganz et al. 1996, p. 6). In this run EPOS II obtained far more data than had been acquired by EPOS I; 10 times more for the U + Th system and 25 times more for the U + Ta system, respectively. In analyzing their data they used cuts identical to those used by EPOS I in both the beam energy and the "wedge cut," the energy sharing between the electron and the positron. They omitted the time-of-flight cut because that would have complicated the comparison between the old and the new results. They noted that the EPOS I peaks had been visible using only the two cuts that EPOS II was using. The EPOS II results for the three most prominent peaks observed by EPOS I; the 608 and 809 keV lines in the U + Th system and the 748 keV line in the U + Ta system are shown in Fig. 43. "No evidence for narrow line structures at sum energies given by the EPOS I experiment has been observed" (Ganz et al. 1996, p. 7). "As indicated in the right column of Fig. [43], where the differences between the measured sum-energy spectra and the normalized event-mixing distribution are displayed, the expected line yields are clearly in contradiction with the new high-statistics data" (Ganz et al. 1996, p. 9). The expected yield depends on the model of the decay of the object chosen. For peaks that were consistent with the two-body decay of a neutral object the EPOS II group used a model in which the electron and positron were emitted back-to-back. For the other peaks they chose a model in which the opening angle between the electron and positron was isotropic. (See discussion in Ganz et al. (1996, pp. 9–10)). The results shown in Fig. 43 are for an isotropic model and Fig. 44 shows a similar plot, but one in which the expected yield is calculated for a back-to-back model. Both of these models were consistent with the previously reported results for the particular systems.

They concluded, "In summary, all our attempts led to the same result. We did not succeed in reproducing with our statistically improved data basis the e^+e^- sum energy lines reported in EPOS I" (Ganz et al. 1996, p. 10).

The difference between the EPOS I and II results was clearly troubling to the group and they discussed various possibilities for explaining the discrepancy. They remarked that the APEX results were consistent with those of EPOS II, and that they covered an angular range for electron-positron emission that was forbidden in EPOS II, eliminat-

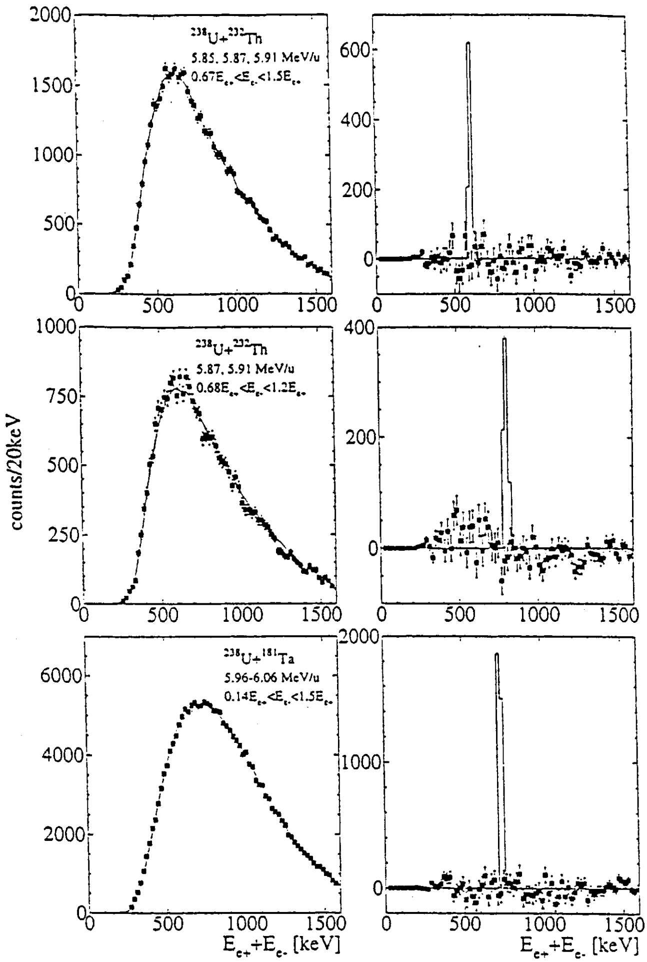

Fig. 43. Sum-energy spectra obtained by EPOS II. The lines in the right panel are the yields expected for the 608, 748, and 809 keV states found previously by EPOS I. From Ganz et al. (1996)

ing that as a possible explanation of the discrepancy. They also concluded that neither differences in acceptance, unknown experimental parameters, or background processes could explain the difference. Still the question remained. Why had EPOS I observed peaks in the sum-energy spectrum?

The EPOS II group noted that judging the statistical significance of peaks enhanced by selection criteria, or cuts, was an extremely difficult problem. They remarked on the sum-energy line that they had reported earlier (Bar et al. 1995).

> The 723 keV line [the energy calibration had changed between the two experiments] was only weakly visible in the total e^+e^- sum-energy spectrum but clearly seen (with a nominal significance of 5σ) when requiring cuts with respect to E_Δ

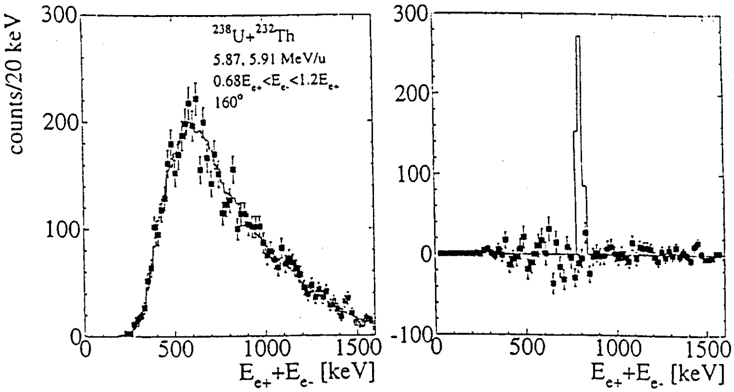

Fig. 44. Sum-energy spectra for $^{238}U + {}^{232}Th$ for leptons with an opening angle of approximately 160° found by EPOS II. The thin line in the right panel is the expected yield of the 809 keV line observed by EPOS I. From Ganz et al. (1996)

> [the energy difference between the electron and the positron] and the scattering angles of the heavy ions. However, not only was the energy of the line at variance with the previous value of 748 keV, also two follow-up, high-statistic experiments performed with an unchanged EPOS II set-up failed to reproduce the line [A similar effect for the line at 620 keV is shown in Fig. 40] (Ganz et al. 1996, pp. 10–11).

The group further investigated the possibility that the effect was due to the cuts, suggested by the disappearance of the line with higher statistics.

> The second example comes from an investigation suggested in Ref. 17 (Roe 1992) where we took advantage of the enlarged data basis collected in the EPOS II experiment. We randomly distributed – on an event-by-event basis – the e^+e^- pairs collected at a certain beam energy into two subsets. While one of these subsets was kept as a reference sample, we searched for narrow line structure in the other subset by choosing different E_Δ and time-of-flight cuts. Surprisingly enough, we were able to find a cut – leading to a spectrum of similar statistics as in a typical EPOS I experiment – which enhances a 2σ-structure visible at 655 keV in the initial subset to a line of $\approx 5\sigma$, which is comparable in width and intensity to those observed in EPOS I.[76] However, applying the identical cut to

[76] The availability of powerful, high-speed computers makes it possible to analyze data with different cuts in a very short time. This can be crucial in detecting a small signal, but it also has dangers, as we have seen. The ability to vary cuts easily and quickly is shown in Figure 45. The graphs show the mass distribution for kaons and pions in the mass region of the D^0 meson. The top row shows the effect of a cut on L/σ, the distance from the primary interaction to the decay vertex divided by the uncertainty in that distance. The prominence of the D^0 peak is enhanced as L/σ gets larger. The bottom row shows those events in which a definite kanon identification has been made. Applying the L/σ cut enhances the peak even more.

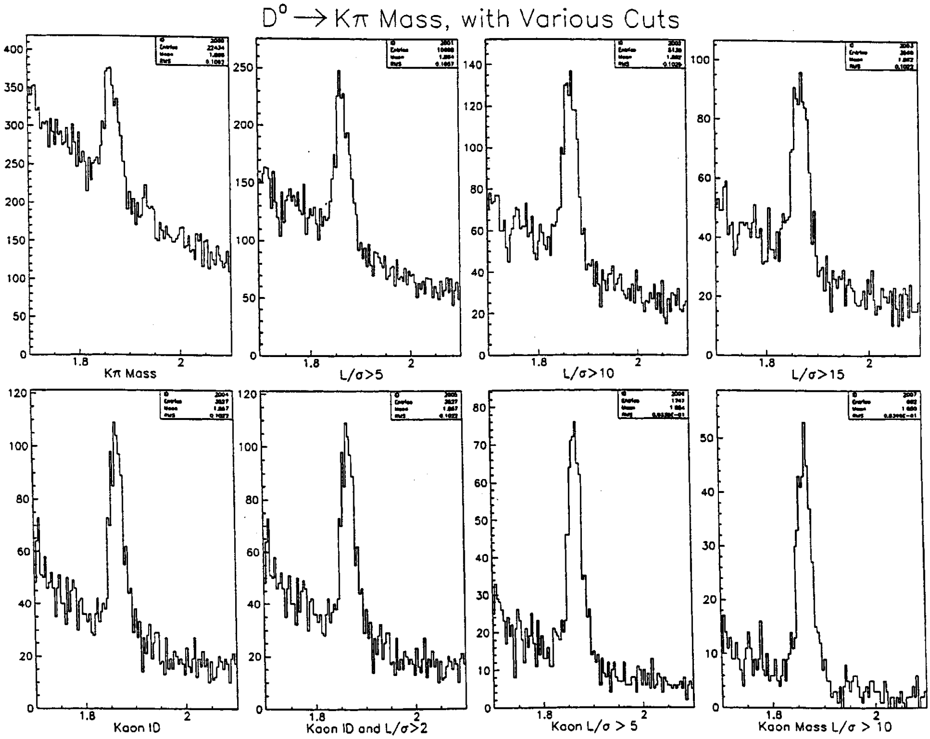

Fig. 45. Graphs of the D° mass obtained for various cuts. Courtesy of Brain O'Reilly

the reference sample does not show any line structure at this energy [Fig. 46]. (Ganz et al. 1996. p. 12).

They concluded that

Both examples [the disappearance of a peak with higher statistics and the division of the data set into two subsets] underline that the statistical significance of spectra obtained by introducing selection criteria, which are acceptable when looking for something unexpected but which cannot be supported later by a coherent physical picture, has to be taken with great precautions. In this situation an independent reproduction based on a considerably larger data set is the only way to confirm the existence of a physical effect. Since we failed to demonstrate the reproducibility of the lines observed by EPOS I and derived cross-section limits which are a factor of up to 10 smaller than the values implied by the previous results, and in view of the negative results obtained by the APEX Collaboration the physical relevance of the EPOS I lines is questionable (Ganz et al. 1996, p. 12, emphasis added).

The ORANGE group also repudiated their earlier results (Leinberger et al. 1997). Their new experiment was designed to look for the ~635 keV peak that was their most statistically significant result (6.5σ). "At improved statistical accuracy [by more than a

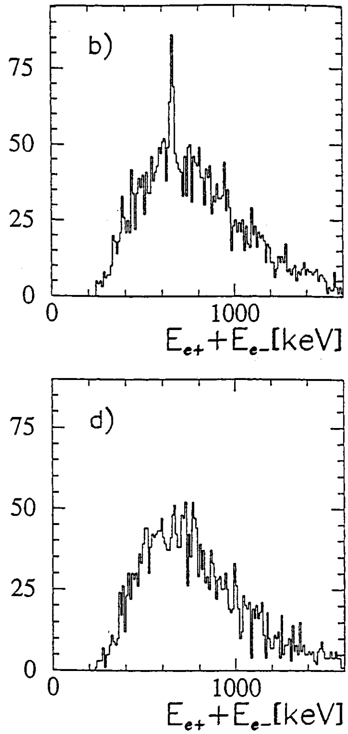

Fig. 46. The upper graph shows the sum-energy spectrum obtaind by tuning the energy-differnce and time-of-flight cuts to maximze the peak. The bottom graph shows the effect of applying the identical cuts to the other half of their data. No peak is visible. From Ganz et al. (1996)

factor of 10], the line couldn't be found in the new data" (p. 16). (Fig. 47). The experimental conditions and the cuts used in the analysis of the data were extremely close to those used previously. The group also eliminated other possible differences in experimental conditions, such as target deterioration, as possible causes for the difference between the old and new results. "Taking into account that we have not found any evidence that the reported line [the earlier result] might be due to trivial effects or background processes, its statistical significance has to be reconsidered" (p. 21). Analysis of their new data had shown that the calculated background used in their earlier experiment was incorrect. Using the new background calculation[77] the previously reported effect was reduced from 6.5σ to $\leq 3.4\sigma$.[78] ORANGE also found, in agreement with EPOS II, that manip-

[77] The previous background calculation fit a polynomial to that part of the spectrum that was not near the observed peaks. The new calculation used a spectrum obtained from event mixing, in which an electron from one event was combined with the positron from another randomly selected event.

[78] Although the occurrence of a 3.4σ effect in a single bin is quite small (the probability is 0.064%), the actual probability is much larger. One has to take into account both the number of

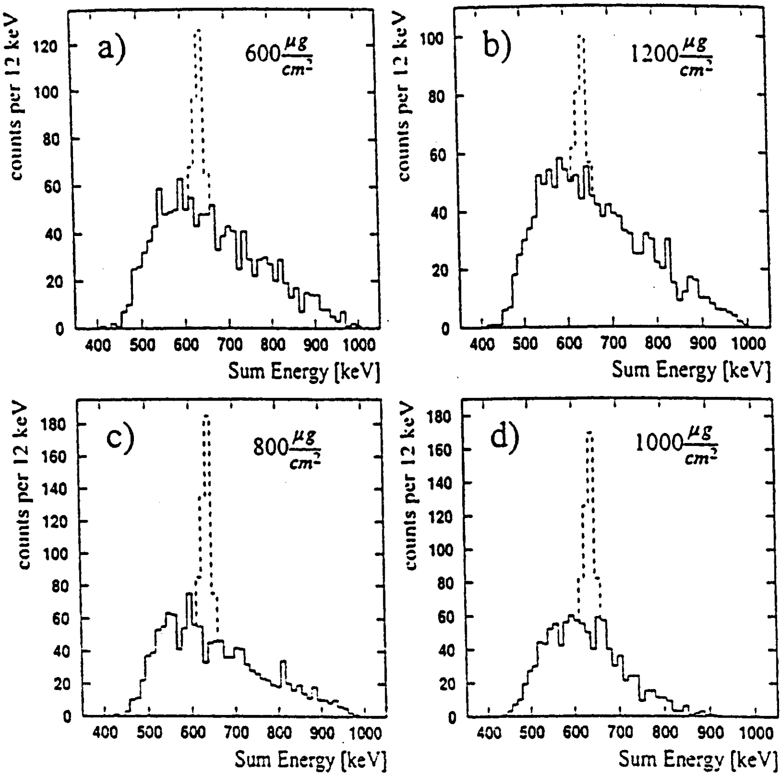

Fig. 47. Sum-energy spectra from U + Ta collisions found by the ORANGE group. The superimposed peaks are those expected for the 635 keV state they had reported previously. From Leinberger et al. (1997)

ulating the cuts could produce electron-positron lines that disappeared with improved statistics. "We shall note in this context that in some runs we found an indication for a

---

bins in the graph as well as the number of graphs plotted. This was illustrated during the 1960s when the search for new particles and resonances using bubble-chamber techniques occupied a substantial fraction of the time and effort of those working in the field. The usual technique was to plot the number of events as a function of the invariant mass and look for peaks above a smooth background. This is quite similar to the technique used to look for peaks in the sum-energy graphs. The usual informal criterion for the existence of a new particle was that it give a 3σ effect above background. This had a probability of 0.27 percent. That was extremely unlikely in any single experiment, but Professor Arthur Rosenfeld of the University of California at Berkeley is supposed to have pointed out that if one considered the number of such experiments done each year and the number of graphs drawn, then one would expect to observe a significant number of such 3σ effects just on statistical grounds. The informal criterion was then changed to 4σ, which had a probability of 0.0064 percent. The story may be apocryphal, but it did have wide circulation among high-energy physicists at the time. For further discussion see Roe (1992, pp. 111–112). As we have already seen, estimating the statistical significance of a peak when selection cuts are being applied is even more difficult.

narrow e^+e^--line in the sum-energy spectra by applying slightly different cuts in the HI [heavy-ion] kinematics. This line, however, could not be reproduced in the subsequent runs taken under similar experimental conditions and better statistics" (p. 21).

Even before the publication of the new ORANGE results, the APEX and EPOS experiments had provided the evidence that convinced virtually everyone working in the field that the sum-energy peaks previously reported by EPOS I, ORANGE, and others did not exist. In June 1996 a meeting was held at Oxford attended by physicists from the EPOS II, ORANGE, and APEX groups. At the end of that meeting the conclusion was that the search was over and that no further experiments were needed.[79]

D. Discussion

The physics community, particularly those working on heavy-ion collisions, has concluded that both the positron lines and the sum-energy peaks observed by EPOS I, ORANGE, and others are not real effects. There is a strong suspicion that they are artifacts created by tuning the selection criteria applied to the data to produce the experimental results, combined with the effects of limited statistics. There is, in fact, a suggestion that such tuning occurred in the acquisition of data in the early experiments.

> Take what the EPOS physicists referred to as the top-hat criterion. Bokemeyer says that the EPOS physicists had noticed that what turned out to be peaks in the final analysis would first appear online as a top-hat shaped bulge in an otherwise smooth spectrum. So the experimenters would start collecting data at a particular energy or with a particular target, and if the spectra were smooth and flat, they would stop the experiment. "We would change the energy or target and try again," says Bokemeyer. "When the spectra started to look like a top-hat, this seemed to be the correct [conditions], and we would continue running without interruption" (Taubes 1997, p. 151).

---

[79] Not everyone in the community of those who worked on the experiments agreed. Greenberg and Cowan criticized the APEX result on the grounds that their energy range was too large and that APEX had overestimated the effect that should have been seen in the APEX experiment on the basis of the EPOS I result (Cowan and Greenberg 1996). APEX disagreed (Ahmad et al. 1996). Greenberg, although a member of the APEX collaboration, withdrew his name from the publication, and has presented a reanalysis of the APEX data which he claims shows peaks similar to those observed earlier. The problem is that the observation of this effect also requires cuts and one might question whether this result is also an artifact produced by the cuts. Griffin has also questioned the conclusions reached by the APEX collaboration and has suggested that there is, in fact, a small peak in their sum-energy spectrum (Griffin 1995; Griffin 1997a; Griffin 1997b). No criticism of the EPOS II result has been published. Greenberg has requested that APEX continue the search, but the group is not willing to do so. Since the publication of these latest results by EPOS II, APEX, and ORANGE further evidence against the existence of the sum-energy peaks and the positron lines has appeared in (Faestermann, Heine and Kienle 1996; Ditzel et al. 1997; Ahmad et al. 1997a; Ahmad et al. 1997b). In particular, the experiments have investigated the question of whether the peaks observed are due to the internal conversion of the γ rays from nuclear transitions. No effects have been seen.

As we have seen in the cases presented earlier, experimenters do not use all of their data in producing a result. Cuts, or selection criteria, are always applied. This is not an unreasonable procedure. When one is looking for a small effect against a much larger background, cuts are needed to enhance the signal. (Recall the K^{+}_{e2} branching ratio experiment). One might, however, legitimately worry that the cuts are being tuned to enhance the effect. We have seen that it is possible to create a peak by tuning the cuts and we have also seen the safeguards taken to guard against this.

The problem becomes even more difficult when, as it was in this episode, the result may depend on a large number of parameters, each of which may be used for selection, and when theory provides no guidance as to what the important parameters might be. Cuts may very well be required to enhance the signal so that an effect can be observed. Greenberg argued that this was indeed a justifiable procedure. "With any kind of new phenomenon, you try to see what conditions would optimize it. Obviously not having any theoretical guidance, we tried various things that would optimize the appearance of the peak" (quoted in Taubes (1997), p. 151). The evaluation of the results was also made more difficult by their apparent, albeit not exact, replication. The question that must be answered, however, is whether the result is real or is an artifact created by the cuts.

The physicists involved in this episode were quite aware of this problem from the very beginning. Recall that Bokemeyer had urged caution in interpreting the first report of positron lines. "However, before drawing any far reaching conclusions from the experimental data presented, we feel that the following questions should be solved. *First of all one must show by further analysis that the structures are not produced by some yet unknown background effects associated with one of the event selecting criteria*. Secondly, the reproducibility of the effect has to be shown" (Bokemeyer et al. 1983, p. 290, emphasis added). The many repetitions of the experiments under both very similar conditions and under varying conditions were attempts to establish the correctness of the results by showing that they were reproducible and were also efforts to acquire a physical understanding of the systems involved. Unfortunately, rather than clarifying the situation, they made it more complex.

The discord was resolved when two experiments with higher statistics, and therefore more evidential weight, those of EPOS II and APEX, found no evidence for the previously reported results. (This was supported by subsequent ORANGE results). In addition, EPOS II demonstrated that a peak found in a limited subset of their data, comparable to that of the earlier experiments, disappeared when the full data set was analyzed. They also showed that by suitable cuts they could create a peak similar to those found previously. The fact that the effect disappeared when identical cuts were applied to the other half of the data suggested that the peak was, in fact, an artifact created by the cuts.[80] This also suggested that the limited-statistics peaks reported earlier might

[80] Reference 17 cited by EPOS II is a textbook written by Byron Roe, *Probability and Statistics in Experimental Physics*. As the author states, "This book is meant to be a practical introduction into the use of probability and statistics for advanced undergraduate students and for graduate students" (p. v). It includes standard uses of both probability in experimental physics and, in particular, devotes several pages to a discussion of the question of when is a signal significant. The author outlines the method of dividing the data set into two subsets in order to answer that question.

also be artifacts. The EPOS II group had shown that they would have detected a peak had one been present and also that one could artificially create such a peak. As Dirk Schwalm, co-spokesperson for EPOS II, remarked, "I think we all overestimated the statistical relevance of the peaks we saw. It sounds a bit silly in the end, 10 years later, but I think that's what happened" (quoted in Taubes (1997), p. 151).[81] The fact that the effects did not also appear in other interactions, such as Bhabha scattering, in which one would have expected them had the sum-energy peaks been real, provided further evidence against the reality of the peaks.

Despite the fact that it was later shown that the experimental results on both positron lines and on the sum-energy peaks in heavy-ion collisions were incorrect, I believe that this episode is, in fact, an example of good science. A wrong result is not bad science. The original experiments were motivated by the interesting theoretical prediction of spontaneous positron production. In fact, one positive result of these experiments is increased understanding of the production of positrons in heavy-ion collisions. (See, for example, Tsertos et al. 1992). The early experiments found suggestive and intriguing results that were later investigated in detail.[82] As discussed earlier the cuts applied in anayzing the experimental data were reasonable, particularly when one is looking for a small effect and there is little theoretical guidance as to the important parameters. Ultimately it was decided that the results were wrong. The decision that there were no low-mass electron-positron states was arrived at, as I have shown, on the basis of experimental evidence and rational discussion and criticism.

VII. Conclusion

In this paper we have seen several types of selection criteria that have been applied to both data and to analysis procedures. The cuts have ranged from Millikan's legitimate

---

> In another case, in an international collaboration, we had a group of enthusiasts who had made various cuts and produced a very unexpected $\mu\pi$ resonance in neutrino interactions. We had lots of arguments about whether it was publishable. Fortunately, we were about to analyze the second half of our data. We froze the cuts from the first half and asked whether they produced the same peak in the second half. This is a fair procedure and is useful if you have enough data.
>
> Play with one-half of the data and then if an effect exists, check it in the second half. It is still necessary to derate the probability by the number of times the second half got checked, but it is a considerable help. In our particular case, the signal vanished, but with similar but not quite the same cuts a new signal could be found in μk . . .(p.112).

The similarity to the case of the sum-energy peaks is obvious.
This technique of dividing a data set into subsets has also been used to estimate systematic errors in an experiment (Wiss and Gardner 1994). The point is that it is a standard technique and not unique to EPOS II.

[81] This is a very embarrassing admission for an experimenter to make. After all, the groups had spent considerable time, effort, and money in pursuing what turned out to be an artifact.

[82] Reinhardt's survey of the literature on the subject includes 282 references. (It is available at http://www.gsi.de).

exclusion of data obtained when he wasn't sure that his experimental apparatus was working properly[83] to the very complex tuning of analysis cuts that produced an artifact in the case of the suggested low-mass electron-positron states. It is clear that there is no single solution to the problem of whether the experimental result is an artifact created by the cuts. What may work in one case may not work in another. There are, however, some general strategies.

Consider, for example, robustness. This is an important method of demonstrating the validity of an experimental result, and for dealing with the problem of cuts. It was, in fact, used in each of the episodes discussed in this paper. In the experiment to measure the K^+_{e2} branching ratio, for example, the experimenters varied both the range cut and the track-matching criterion over reasonable intervals and showed that the branching ratio found was robust under those variations. In the case of both gravity waves and the 17-keV neutrino, robustness was again important. In the gravity wave episode, Weber's critics used both their own preferred analysis algorithm as well as Weber's non-linear algorithm and showed that they still found no gravity wave signal. This was one of the strong arguments in favor of the critics' results and against the correctness of Weber's result. Similarly, in the case of the 17-keV neutrino, several experimenters used both a wide energy range and a narrow energy range in their analysis and demonstrated that their conclusions did not change. In the decisive experiments that showed that the 17-keV neutrino did not exist the experimenters demonstrated that the choice of analysis procedure was not a problem in their experiments. We also saw that Simpson's apparent failure to use robustness as a criterion led to his interpretation of an artifact of data analysis as a real effect in his reanalysis of Ohi's data. In the case of Millikan's measurement of the charge of the electron, robustness was provided by subsequent measurements of that charge.

Robustness did not, however, provide an unambiguous solution to the problem in the episode of the low-mass electron-positron states. This was because the results obtained, although similar, seemed to be extremely sensitive to the experimental conditions, such as time of flight, bombarding energy, scattering angle, and the equaltity of electron and positron energies (the "wedge cut"). Varying those conditions seemed to make the effects vary or disappear. They lacked robustness. Were the variations a real sensitivity to the

[83] It should be emphasized that this data selection cut by Millikan is not unique. Few experiments work properly the first time that are turned on, and no experimenter would accept data unless they were convinced that the apparatus was working properly. In the case of the K^+_{e2} branching ratio experiment I know, because I was a participant in the experiment, that data was excluded when the apparatus was not working properly. I believe that this was also true for the other episodes. Consider a problem that developed in the experiment that first demonstrated the violation of combined particle-antiparticle and space-reflection symmetry (CP violation) (Christenson et al. 1964). For details see Franklin (1986, pp. 83–87). An interaction with other nearby experiments not only stopped the taking of data, but also led to excluding the data taken while the problem existed. These runs were interrupted by discovery that bending magnet of Frisch at 6 BeV gives $\sim$20/1 ratio of [counter] 3 to [counter] 2. This is intolerable. Now they have reduced beam and we resume running pending solution. (Quote from the laboratory notebook of the Fitch-Cronin experiment)." The setting of a magnet in the adjacent experiment run by an MIT group dramatically changed, and not for the better, the operation of the Princeton experiment. It made the data taken under those conditions unreliable. It wasn't "good"data.

conditions or were they artifacts? There are, after all, many phenomena in science which exhibit such sensitivity. Consider the discovery of the J/Ψ particle. This is a particle with an extremely narrow energy width. One of the experiments that originally found the particle used colliding electron and positron beams. Only when each beam had half the energy of the mass of the J/Ψ was the dramatic cross-section increase that signaled the presence of the particle seen. Changing the beam energies slightly caused the phenomenon to disappear. Experimenters thought that this type of effect might be occurring in the heavy-ion collisions that produced the low-mass electron-positron states. In this episode, later more careful analysis showed that tuning the cuts could produce such results. They were magnifications of statistical fluctuations produced by tuning the selection criteria.

Nevertheless the fact that, in these heavy-ion collisions, similar results were obtained in several experiments using different detectors, different projectile and target nuclei, and at similar, although not identical, energies, gave the results sufficient credibility to encourage further investigation of this subject.

This raises the interesting question of how similar two experiments must be to count as replications and how close experimental results to count as confirmations. As shown in Fig. 35 a large number of possible states were found in the different experiments. The experimenters interpreted the results as evidence for three such states and used the fact that they were obtained in "different" experiments to support their existence.[84] In retrospect, they were wrong. There is no easy solution to this problem. How similar the conditions or effects must be can be decided only on a case by case basis. In this episode the sensitivity to experimental conditions was shown to be an artifact. In the case of the J/Ψ the sensitivity of the result to the experimental conditions led to an important discovery.

Replication, another form of robustness, also plays an important role in guarding against artifacts created by cuts. In the case of Millikan, unlike the other episodes discussed, both his data exclusion and his varying analysis procedures were private and thus unavailable to the scientific community. Here the robustness of the value of *e* obtained in both similar and different experiments argued for the correctness of Millikan's result and acted as a safeguard against his selectivity. This is usually the case in experiments with important theoretical implications. For example, in the discovery of the existence of the intermediate vector boson there were two experiments, UA1 and UA2, that each demonstrated the existence of the particle. In the case of the SLAC E122 experiment that demonstrated the existence of parity violation in electroweak interactions the fact that there was only a single experiment done made the epistemological arguments in support of the correctness of the result crucial. The experiment has not been replicated, but the care with which it was done and anayzed has persuaded the physics community that the result is correct. (For details see Franklin 1990, Ch. 8). The failure, however, to reproduce the low-mass electron-positron effects in Bhabha scattering, a different physical system, but one in which the same effects were expected to be seen, also cast doubt on the results.

The fact that one could show that cuts could create the observed effect also played a major role in these episodes. Thus, Kafka, analyzing his own data and varying his

---

[84] For a discussion of why "different" experiments provide more support for a hypothesis than does the repetition of the "same" experiment see Franklin and Howson (1984).

threshold criterion showed that he could create an apparent gravity wave signal. The same effect was shown by Levine and Garwin using a computer simulation. In the episode of the 17-keV neutrino, Bonvicini demonstrated, also using a Monte Carlo calculation, that analysis cuts combined with limited statistics could produce effects that might mask or mimic the presence of the proposed particle. It should be emphasized, however, that demonstrating that an effect can be produced by applying selection criteria can only cast doubt on an experimental result. It cannot demonstrate that the result is incorrect. In the case of both gravity waves and the low-mass electron-positron states, other arguments were both needed and provided. For the gravity-wave episode, as we have seen, there were other difficulties with Weber's result. In addition, six other experiments failed to replicate his result. In the episode of the low-mass electron-positron states, EPOS II demonstrated, using real data, that cuts could create the observed effects. This, combined with the lack of robustness of the results, led to the reasonable conclusion that the initially observed effects were artifacts produced by tuning the cuts. Conversely, arguing, as was the case for the K^+_{e2} branching ratio experiment, that the applied cuts could not create the observed effect, increased confidence in the result.

Sometimes one can argue that an experimental result is not an artifact by the use of a surrogate signal. If the apparatus can detect such a signal then it argues that both the experimental apparatus, along with the analysis procedure, are working properly. This was the case in the episodes of both gravity waves and the 17-keV neutrino. Weber's critics were able to detect a pulse of acoustic energy injected into the antenna that mimiced the effect expected for gravity waves. The Argonne group was able to detect the kink created by the composite spectrum of ^{35}S and ^{14}C which served as a surrogate for the effect expected for the 17-keV neutrino. Such a procedure tests the proper operation of the experimental apparatus and the analysis procedure, including the cuts.

Should the fact that there is no single algorithm or procedure to guard against results that are artifacts of the selection criteria used to produce them cause us to doubt both experimental results and the science based on those results? I think not. Difficulty is not impossibility. Although, as we have seen, the validty of results may be difficult to establish, it is not impossible. In each of these episodes the question of whether the result was an artifact was answered. Sometimes, as in the case of Weber's gravity wave results, the 17-keV neutrino, and the low-mass electron positron states, the observed effect was an artifact. In the others it wasn't. It would be an error to conclude from the fact that three out of the five cases discussed here had results that were artifacts of the selection criteria that this is typical of experimental results in physics. The episodes were chosen precisely because there were discordant results and because selection criteria were important. The K^+_{e2} branching ratio experiment is the norm, not the exception. Cuts may be ubiquitous, but they are not fatal.

References

Ahmad, I., S.M. Austin, B.B. Back, et al. (1995a). "Positron Production in Heavy Ion Collisions: Current Status of the Problem". *Nuclear Physics A* 583: 247–256

Ahmad, I., S.M. Austin, B.B. Back, et al. (1995b). "Search for Narrow Sum-Energy Lines in Electron-Positron Pair Emission from Heavy-ion Collisions near the Coulomb Barrier". *Physical Review Letters* 75: 2658–2661

Ahmad, I., S.M. Austin, B.B. Back, et al. (1996). “Ahmad et al. Reply”. *Physical Review Letters* 77: 2839

Ahmad, I., S.M. Austin, B.B. Back, et al. (1997a). “Internal Pair Conversion in Heavy Nuclei”. *Physics Review C* 55: R2755–R2759

Ahmad, I., S.M. Austin, B.B. Back, et al. (1997b). “Search for Monoenergetic Positron Emisi-ion from Heavy-Ion Collisions at Coulomb-Barrier Energies”. *Physical Review Letters* 78: 618–621

Albrecht, H., U. Binder, P. Bockmann, et al. (1986). “Search for Exotic Decay Modes of the Upsilon (1s)”. *Physics Letters B* 179: 403–408

Baba, C.V.K., D. Indumathi, A. Roy, et al. (1986). “Search for a Light Neutral Particle in the Decay of the 3.68 MeV State in ^{13}C”. *Physics Letters B* 180: 405–408

Backe, H., W. Bonin, E. Kankeleit, et al. (1983a). “Positrons from Heavy Ion Collisions - Experiments at the UNILAC”. *Quantum Electrodynamics of Strong Fields. New York: Plenum Press.* W. Greiner (Ed.). New York, Plenum Press: 107–132

Backe, H., L. Handschug, F. Hessberger, et al. (1978). “Observation of Positron Creation In Superheavy Ion-Atom Collision Systems”. *Physical Review Letters* 40: 1443–1446

Backe, H., P. Senger, W. Bonin, et al. (1983b). “Estimates of the Nuclear Time Delay in Dissipative U + U and U + Cm Collisions Derived from the Shape of Positron and d-Ray Spectra”. *Physical Review Letters* 50(1838–1841)

Balantekin, A.B., C. Bottcher, M.R. Strayer, et al. (1985). “Production of New Particles in Heavy-ion Collisions”. *Physical Review Letters* 55: 461–464

Bar, R., A. Balanda, J. Baumann, et al. (1995). “Experiments on e^+e^--Line Emission in HI Collisions”. *Nuclear Physics A* 583: 237–246

Bargholtz, C., L. Holmberg, K.E. Johansson, et al. (1987). “Investigation of Anomalous Spectral Structure in Low-Energy Positron Scattering”. *Journal of Physics G* 13: L265–L270

Biedenharn, L.C. (1983). “Summary Talk: Theoretical”. *Quantum Elecrodynamics of Strong Fields*. W. Greiner (Ed.). New York, Plenum Press: 837–851

Bokemeyer, H., K. Bethge, H. Folger, et al. (1983). “Search for Spontaneous Positron Production in Heavy-Ion Collisions”. *Quantum Elecrodynamics of Strong Fields*. W. Greiner (Ed.). New York, Plenum Press: 273–292

Bokemeyer, H., P. Salabura, D. Schwalm, et al. (1989). “Correlated Electron-Positron Emission in Heavy-Ion Collisions”. *Tests Of Fundamental Laws in Physics*. O. Fackler and J. Tran Thanh Van (Ed.). Gif sur Yvette, Editions Frontieres: 77–92

Bonvicini, G. (1993). “Statistical Issues in the 17-keV Neutrino Experiments”. *Zeitschrift für Physik A* 345: 97–117

Borge, M.J.G., A. De Rujula, P.G. Hansen, et al. (1986). “Limits on Neutrino-Mixing from the Internal Bremsstrahlung Spectrum of ^{125}I”. *Physica Scripta* 34: 591–596

Bowcock, T., R.T. Giles, J. Hassard, et al. (1986). “Upper Limits for the Production of Light Short-Lived Particles in Radiative Upsilon Decay”. *Physical Review Letters* 56: 2676–2679

Bowen, D.R., A.K. Mann, W.K. McFarlane, et al. (1967). “Measurement of the K^+_{e2} Branching Ratio”. *Physical Review* 154: 1314–1322

Brown, C.N., W.E. Cooper, D.A. Finley, et al. (1986). “New Limit on Axion Production in 800 GeV Hadronic Showers”. *Physical Review Letters* 57: 2101–2104

Bryman, D.A. and E.T.H. Clifford (1986). “Exotic Muon Decay MU -> e + x”. *Physical Review Letters* 57: 2787–2788

Chodos, A. and L.C.R. Wijewardhana (1986). “Production Mechanisms for a New Neutral Particle below 2 MeV”. *Physical Review Letters* 56(302–305)

Christenson, J.H., J.W. Cronin, V.L. Fitch, et al. (1964). “Evidence for the 2π Decay of the K°_2 Meson”. *Physical Review Letters* 13: 138–140

Clemente, M., E. Berdermann, P. Kienle, et al. (1984). “Narrow Positron Lines from U-U and U-Th Collisions”. *Physics Letters* 137B: 41–46

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

Connell, S.H., R.W. Fearick, R.F.A. Hoernle, et al. (1988). "Search for Low-Energy Resonances in the Electron-Positron Annihilation-in-Flight Photon Spectrum". *Physical Review Letters* 60: 2242–2245

Cowan, T., H. Backe, M. Begemann, et al. (1985). "Anomalous Positron Peaks from Supercritical Collision Systems". *Physical Review Letters* 54: 1761–1764

Cowan, T., H. Backe, K. Bethge, et al. (1986). "Observation of Correlated Narrow-Peak Structures in Positron and Electron Spectra from Superheavy Collision Systems". *Physical Review Letters* 56: 444–447

Cowan, T.E. and J.S. Greenberg (1996). "Comment on the APEX e^+e^- Experiment". *Physical Review Letters* 77: 2839

Cowan, T.E., J.S. Greenberg, H. Backe, et al. (1987). "Narrow Correlated Electron-Positron Peaks from Superheavy Collision Systems". *Physics of Strong Fields*. W. Greiner (Ed.). New York, Plenum Press: 111–194

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

Davier, M., A. Jeanjean and H. Nguyen-Ngoc (1986). "Search for Axion-Like Particles in Electron Bremsstrahlung. *Physics Letters B* 180: 295–298

DeBoer, F.W.N., K. Abrahams, A. Balanda, et al. (1986). "Search for Short-Lived Axions in a Nuclear Isoscalar Transition. *Physics Letters B*180: 4–8.

Ditzel, E., J.M. Hoogduin, H. Backe, et al. (1997). "Absence of E0 Transitions around 1.8 MeV after Collisions of ^{238}U with ^{181}Ta". *Zeitschrift für Physik A* 358: 11–14

Eichler, R., L. Felawka, N. Kraus, et al. (1986). "Limits for Short-Lived Neutral Particles Emitted in μ^+ or π^+ Decay". *Physics Letters B* 175: 101–104

Erb, K.A., I.Y. Lee and W.T. Milner (1986). "Evidence for Peak Structure in e^+ + Th Interactions". *Physics letters B* 181: 52–56

Faestermann, T., F. Heine and P. Kienle (1996). "Are the GSI Positron Electron Lines Due to Internally Converted E0 Transitions?". *Zeitschrift für Physik A* 354: 213–214

Fairbank Jr., W.M. and A. Franklin (1982). "Did Millikan Observe Fractional Charges on Oil Drops?". *Amercian Journal of Physics* 50: 394–397

Ford, K.W. (1968). *Basic Physics*. Lexington: Xerox

Franklin, A. (1981). "Millikan's Published and Unpublished Data on Oil Drops". *Historical Studies in the Physical Sciences* 11: 185–201

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

Franklin, A. (1990). *Experiment, Right or Wrong*. Cambridge: Cambridge University Press

Franklin, A. (1993). *The Rise and Fall of the Fifth Force: Discovery, Pursuit, and Justification in Modern Physics*. New York: American Institute of Physics

Franklin, A. (1994). "How to Avoid the Experimenters' Regress". *Studies in the History and Philosophy of Science* 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. and C. Howson (1984). "Why Do Scientists Prefer to Vary Their Experiments?". *Studies in History and Philosophy of Science* 15: 51–62

Garwin, R.L. (1974). "Detection of Gravity Waves Challenged". *Physics Today* 27(12): 9–11

Ganz, R., R. Bar, A. Balanda, et al. (1996). "Search for e^+e^- Pairs with Narrow Sum-Energy Distributions in Heavy Ion Collisions". *Physics Letters B* 389: 4–12

Greenberg, J.S. (1983). "Summary: Experimental Aspects. *Quantum Elecrodynamics of Strong Fields*. W. Greiner (Ed.). New York, Plenum Press: 853–892

Greenberg, J.S. and W. Greiner (1982). "Search for the Sparking of the Vacuum". *Physics Today* 35(8): 24–32

Greiner, W., Ed. (1983). *Quantum Electrodynamics of Strong Fields*. New York: Plenum Press

Greiner, W., Ed. (1987). *Physics of Strong Fields*. New York: Plenum Press

Greiner, W. and J. Reinhardt (1995). "Quo Vadis Electron-Positron Lines?". *Physica Scripta* T56: 203–219

Griffin, J.J. (1995). "Sharp (e^+e^-) Pairs: Reason for Caution". *Acta Physica Slovaca* 45: 661–672

Griffin, J.J. (1997a). Comment on "A Search for Narrow Sum Energy Lines in Electron-Positron Pair Emission from Heavy Ion Collisions Near the Coulomb Barrier". College Park, MD, University of Maryland

Griffin, J.J. (1997b). APEX: Positive Evidence for Sharp 800 keV Pairs from Heavy Ion Collisions Near the Coulomb Barrier. College Park, MD, University of Maryland

Hallin, A.L., F.P. Calaprice, R.W. Dunford, et al. (1986). "Restrictions on a 1.7 MeV Axion from Nuclear Pair Transitions". *Physical Review Letters* 57: 2105–2108

Henderson, S.D., P. Asoka-Kumar, J.S. Greenberg, et al. (1992). "Search in *s* Channel for Production of 1–2 MeV/c^2 Long-Lived e^+e^- Resonances". *Physical Review Letters* 69: 1733–1736

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

Hime, A. (1992). "Pursuing the 17 keV Neutrino". *Modern Physics Letters A* 7: 1301–1314

Judge, S.M., B. Krusche, K. Schreckenbach, et al. (1990). "Search for Long-Lived Neutral Resonances in Bhabha Scattering around 1.8 MeV/c^2". *Physical Review Letters* 65: 972–975

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

Kienle, P. (1983). "Pair Creation in Heavy Ion-Atom Collisions". *Quantum Elecrodynamics of Strong Fields*. W. Greiner (Ed.). New York, Plenum Press: 293–316

Kienle, P. (1989). "Electron-Positron Pair Creation in Heavy Ion Collisions Investigated with 'Orange' Spectrometers". *Tests Of Fundamental Laws in Physics*. O. Fackler and J. Tran Thanh Van (Ed.). Gif sur Yvette, Editions Frontieres: 63–75

Koenig, W., E. Berdermann, F. Bosch, et al. (1989). "On the Momentum Corelation of (e^+e^-) Pairs Observed in U + U and U + Pb Collisions". *Physics Letters B* 218: 12–16

Koenig, W., F. Bosch, P. Kienle, et al. (1987). "Positron Lines from Subcritical Heavy Ion-Atom Collisions". *Zeitschrift für Physik A* 328: 129–145

Konaka, A., K. Imai, H. Kobayashi, et al. (1986). "Search for Neutral Particles in Electron-Beam-Dump Experiment". *Physical Review Letters* 57: 659–662

Kozhuharov, C., P. Kienle, E. Berdermann, et al. (1979). "Positrons from 1.4 - GeV Uranium Atom Collisions". *Physical Review Letters* 42: 376–379

Leinberger, U., E. Berdermann, F. Heine, et al. (1997). "New Results on e^+e^--line Emission in U + Ta Collisions". *Physics Letters B* 394: 16–22

Lorenz, E., G. Mageras, U. Stiegler, et al. (1988). "Search for Narrow Resonance Productions in Bhabha Scattering at Center-of-Mass Energies Near 1.8 MeV". *Physics Letters B* 214: 10–13

Mageras, G., P. Franzini, P.M. Tuts, et al. (1986). "Search for Light Short-Lived Particles in Radiative Upsilon Decays". *Physical Review Letters* 56: 2672–2675

Maier, K., W. Bauer, J. Briggmann, et al. (1987). "Experimental Limits for Narrow Lines in the Excitation Function of Positron-Electron Scattering around $E^* = 620$ keV and $E^{*=810}$ keV". *Zeitschrift für Physik A* 326: 527–529

Maier, K., E. Widmann, W. Bauer, et al. (1988). "Evidence for a Resonance in Positron-ElectronScattering at 810 keV Centre-of-Mass Energy". *Zeitschrift für Physik A* 330: 173–181

Meyerhof, W.E., J.D. Molitoris, K. Danzmann, et al. (1986). "Search for a Correlated Narrow-Peak Structure in the Two-Photon Spectrum from 6.0 MeV/Nucleon U + Th Collisions". *Physical Review Letters* 57: 2139–2142

Millikan, R.A. (1911). "The Isolation of an Ion, A Precision Measurement of Its Charge, and the Correction of Stokes's Law". *Physical Review* 32: 349–397

Millikan, R.A. (1913). "On the Elementary Electrical Charge and the Avogadro Constant". *Physical Review* 2: 109–143

Millikan, R.A. (1917). *The Electron*. Chicago: The University of Chicago Press

Mills Jr., A.P. and J. Levy (1987). "Search for a Bhabha-scattering Resonance near 1.8 MeV/c^2". *Physical Review D* 36: 707–712

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 (Ed.). 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

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. (1993b). "0.073% (95% CL) Upper Limit on 17 keV Neutrino Admixture". *XXVI International Conference on High Energy Physics*. J. R. Sanford (Ed.). Dallas, American Institute of Physics. 1: 1128–1135

Ohshima, T., H. Sakamoto, T. Sato, et al. (1993a). "No 17 keV Neutrino: Admixture < 0.073% (95% C.L.)". *Physical Review D* 47: 4840–4856

Peckhaus, R., T.W. Elze, T. Happ, et al. (1987). "Search for Peak Structure in e^+ + Th Collisions". *Physical Review C* 36: 83–86

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

Reinhardt, J., B. Muller, W. Greiner, et al. (1987). "Are the GSI Events Caused by Particle Decay?". *Physics of Strong Fields*. W. Greiner (Ed.). New York, Plenum Press: 315–347

Reinhardt, J., A. Scherdin, B. Muller, et al. (1987). "Resonant Bhabha Scattering at MeV Energies". *Zeitschrift für Physik A* 327: 367–381

Rhein, M., B. Blank, E. Bozrk, et al. (1989). "Electron-Positron Coincidence Experiment at the TORI-Spectromenter". *Tests Of Fundamental Laws in Physics*. O. Fackler and J. Tran Thanh Van (Ed.). Gif sur Yvette, Editions Frontieres: 195–201

Roe, B.P. 1992. *Probability and Statistics in Experimental Physics*. New York: Springer-Verlag

Sakai, M. (1989). "Reply to "Comment on 'Electron Peaks in e^+ + Th, U, and Ta Interactions' " ". *Physical Review C* 40: 1841

Sakai, M., Y. Fujita, M. Imamura, et al. (1988). "Electron Peaks in e^+ + Th, U, and Ta Interactions". *Physical Review C* 38: 1971–1974

Salabura, P., H. Backe, K. Bethge, et al. (1990). "Correlated e^+e^- Peaks Observed in Heavy-ion Collisions". *Physics Letters B* 245: 153–160

Savage, M.J., R.D. McKeown, B.W. Filippone, et al. (1986). "Search for a Short-Lived Neutral Particle Produced in Nuclear Decay". *Physical Review Letters* 57: 178–181

Schafer, A., J. Reinhardt, B. Muller, et al. (1985). "Is There Evidence for the Production of a New Particle in Heavy-Ion Collisions". *Journal of Physics G* 11: L69–L74

Schreckenbach, K. (1989). "Sensitive Search for resonances in Bhabha Scattering Around the Invariant Mass of 1.8 Mev/c^2". *Tests Of Fundamental Laws in Physics*. O. Fackler and J. Tran Thanh Van (Ed.). Gif sur Yvette, Editions Frontieres: 165-170

Schweppe, J., A. Gruppe, K. Bethge, et al. (1983). "Observation of a Peak Structure in Positron Spectra from U + Cm Collisions". *Physical Review Letters* 51: 2261–2264

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

Siegel, D. (1979). "Review of *The Scientific Imagination* by Gerald Holton". *American Journal of Physics* 47: 476–477

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 (Ed.). Gif sur Yvette, Editions Frontieres: 565–577

Sona, P., M. Bini, T. Fazzini, et al. (1989). "A Sensitive Search for the Emission of a Neutral Particle in the Decay of the First Excited State in ^{16}O". *Tests Of Fundamental Laws in Physics*. O. Fackler and J. Tran Thanh Van (Ed.). Gif sur Yvette, Editions Frontieres: 217–220

Taubes, G. (1997). "The One That Got Away?". *Science* 275: 148–151

Tsertos, H. (1989). "Comment on "Electron Peaks in e^+ + Th, U, and Ta Interactions"". *Physical Review C* 40: 1839–1840

Tsertos, H., E. Berdermann, F. Bosch, et al. (1985). "On the Scattering-Angle Dependence of the Monochromatic Positron Emission from U + U and U + Th Collisions". *Physics Letters* 162B: 273–276

Tsertos, H., E. Berdermann, F. Bosch, et al. (1992). "Systematic Studies of Positron Production in Heavy-Ion Collisions near the Coulomb Barrier". *Zeitschrift für Physik A* 342: 79–94

Tsertos, H., P. Kienle, S.M. Judge, et al. (1991). "Experimental Exclusion of Neutral Resonances in Bhabha Scatttering at MeV Energies". *Physics Letters B* 266: 259–263

Tsertos, H., C. Kozhuharov, P. Armbruster, et al. (1988a). "Limits for Resonant Bhabha Scattering Around the Invariant Mass of 1.8 MeV/c^2". *Zeitschrift für Physik A* 331: 103–104

Tsertos, H., C. Kozhuharov, P. Armbruster, et al. (1988b). "Sensitive Search for Neutral Resonances in Bhabha Scattering Around 1.8 MeV/c^2". *Physics Letters B* 207: 273–277

Van Klinken, J., F. Bosch, U. Kneissl, et al. (1989). "Search for Resonant Bhabha Scattering". *Tests Of Fundamental Laws in Physics*. O. Fackler and J. Tran Thanh Van (Ed.). Gif sur Yvette, Editions Frontieres: 147–163

Van Klinken, J., W.J. Meiring, F.W.N. deBoer, et al. (1988). "Search for Resonant Bhabha Scattering Around an Invariant Mass of 1.8 MeV". *Physics Letter B* 205: 223-227

Von Wimmersperg, U., S.H. Connell, R.F.A. Hoernle, et al. (1987). "Observation of Bhabha Scattering in the Center-of-Mass Kinetic-Energy Range 342–845 keV. *Physical Review Letters* 59: 266–269

Wang, T.F., I. Ahmad, S.J. Freedman, et al. (1987). "Search for Sharp Lines in e^+-e^- Coincidences from Positrons on Th". *Physical Review C* 36: 2136–2138

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

Widmann, E., W. Bauer, S. Connell, et al. (1991). "Limits for Two-Photon and e^+e^- Decay Widths of Positron-Electron Scattering Resonances for sqrt s = 1.78 to 1.92 MeV"

Wiss, J. and R. Gardner (1994). "Estimating Systematic Errors". Urbana, University of Illinois

Wong, C.Y. and R.L. Becker (1986). "Scalar Magnetic (e^+e^-) Resonance as Possible Source of an Anomalous e^+ Peak in Heavy-Ion Collisions". *Physics Letters B* 182: 251–255

Wu, X.Y., P. Asoka-Kumar, J.S. Greenberg, et al. (1992). "Search for Low-Mass States in e^+e^- Scattering". *Physical Review Letters* 69: 1729–1732

University of Colorado at Boulder
Department of Physics, CB390
Boulder, CO 80309–0390, USA

(Received September 28, 1998)