"High-dilution" experiments a delusion

The now celebrated report by Dr J. Benveniste and colleagues elsewhere is found, by a visiting Nature *team, to be an insubstantial basis for the claims made for them.*

THE remarkable claims made in *Nature* (**333**, 816; 1988) by Dr Jacques Benveniste and his associates are based chiefly on an extensive series of experiments which are statistically ill-controlled, from which no substantial effort has been made to exclude systematic error, including observer bias, and whose interpretation has been clouded by the exclusion of measurements in conflict with the claim that anti-IgE at "high dilution" will degranulate basophils. The phenomenon described is not reproducible in the ordinary meaning of that word.

We conclude that there is no substantial basis for the claim that anti-IgE at high dilution (by factors as great as 10^{120}) retains its biological effectiveness, and that the hypothesis that water can be imprinted with the memory of past solutes is as unnecessary as it is fanciful.

We use the term "high dilution" reluctantly; these solutions contain no molecules of anti-IgE, and so are not solutions in the ordinary sense. "Solute-free solution" would similarly be illogical.

Our conclusion is based on a week-long visit to Dr Benveniste's laboratory, the INSERM unit for immunopharmacology and allergy (otherwise INSERM 200) at Clamart, in the western suburbs of Paris, during the week beginning 4 July. Among other things, we were dismayed to learn that the salaries of two of Dr Benveniste's coauthors of the published article are paid for under a contract between INSERM 200 and the French company Boiron et Cie., a supplier of pharmaceuticals and homoeopathic medicines, as were our hotel bills.

Benveniste's results are being widely interpreted as support for homoeopathic medicine. In the light of our investigation, we believe that such use amounts to misuse.

Our visit and investigation were preconditions for the publication of the original article. We acknowledge that we are an oddly constituted group. One of us (J.R.) is a professional magician (and also a MacArthur Foundation fellow) whose presence was originally thought desirable in case the remarkable results reported had been produced by trickery. Another of us (W.W.S.) has been chiefly concerned, during the past decade, in studies of errors and inconsistencies in the scientific literature and with the subject of misconduct in science. The third (J.M.) is a journalist with a background in theoretical physics. None of us has first-hand experience in the field of work at INSERM 200.

We acknowledge that we might well have found ourselves unable to get to grips with the work of the laboratory. But, on the basis of our experience, we are confident that the design of the experiments reported by INSERM 200 is inadequate as a basis for the claims made last month and that the defects we shall catalogue are a sufficient explanation of the remarkable results then reported.

We believe that experimental data have been uncritically assessed and their imperfections inadequately reported. We believe that the laboratory has fostered and then cherished a delusion about the interpretation of its data.

We are grateful to Dr Jacques Benveniste for his openness in discussing most of the questions we raised with him. He allowed us to borrow and to photocopy the relevant laboratory notebooks, which were invaluable for our investigation. We have every reason to believe that Dr Benveniste was (and, perhaps, still is) convinced of the reality of the phenomena reported in his article. We are also in the debt of several of Dr Benveniste's colleagues, especially to Dr Elisabeth Davenas. On her fell most of the burden of demonstrating the standard dilution experiments and of repeating them in a blinded protocol under our scrutiny. We know that our report will be a disappointment to the laboratory. We are sorry.

What follows is a narrative account of our visit and a summary of our conclusions.

Our investigations concentrated exclusively on the experimental system on which the publication was based. During our week in Paris, we resisted several proffered opportunities to examine other systems in which high dilution is claimed

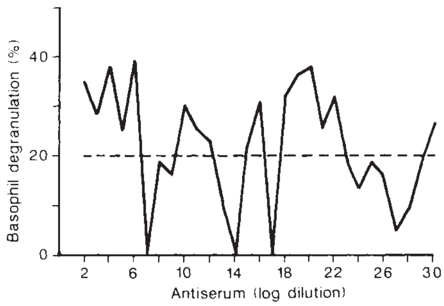

Fig. 1 *A demonstration degranulation, the first of the three open experiments.*

not to diminish the biological effectiveness of a molecule.

The experimental system has evolved from a test for assessing the susceptibility of people to specific allergens. The guiding principle is that blood-borne allergens have the specific effect of interacting with the leukocytes known as basophils, causing them to degranulate — that is, to release the contents of cytoplasmic granules carrying histamine and other active substances provoking the symptoms of asthma and hay-fever.

These allergic reactions are apparently mediated at least in part by IgE molecules attached to the surfaces of basophils (in the blood) or mast cells (in tissues). Normally, degranulation is triggered by the interaction of anchored IgE molecules with an antigen, but the same effect can be brought about by the use of anti-IgE — antibody prepared by injecting human IgE into an animal of another species. (INSERM 200 uses goat anti-IgE at a concentration of 1 mg cm^{-3} sold by the Dutch company Nordic.)

The laboratory notebooks provide ample evidence that this expected degranulation is a maximum between log(dilution) 2 and 4.

Benveniste described the published procedure as a "simple experiment". A buffered solution of anti-IgE is serially diluted by a factor of 10 by transferring measured volumes from one test-tube to another. Pipette tips are discarded after each transfer. Measured volumes of re-suspended white cells derived from human blood are transferred to wells in a polystyrene plate. To each of these is added a measured volume of serially diluted anti-IgE or buffer as a control. The wells are incubated for 30 minutes at 37°C. An acidic solution of toluidine blue, which stains intact but not degranulated basophils red, is added and the numbers of recognizable basophils counted on a haemocytometer slide. Anti-IgG, which does not degranulate basophils, is used as a control.

We were surprised to learn that the experiments do not always "work". There have been periods of several months at a time during which solutions at high dilution have not degranulated basophils. Indeed, the laboratory had just emerged from such a period. (Speculation at the laboratory is that the distilled water may have been contaminated, or otherwise made unsuitable.) It also appears that

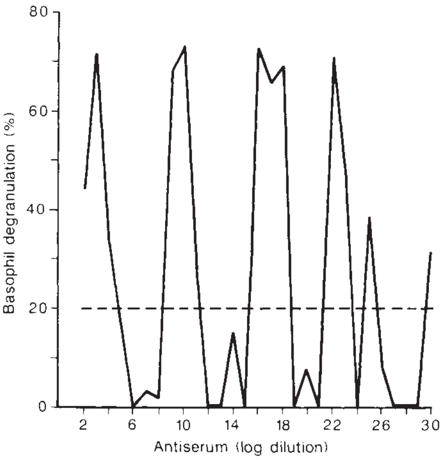

Fig 2 *The fourth demonstration experiment (read "blind") with unexpectedly high peaks (see text).*

bloods that "do not degranulate" are often encountered; we were informed that, in this event, data are recorded but not included in analyses prepared for publication. Even so, the source of blood for the experiments is not controlled, except that an attempt is made not to use blood from people with an allergy.

We witnessed a total of seven runs of this experiment, of which three were routine repetitions of the standard procedure. For the fourth experiment, samples of diluted IgE were transferred by one of us (W.W.S.) to wells in a plastic plate in a random sequence and then read blind by Dr Davenas. All four of these experiments, the last after decoding, gave results described as positive by Benveniste. But three further sequences of counts of stained basophils in three further strictly blind experiments gave negative results (see below).

Figure 1 shows results gathered in the first group of experiments. The ordinate is the decrease (compared with the control) of the numbers of stained basophils at dilutions ranging from 10^{-2} to 10^{-30}. In each case, the left-most peak is that expected from the interaction between anti-IgE and IgE bound to basophils. The number of stained basophils increases to near its control value at log(dilution) of between 5 and 7 (0 per cent degranulation, called "achromasie"); the unexpected phenomenon is that the graphs then reach a series of three or four further peaks with increasing dilution.

These are the successive peaks of activity said in the original article to occur in a periodic fashion, and whose position was said to be reproducible. It is clear from the four graphs that this claim is not obviously supported by this data. The laboratory notebooks confirm that the position of the peaks varies from one experiment to another.

The data in the fourth experiment appear different from those recorded earlier in the laboratory. Indeed, Benveniste volunteered that "we've not seen one like this before". The odd feature of the curve is that the activity of the diluted anti-IgE is, at its peak, identical with that of anti-IgE at log(dilution) 3 — presumably the point at which the natural degranulating effect of anti-IgE is a maximum.

We raised with Dr Benveniste and his colleagues the obviously relevant question of the sampling error. We were astonished to learn, in the discussion of our conclusions at the end of our visit, that neither Dr Benveniste nor his colleagues seemed to be aware of what sampling errors are. We provided a simple explanation, complete with an account of what happens when one pulls a handful of differently coloured balls from a bag, to argue that the sampling error of any counting measurement must be of the order of the square root of the number to be counted. On several occasions, Benveniste called these "theoretical objections".

Ironically, he is himself one of the three authors of a paper published in 1981, in which just this issue had been addressed in a superficially similar situation (Petoit, J.F., Sainte-Laudy, J. & Benveniste, J. *Ann. Biol. clin.* **39**, 355; 1981), and which appears to be the justification of the dotted line drawn at about 20 per cent (corresponding to two standard deviations) on the per cent degranulations of intact basophils after axis.

That brief paper deals exclusively with the effect of sampling errors (not other kinds of errors) on the interpretation of measurements of intact basophils after white-cell suspensions had been allowed to react with allergens via their attached IgE molecules. Even now, at the Clamart laboratory, provision is made for the measurement of two control samples. Among other things, the paper provides a statistical test for telling when the difference between the two control values is statistically significant at the 5 per cent level, in which case people using the procedure as a diagnostic test of allergy are advised to start their experiment all over again.

At INSERM 200, there seems to have grown up a less formal way of dealing with problems of this kind; when the reading of a diluted sample is greater than the control counts, the experimenter often counts the control sample again, on the grounds that the first reading "must have been wrong". This happened when Dr Davenas was counting the first of the first group of experiments.

This procedure exaggerates to some extent the amount of basophil degranulation measured with reagents at high dilution. The practice makes the control values unreliable, and is a significant pointer to the laboratory's disregard of statistical principles.

In these circumstances, it is natural that we should eagerly have accepted Benveniste's invitation to devise a blind experiment. We set out to devise a procedure that would be watertight. We asked that three samples of blood should be run. The serial dilutions would be prepared by Dr Davenas, secretly coded by us before being transferred to wells for incubation and staining by her.

In a small laboratory, procedures like this are inevitably and understandably disruptive. At INSERM 200, the sense of melodrama was further heightened by the general recognition of the importance of the trial, and by the precautions necessary to ensure that the code would not be known to others than ourselves as well as by the need that the one of us with a reputation for sleight of hand (J.R.) could be shown to have been kept away from the test-tubes containing the serial dilutions.

This was done by arranging that Davenas should carry the diluted anti-IgE solutions in stoppered test-tubes to a separate room, where their contents would have been transferred to previously labelled tubes as determined by counters drawn at random from a bag. The coding procedure was monitored by a video camera operated by Randi, who was thereby prevented from touching anything else. (We have a record of the proceedings on an unbroken reel of tape.)

We made two last-minute changes in the planned procedure. First, we included 5 control tubes containing only buffer.

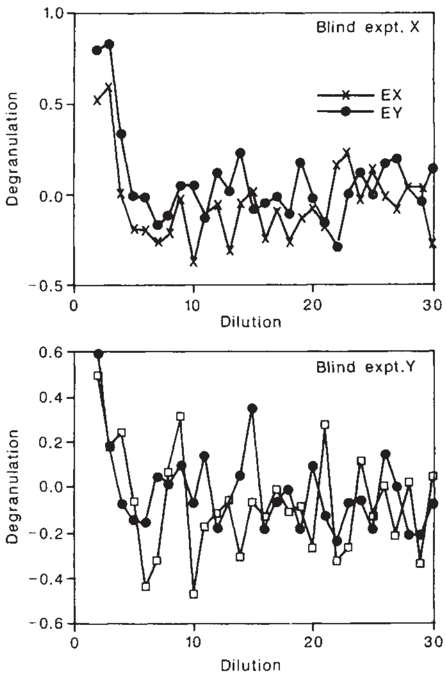

Fig. 3 *Records for the first two blind experiments (5–7 inclusive), showing sampling noise only below the expected decline of degranulation with increasing dilution. Note that the ordinate extends below zero on the degranulation scale (to accommodate sampling errors above as well as below the control values).*

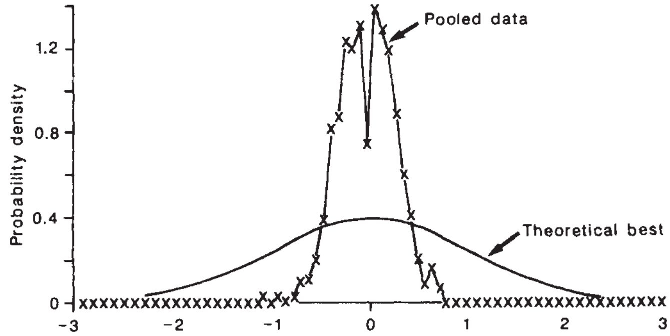

Fig. 4 *Comparison of measured departures of duplicate normalized readings from their means with the gaussian distribution expected.*

Second, having been warned that homoeopathists might regard the data as invalid if solutions were decanted from one set of tubes to another, we removed the numbers written with a felt pen on the original tubes, replacing them with numbered labels which Randi assured us were tamper-proof. The code itself was eventually folded in aluminium foil, enclosed in an envelope specially sealed by Randi and then taped to the laboratory ceiling for the duration of the experiment.

We also arranged a second step of coding just before the slides were counted. One of (W.W.S.) took responsibility for pipetting, after securing the agreement of Davenas and Benveniste that his technique was satisfactory. Both the laboratory procedures and the codes themselves were recorded on video tape. The plates containing the stained cell suspensions were stored in a box (sealed by Randi) in a cold room until read, in random order. The second plate took longer to read, partly because each well was read in duplicate by each observer (Dr Davenas and her colleague, Dr Francis Beauvais), partly because the cells of the second plate were only faintly stained and were thus difficult to read.

Whatever the three runs would provide, we were especially anxious to derive some objective estimate of the intra- and inter-observer measurement errors. We had been told at the outset, by Benveniste, that Dr Davenas was not merely exceptionally devoted to her work but the one in whose hands the experiment most often "works". He said that she usually "counts more cells" than other people. Dr Beauvais, who was also said to be exceptionally skilled, read the slides separately from Dr Davenas, but at the same time. On this occasion, the sampling errors missing from most of the laboratory records did indeed appear.

The duplicate measurements in our strictly blinded experiments were especially important. First, they show that sampling errors do indeed exist, and are not "theoretical objections". Second, they show that the two observers were counting as accurately as could be expected, which gives the lie to the later complaint that the results of the double-blind experiments might be unreliable because the observers had been exhausted by our demands.

Others working in this field recognize the difficulty of counting basophils (roughly 1 in 100 among leukocytes), preferring instead to measure the histamine released on degranulation. This practice is not followed at INSERM 200 because, we were told, of previous failure to record histamine release (as distinct from the disappearance of stained basophils) at high dilution (whence the term "achromasie").

We began to break the codes by lunchtime on our last day, the Friday. When the slides had been matched to the wells from which their samples had derived, but before the appropriate dilutions had been assigned to them, there was a great sense of light-heartedness in the laboratory, no doubt at the prospect that the ordeal would soon be at an end. Benveniste, glancing at the half-decoded data, even offered to predict where the peaks and troughs would fall in the data. His offer was accepted. But his predictions proved to be entirely wrong.

We asked at this stage for criticisms of the conduct of the trials, but were given none. To the question what would be said if the two observers had recorded degranulation peaks, but at different high-dilution values, Benveniste said that would still constitute success.

Opening sealed envelopes is Randi's expertise. He found that the sealed flap of the envelope had detached itself at a surprisingly straight angle when the scotch tape attaching the code to the ceiling was pulled away, but inspection of the aluminium foil allowed him to pronounce himself satisfied that the code had not been read. Then came the decoding — one person singing out numbers to another.

So do the numbers make sense? Six numbers into the record of the first plate to be read, Benveniste said "that patient isn't degranulating, try another". So we did — first the parallel readings by Dr Beauvais, then the remaining two experiments. In the event, the results of all three experiments were similar. The anti-IgE at conventional dilutions caused degranulation, but at "high dilution" there was no effect. Blood from three sources in a row degranulated at ordinary dilutions but not at homoeopathic dilutions. Each of the three experiments was a failure.

Conclusions

We conclude that the claims made by Davenas *et al.* are not to be believed. Our conclusion, not based solely on the circumstance that the only strictly double-blind experiments we had witnessed proved to be failures, may be summarized as follows:

■ **The care with which the experiments reported have been carried out does not match the extraordinary character of the claims made in their interpretation.** What we found, at Clamart, was a laboratory procedure possibly suitable for the application of a well-tested bioassay, but unsuitable as a basis for claiming that anti-IgE retains its biological activity even at a log(dilution) of 120. In circumstances in which the avoidance of contamination would seem crucial, no thought seemed to have been given to the possibility of contamination by misplaced test-tube stoppers, the contamination of unintended wells during the pipetting process and general laboratory contamination (the experiments we witnessed were carried out at an open bench). We have no idea what would be the effect on basophil degranulation of the organic solvents and adhesives backing the scotch tape used to seal the polystyrene wells overnight, but neither does the laboratory.

The design of the experiments hardly matches the nature of their interpretation.

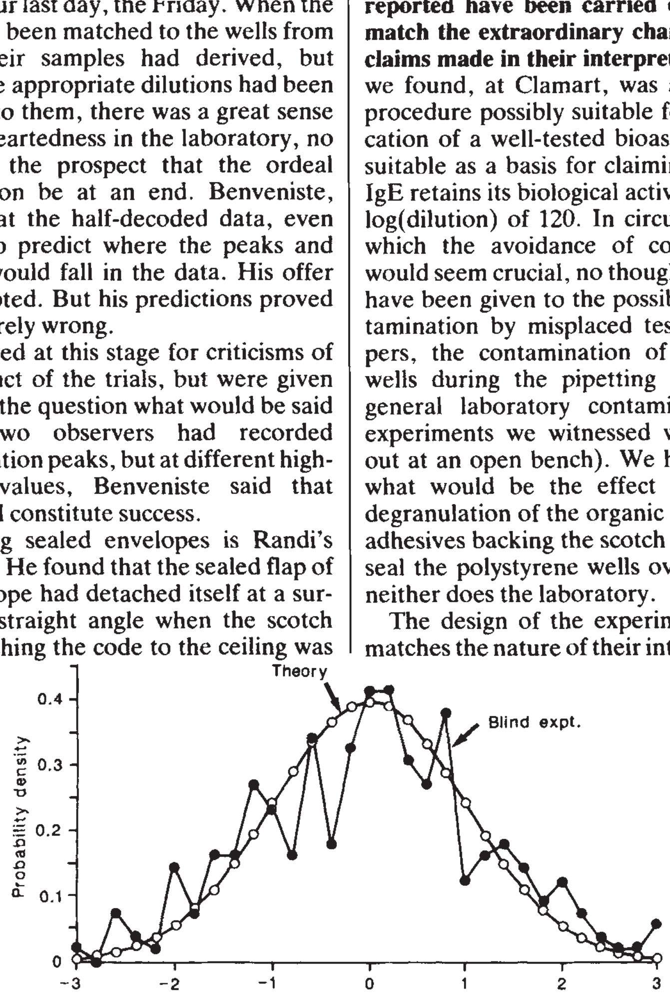

Fig. 5 *Same as Fig. 4 except that data derive from duplicated readings within the blind experiments only.*

For example, one would have thought that counting wells at least in duplicate would have been an elementary precaution against gross errors. The second of our strictly blinded experiments seems to be one of the few in which something of this kind had been attempted.

The laboratory seems to have been curiously uncritical of the reasons why its experiments do not, on many occasions, "work". For example, we were told that the best results were obtained when cells were left in the cold-room overnight before counting, but there has been no investigation of that phenomenon, or of the reports that taking a second sample from a single well gives odd results (an effect not apparent in our double-blind experiments).

■ **The phenomena described are not reproducible, but there has been no serious investigation of the reasons.** We have referred to the fact that some blood yields negative results, and that there are periods of time when no experiments work. But the laboratory notebooks show great variability in the positions at which peaks occur.

■ **The data lack errors of the magnitude that would be expected, and which are unavoidable.** This is best illustrated by Fig. 3, whose two graphs have been constructed from data recorded by Dr Davenas from samples supposedly identical with each other, usually measurements of control samples but also including some duplicate runs. The recorded values have been normalized by subtracting the mean and dividing by the square root of the mean (the expected sampling error). If the only source of error were sampling error, the standard deviation of the plotted curve should be unity (1). Other sources of error, for example, experimental variability, could only increase the standard deviation. But Fig. 3 shows that repeat observations agree more closely than would be expected from the underlying distribution. This is a well-known effect that sometimes affects duplicate readings by the same individual, but the magnitide of the effect in this case calls into question the validity of the readings. This artefact is nevertheless not apparent in the blinded duplicated readings.

■ **No serious attempt has been made to eliminate systematic errors, including observer bias.** It is true that the laboratory notebooks record experiments in which anti-IgG has been used as a control; we were surprised to find that the IgG control run reported by Davenas *et al.* (their Fig. 1*b*) was carried out at a different time from the run with IgE published in the same figure.

Most of the data recorded in the laboratory notebooks derive from experiments in which the same person has been responsible for the sequential dilution, plating out and counting. Given the

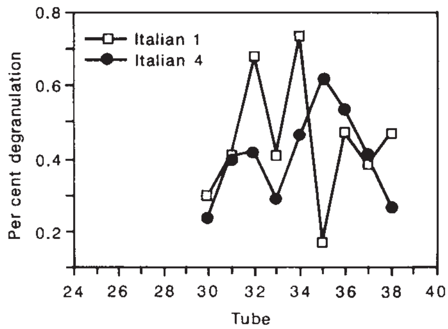

Fig. 6 *Two duplicate Italian runs showing high degranulation, but discordantly.*

shared belief at Clamart in the reality of the phenomenon reported last month, and its potential importance, it is mystifying that duplicate and blind counting is not routine.

■ **The climate of the laboratory is inimical to an objective evaluation of the exceptional data.** So much is readily apparent from the way in which experiments are described as successes and failures, by the use of the word "working" to describe experiments yielding a positive result, and by the several speculations we were offered, without experimental evidence in their support, to explain the several failures the laboratory has experienced. The folklore of high-dilution work pervades the laboratory, as epitomized by the suggestion that decanting diluted solution from one tube to another might spoil the effect and the report that the repeated serial dilution by factors of three and seven (rather than ten) always yields negative results.

Collaborations

We have not been able to pay as much attention as we would have wished to the data collected at other laboratories and cited in Davenas *et al.*, but we have examined documentary evidence available.

Supporting data were said to have come from Rehovot (Israel), Milan and Toronto. Dr Benveniste told us we could not see the Toronto data, described as preliminary, without the consent of the authors, who could not be telephoned.

The data gathered in Israel and Milan are, apparently, significant. Figure 6 is typical of the data from Milan. Though there are no duplicate measurements and therefore no direct evidence of sampling error, there is also some evidence of degranulation at high dilution. Without knowing more about the circumstances, we are unable to comment.

The Israeli data are more extensive. The first trials were in March 1987, during a visit to Rehovot by Dr Davenas. The most remarkable of several successful trials was her correct identification of seven high-dilution tubes out of ten presented to her blind. Even so, the report (to Benveniste) of the trials was cautious. Later, analysis of the tubes which had tested positive in this trial revealed not merely immunoglobins but other protein contaminants apparently identical with materials in the original IgE vial. One of the participants (Professor Meir Shinitsky of the Weizmann Institute) then withdrew as a putative co-author.

Since then, there have been two developments in Israel — a series of experiments carried out independently of Benveniste's laboratory and a further blinded experiment. Data from the latter are unfortunately not available. Maitre Simart, a legal official at Clamart who held the codes, is said not to have had time to decode them.

These measurements are nevertheless, to judge from the documents we have seen, stronger evidence than any we found at Dr Benveniste's laboratory to support his claims. But we do not have the information to evaluate them.

Postscript

We presented the substance of these conclusions to Dr Benveniste and his colleagues immediately after the strictly blinded experiments were decoded. The discussion that followed was inevitably tense. Benveniste acknowledged that his experimental design may not have been "perfect", but insisted (not for the first time) that the quality of his data was no worse than that of many papers published in *Nature* and other such journals.

One of us (J.M.) said it would be best if Benveniste would withdraw the published article, or at least write to *Nature* to qualify his findings and their interpretation, in which case we would not publish this report. It was mutually agreed that nothing would be said publicly until 28 July. But Benveniste said that the laboratory would work through the weekend "and all next week" to prove the reality of the pheonomenon.

Our greatest surprise (and disappointment) is that INSERM 200 seems not to have appreciated that its sensational claims could be sustained only by data of exceptional quality. Randi put the point best, during our Friday discussion, by saying: "Look, if I told you that I keep a goat in the backyard of my house in Florida, and if you happened to have a man nearby, you might ask him to look over my garden fence, when he'd say 'That man keeps a goat'. But what would you do if I said, 'I keep a unicorn in my backyard'?" We have no way of knowing whether the point was taken.

Eventually, there was no more to say. We shook hands all round, sped past the common-room filled with champagne bottles destined now not to be opened and into the lens of a news agency photographer summoned for the happier event.

John Maddox
James Randi
Walter W. Stewart