Waves caused by extreme dilution

From this issue of Nature, *the past several weeks' correspondence on the Benveniste affair will be closed. The editor of* Nature *here discusses some of the issues that have arisen.*

IN FIFTEEN out of the past twenty-two years as editor of *Nature*, I have known nothing like the controversy touched off by the publication last June of an article by Dr Jacques Benveniste and a group of his associates (Davenas, E. *et al. Nature* **333**, 816; 1988) claiming that solutions of anti-IgE retain their biological effectiveness even when indefinitely diluted, and by the publication, a month later, of the report of an on-site investigation by three people, myself among them, who have no expertise in the field of allergic immunology (*Nature* **334**, 287; 1988). I have learned a great deal from the controversy, as have many of my colleagues, although I do not pretend fully to understand why such great passions have been aroused.

What follows is a discussion of some of the issues that have arisen. Because it is intended that *Nature*'s correspondence columns should now revert to their normal uses, only matters raised in the original papers, the subsequent correspondence and certain newspaper comments have been dealt with. New material has been avoided. It is hoped that readers will nevertheless find this account an explanation of some issues not previously fully dealt with. It may be found also to point to some questions about the functioning of the scientific literature and even the practice of science for which there are no simple answers.

The essence of Benveniste's claims is easily summarized. The immunoglobin IgE is one of the mediators of allergic reactions in mammals, at least in part through its attachment to the surfaces of mast cells and of the subclass of lymphocytes known as basophils. Allergens, for example pollen grains, in the blood of persons carrying IgE of appropriate specificity will react in such a way as to trigger first the expulsion of (exocytosis), and then the release of histamine and other humoral factors from, the vesicles containing them in both mast cells and basophils. The same effect can be brought about by antibodies against IgE, on at least one view because they induce the formation of a network of crosslinked IgE molecules on the cell surface.

Anti-IgE antibody is prepared by injecting human IgE into some other mammalian species; Benveniste's material, manufactured by the Dutch company Nordic Immunology, is prepared in goats. The antibody can be used as a reagent in the assessment of allergic reactions in human patients, for example, in whom the gross amount of circulating IgE may be an indicator of a patient's allergic status.

Experiments

The standard measurements at Benveniste's laboratory, INSERM 200 at Clamart in Paris, as observed in July and recorded in earlier laboratory protocols, involves the interaction of anti-IgE solutions with samples of white blood cells taken from the blood of human volunteers. Intact basophils are known to be stained red by their interaction with an acidified solution of toluidine blue. The

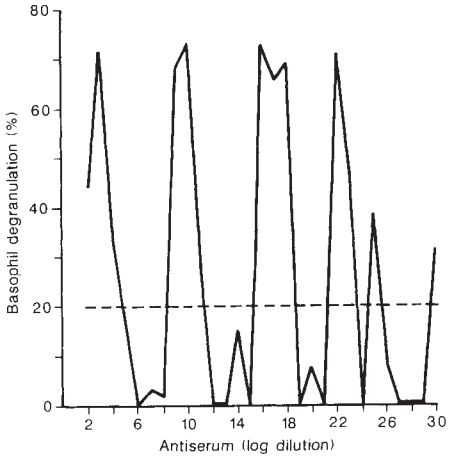

Fig. 1 The fourth of seven demonstration experiments (read 'blind') with unexpectedly high peaks.

standard experiment consists first of allowing suspensions of white blood cells in buffer (a version of Tyrode's solution) to interact with serial dilutions of IgE in the same material and then of counting the surviving intact basophils after staining with toluidine blue.

The general conclusions will be well-remembered. In experiments in which anti-IgE solutions undergo a sequence of ten-fold dilutions (simply accomplished by measuring a fixed amount of inert solvent into each of a sequence of test-tubes and then transferring one ninth of the contents of one tube into the next in line), some of the diluted solutions are found to retain the effectiveness of undiluted IgE in degranulating basophils. The experiments we witnessed in Paris stopped short at 35 successive dilutions, but the original paper referred to experiments in which 60 successive ten-fold dilutions had been carried out, and in which dilutions of 10^{120} had been reached by successive 100-fold dilutions.

The significance of these numbers may not be fully appreciated (although it was clearly explained in the original paper). Avogadro's number is 6.02×10^{23} and the concentration of the commercially supplied anti-IgE is 1 mg cm^{-3}, which corresponds to a molar concentration of 2.2×10^{-6}M. One cm^3 of the commercial preparation will therefore contain some 1.5×10^{15} molecules of anti-IgE.

Provided that the number of antibody molecules transferred from one tube to the next is proportional only to their concentration in the first tube and the volume of the liquid transferred, after 15 ten-fold dilutions, 1 cm^3 of solution will on the average contain 1.5 molecules of anti-IgE — the actual number will be a matter of chance governed, no doubt, by a Poisson distribution. Further dilution beyond this point will yield solutions in which the average number of antibody molecules in 1 cm^3 will be less than one.

It is not for nothing that one correspondent has raised the question of how these extreme dilutions square with the supposedly finite size of the Universe. If the cosmological view that there are 10^{77} baryons (protons and neutrons) in the Universe is correct, the whole bulk of the Universe even if if converted into water would be insufficient to dilute the contents of a 1 cm^3 vial of anti-IgE to the highest dilutions of 10^{120} reported by Benveniste.

The second striking feature of the original paper is the assertion of a rhythmic appearence and disappearence of the degranulation effect at successive dilutions. In the original paper, it was reported that "the repetitive waves of anti-IgE-induced degranulation were reproducible, but the peaks of degranulation could shift by one or two dilutions with every fresh sequential dilution of anti-IgE and depended on the blood sample". The phenomenon is said to depend crucially on the agitation of the freshly diluted solutions at each stage.

No firm explanation of the phenomenon was put forward, but it was suggested that "water could act as a 'template' for the molecule . . .", in effect a kind of memory for molecules long-since diluted away. No model to account for the rhythmic fluctuations of degranulation has yet been put forward. As a reminder of the data, Fig. 1 shows a typical plot of basophil degranulation against anti-IgE dilu-

tion — in reality, this represents the data the fourth of seven of the sets of measurements gathered during July.

While others working in the field say that it would have been better to have worked with mast cells than with basophils, or that basophil degranulation can be more reliably estimated by measuring histamine release quantitatively than by counting stained cells, it should be recalled that the INSERM 200 group acknowledged in July that it had been found that the disappearence of stained cells after reaction with anti-IgE at extreme dilution does not correlate well with histamine release, and cannot therefore be used as a measure of phenomenon.

Nothing in this account of the standard experiments would give offence at INSERM 200.

General criticisms

At the outset, the torrent of correspondence generated by these publications consisted mostly of the responses to the invitation to readers to suggest how the data reported might have arisen normally, perhaps as artefacts, but from the start there were also some general criticisms of *Nature*'s handling of the affair. Some of the questions raised are summarized below, together with my responses. Because many of the questions raised by correspondents and others are contingent on each other (as in "If you had to publish, why not investigate first?"), this dissection of the problem may over-simplify the issues that most trouble readers. It will become apparent that the responses hang to some degree on a conception of *Nature*'s function of which even regular readers may not be fully aware.

Why publish at all? or, in a stronger version, *Why publish with reservations so explicit as to suggest the conclusions should not be believed?* All journals know that, while most articles submitted for publication can be dealt with quickly, there are some on which referees cannot offer helpful advice. Over a period of nearly two years, there were in this case four referees, of whom three provided detailed reports on various occasions. The general tenor of all comments was that, if the data were correct, INSERM 200 must indeed have discovered a remarkable phenomenon, but that the conclusions were so remarkable that the experiments must in some way be flawed.

Throughout this process, Dr Jacques Benveniste (with whom all correspondence was conducted) appeared readily to follow all suggestions offered, in particular arranging that measurements should be independently repeated; he chose laboratories in Israel and Milan. Copies of pages from the Clamart laboratory notebooks as well as data from Israel and Milan were sent for scrutiny by referees.

In retrospect, it is nevertheless clear that the the paper as published contained internal inconsistencies, notably that quoted errors were much smaller than the expected sampling errors. More seriously, perhaps, we were unimaginative with our suggestion of how more stringent controls might have been devised. I should also have been more cautious when, having rejected the paper for what my colleagues hoped would be the last time, Dr Benveniste telephoned indignantly to protest that *Nature* was proposing to suppress news of one of the greatest discoveries of the twentieth century. I forget whether he compared his dilemma to that of Galileo on that occasion or in a conversation during the later visit to Paris.

To my complaint along the lines of "But you don't even consider how these extraordinary results might be conventionally explained" Dr Benveniste answered with a version of the Russian verbal shoulder-shrug "No problem!"; the section of the published paper about the long-term memory of water followed only a few days later.

Journals are these days painfully aware of the accusation that their collective influence is to inhibit innovation. The accusation is untrue, given the competition among journals for soundly based innovatory publications. *Nature*, moreover, consciously publishes a regular sprinkling of heterodoxy, mostly as "Commentary" or "Scientific Correspondence". This is an important function even when there is little chance that the heterodox will become the orthodox; people may find it instructive to know what is happening on the fringes of their interests.

But the accusation of suppression by the head of a government-supported laboratory cannot be dismissed lightly. And while, in an ideal world, exasperation should play no part in editorial decisions, it must surely on occasion be proper that a person convinced that heterodox conclusions are correct should be given the opportunity for which he asks to see what happens when they are tested publicly.

Those were the circumstances in which it was agreed between Dr Benveniste and I that the paper would be published, but that publication would be followed by an on-site investigation by a nominated group. From the outset, Dr Benveniste knew the composition of the group and that it would be sceptical of his conclusions. My concern was that the publication of his paper (certain to excite the interest of the homoeopathic community) should be followed as quickly as possible by the appearence of a report of the investigation. The actual timing of the visit to Paris was decided by Dr Benveniste and Mr James Randi; for *Nature*'s purposes, it was inconveniently soon.

Why not investigate first and publish only afterwards? Journals do not normally undertake investigations of contributors' laboratories, and for good reason: they have neither the resources nor the skill, and they cannot command them from elsewhere. The decision to take up Dr Benveniste's long-standing invitation to "send" a group of people to Paris was mine, prompted largely by the exceptional circumstances — Benveniste's evident convinction of the correctness of his conclusions, his insistence that his data should see the light of day and our suspicion that one explanation might be that the data had been generated by a hoaxer in his laboratory. It seemed best that the material should be published before and not after the investigation. With hindsight, it would probably have been sufficient that there should have been an informal preliminary visit to the laboratory, but that was not self-evident at the outset.

From correspondence and comment in the general press, the procedure that we actually followed seems to have raised objections on two counts. First, it is said, it has turned a serious issue into a "circus". Second, the procedure was essentially a trick by means of which Dr Benveniste and his colleagues exposed themselves to gratuitous criticism (or were "stitched up", as the *New Scientist* inelegantly put that point in August).

Bargain

In reality, the arrangement was an explicit bargain. Investigation was a pre-condition of publication (although Dr Benveniste appears not to have informed his colleagues of that circumstance). But even though avowedly sceptical of the conclusions, I had no reason for confidence that a necessarily brief visit to Paris would have yielded support for that scepticism. That (in our opinion) it did has inevitably engendered something of an air of melodrama (the extent of which is nevertheless surprising). But those who have advanced this criticism overlook the danger nearer the front of my mind in June — the chance that a brief visit to INSERM 200 would provide substance only for a report recording that experiments were carried out, and with the results already anticipated in *Nature*.

To have inverted the bargain would have given Dr Benveniste full freedom of decision; given a favourable or even non-commital report, he would have claimed publication, but otherwise could have withdrawn. Locking both of us into an agreement to publish simultaneously would have led either to an implicit endorsement of an unbelievable result or to a nonsense — the simultaneous publication of an extraordinary claim and its purported refutation. (The circumstances are different from those of the earlier simultaneous publication of such a claim, that a protein called scotophobin can serve as a vehicle for the transfer of learned be-

haviour from one rat to another, and its refutation, in that the refutation sprang from the internal inconsistencies of the documents submitted for publication.) Even as things were arranged, the investigation might have found no obvious fault with the conduct of the experiments — an outcome of which Dr Benveniste appeared confident until almost the end.

Nature has no ambition to lead a pack of vigilantes seeking to rid the scientific literature of error, and worse; it is rather concerned that over-zealousness may yet make honest mistakes culpable. But there is also a distinction to be drawn between erroneous science, however motivated, and science that is conducted carelessly, allowing sharp inferences to be drawn from insubstantial data.

I am puzzled that Dr Benveniste is as indifferent as appears to be the case, both in several conversations in Paris and in his two comments on our report, of the complaint that he and his colleagues were unaware of the importance of sampling errors. At our final conversation on 8 July, it was clear that the relevance of the point was simply not understood, and discounted as a "theoretical objection".

To the extent that allowance for sampling errors would have made some apparently statistically significant measurements statistically insignificant, it may seem proper simply to set those disputed data aside, relying on the remainder to sustain the startling hypothesis. This is what Dr Benveniste does in referring to "confirmation of our data by two positive experiments", where the Fig. 2 referred to is reproduced here as Fig. 1.

But what if indifference to an ubiquitous source of error should lead to the uncritical acceptance of data which appear to be more consistent among themselves than the simple arithmetic of sampling error would allow? Fig. 4 in our report (reproduced here as Fig. 2) is compiled from all multiple measurements of the same samples recorded in the notebooks. Its striking feature is that the distribution of the discrepancies of measurement is, for whatever reason, narrower than the gaussian distribution expected for sampling errors. The bite chewed out of the measured curve at its mid-point even suggests that identical measurements of samples of the same suspension of blood cells are less frequent than sampling theory would allow (and dictate). These distributions suggest that observer bias seriously affects the whole series of data: in the circumstances, it is difficult to tell what confidence there can be in even the most apparently clear-cut graph, even the data represented here in Fig. 1.

Figure 1, on the other hand, has been taken by Dr Benveniste in his accompanying comment and elsewhere as a proof that even a hostile investigation threw up data consistent with his hypothesis and he denies (contrary to the recollection of all three of us) that he remarked "We've never seen one like that before". But inspection shows that there is a significant difference: in Fig. 1, the data points are either very low (no degranulation) or near the maximum (70 per cent or more degranulation). In other published curves and in the laboratory notebooks, the much more common pattern is that there are intermediate data-points, lying improbably often on straight lines between a peak and a trough.

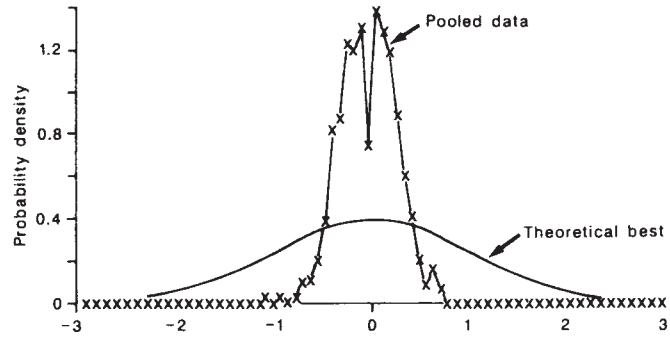

Fig. 2 Comparison of measured departures of duplicate normalized readings from their means with the gaussian distribution expected (Fig. 4 of the report by J. Maddox *et al.*)

In the scientific literature, outright fraud is rare, honest error is probably much more common but is commonly corrected either by its authors (or, more eagerly, by their competitors) or, more often still, is ignored, to the general confusion. But the fate of careless science is usually more seemly — it earns its authors a modest degree of acclaim, but is in the end usually forgotten. Dr Benveniste's data have been used to sustain an interpretation that cannot easily be overlooked, but they are also in our judgement carelessly gathered and carelessly interpreted. In my opinion, if these complaints are true, it is a matter of some importance for the scientific community that the extent of these lapses from good practice should be more widely known.

Conviction

There is every reason to believe that Dr Benveniste is convinced of the reality of the phenomenon he describes, and that he is convinced that the data have been gathered properly. But the essence of the report of our investigation is that the Clamart team did not appreciate that sampling errors are unavoidable in their measurements, had not investigated the causes of qualitative variability in their experiments and worked in an intellectual climate conducive to observer bias. Dr Benveniste appears not to have dealt with these charges either in his two published statements or in statements quoted in the general press.

Why send a team of amateurs? Although the composition of the visiting group was known to INSERM 200 at the outset, a fuller explanation is called for, especially because the composition of the group is another of the reasons why the visit to Paris has been called a "circus". That the composition of the group was unorthodox was acknowledged in July. Even so, its members' credentials are not inconsiderable.

The short answer to the question is that if a group of mere amateurs can so quickly discover procedural errors of such importance, that is sufficient justification.

More fully, Walter W.Stewart was one of the most assiduous of the referees of the Benveniste manuscript in its various forms. Over the years, he has demonstrated a flair for spotting inconsistencies in intricate arguments. Stewart has demonstrated, in his handful of published works, great skill as a laboratory scientist. His friends (and his employers, the National Institutes of Health, no doubt) would wish that the zeal he has shown in the pursuit of error in the literature had been devoted to some substantial scientific project. Stewart's presence in Paris has been understood by some in the United States to be an endorsement by *Nature* of his role in the recent pursuit of Dr David Baltimore for not having disowned an alleged error in a published paper (see *Nature* **333**, 795; 1988), which is mistaken.

James Randi, whose presence seems most to have offended Dr Benveniste, is more than a mere stage performer. His role in demonstrating that Uri Geller's illusions could be accomplished by conjuring is now an acknowledged public service, since when he has devoted a large part of his considerable energy to exposing other exploitations of public credulity. That he is a Macarthur Fellow is well-known, that his first job was as a laboratory assistant at the Banting and Best laboratory in Toronto is less so. His presence in Paris, originally suggested as a means of telling whether Dr Benveniste was being hoaxed, proved much more valuable than that. Over the years, Randi has developed a talent for planning in advance how to carry out procedures and to prove that they have been followed faithfully. Even though he announced on 5 July that his task was over, he applied his great intelligence to every detail of the procedures carried out at Clamart. It was he, for example, who warned us that if

Stewart were to carry out the pipetting of blood samples (one of the obvious ways in which readings might be corrupted by accident or design), it was essential that one of Dr Benveniste's colleagues should keep him closely under watch. Unfortunately, Dr Benveniste appeared not to appreciate that Randi is more than a mere conjuror.

Negotiations

Incredible though it may seem, my own intention has been to spend only a day or two in Paris at the beginning of the investigation, helping to negotiate the programme of work for the days ahead and reaching an understanding with everybody concerned about such matters as communicating with the press (an undertaking that Dr Benveniste, unfortunately, was unable to keep). But it quickly became apparent that the process of negotiation would have to be continuous.

It goes without saying that the group's report purports to provide only a sufficient reason why the data cannot be held to sustain the conclusions built on them, and not an explanation of those that appear to stand out above the noise. It may well be that there is also some kind of artefact buried in the data, but if the significance of the data is consistently overestimated by the neglect of sampling errors, a search for artefacts will require the recompilation of a comparable dataset under more rigorous conditions.

But in one important respect, our investigation and report were deficient: we were unconstructive in our comments. On reflection, it is plain that there are several important sources of variability in the measurements that require fuller investigation. Thus there seemed to be a curious lack of intellectual curiosity at Clamart about the reasons why some bloods do not degranulate on some days, why some violent agitation — "vortexing" — is necessary after each dilution, why the results of experiments are better after blood cells have stood in the cold room overnight, and so on. All these would seem crucial both to understanding and underpinning the physical conclusion that extreme dilution has no effect, as would be the further investigation of whether the distribution of anti-IgE is homogenous in the dilution procedure followed in the experiments (one such experiment, with radio-labelled antibody, is referred to in Davenas *et al.*).

What of the other laboratories? Our neglect of what Dr Benveniste considers to be confirmatory evidence appears to have caused him the most distress, and for reasons which are readily understood. Three (not five) collaborating groups were referred to in the original paper.

The data available from the Israeli work is the most explicit but also somewhat confusing. We know of three separate phases of investigation — an attempt to repeat the Clamart experiments (with negative results), a further trial in the presence of Elisabeth Davenas (which yielded positive results but also, unfortunately, accusations of deception by some members of the Israeli group) and a further trial organised remotely from Paris under the supervision of the Clamart bailiff, M. Simart.

The data from the second trial are undoubtedly significant; we said so. There is a profound misunderstanding about the third series of measurements, whose incompleteness came to light when we failed to find the decoded data in the notebooks we had borrowed. Our recollection is that Dr Davenas said at our meeting on 8 July that M. Simart had been too busy to decode them, and that Dr Benveniste said something to the effect that "I'll get them from him on Monday". But now, members of the Paris and the Israeli groups have said that the data were already decoded, in which case we have not seen them (or have mistaken them for other data).

Difficulties

In the hope of circumventing these difficulties, I wrote to all Dr Benveniste's coauthors some weeks ago, asking whether they wished to comment on the past few months' events, for publication or otherwise. One member of the Israeli group has replied, presumably on behalf of all, but unfortunately has stipulated that the contents of his letter must be confidential. The Milan group says it considers *Nature*'s handling of its contributions "unfair" and, in passing, that it had not been aware of the referees' criticisms until the publication of Davenas *et al.* in June. The Toronto group (whose work was cited as confirmatory) replied with a series of legitimate questions about our report, to which there has not been time to provide replies; but the group appears to be embarking on an objective study of the phenomenon in what one member described as a "Popperian spirit".

Nature will be glad to publish as Scientific Correspondence the general conclusions of any or all of these groups when they are ready.

Has it been worthwhile? The short answer is "No". Too many people have been caused too much distress, particularly at INSERM 200, and too much time has been consumed by the circus when there have been better things to do. But again the longer answer is that there are several ways in which the affair has been instructive, certainly for those involved and, it must be hoped, for many readers.

My own guess is that Dr Benveniste's colleagues will now be counting basophils in replicate, following the standard procedure for controlling sampling errors, and will be eliminating unavoidable observer bias by making blind measurements a routine. I expect that the results will not differ substantially from those obtained in the three blind experiments (each with two observers) at Clamart on 9 and 10 July; it will be extremely interesting if it should be otherwise, but no doubt Dr Benveniste would prefer to publish that intelligence in some other journal.

Fallibility

Nature has learned much about the fallibility of the refereeing process, which is not to suggest that the referees involved in this extreme dilution case have been anything but excellent. To put the matter uncontentiously, there are circumstances in which it is difficult to tell from the inspection of a manuscript, and of proferred supplementary material, whether some kinds of claims are valid. This does not imply that the refereeing process is uniformly unreliable, or even consistently full of holes, but that there must be circumstances in which the decision whether or not to publish is subjective. There is no surprise in that for those with editorial responsibility, but — even if only marginally — it gives the lie to the myth that what appears in the journals is the next best thing to the truth.

One of the uncomfortable findings of the high-dilution matter is nevertheless that *Nature* appears to be shackled in the public mind to the totem-pole of absolute veracity. That seems partly to explain much of the reaction of the general press to the events of the past few months. Its interest has been intense but superficial. On one occasion I found myself being questioned closely by a reporter using an *aide-memoire* supplied by Dr Benveniste. It is sad that the notion that truth is black or white takes so long to disappear.

Critics from among the research community have had different motives. Publication has become one of the chief means by which standards of quality are set and maintained in the research profession. *Nature*, with no formal connection with academies and professional societies, has nevertheless by tradition acquired a licence to be a part of the process, to the general enlightment of its readers. There is a sense in which part of the price of the independence is to risk innovation, which explains why *Nature* sometimes appears to behave brashly. But what has plainly seemed to some to have been a mockery of the standards-setting process, however instructive in other ways, is bound to have engendered unusual passion. Thus it is that the events of the past few months are unlikely to be repeated regularly.

So what is the truth about INSERM 200's claims on behalf of high-dilution anti-IgE? One correspondent chided us with having impeded the discovery of the true explanation. My own conviction is that it remains to be shown that there is a phenomenon to be explained.

John Maddox