Chapter 11. Law of small numbers… big consequences

"Statistics is indeed an eminently cheerful science, which requires no mental overwork".

Alphonse Allais. *Ne nous frappons pas* (1901).

Despite the optimistic quotation of the French humorist Alphonse Allais, the two chapters that follow are the most technical ones of this book. Their reading requires a minimum knowledge in statistics. I nevertheless invite the readers who are not fond of mathematics (they are not however of a high level; they are only mathematics for biologists…) to read them, even if he/she jumps the too difficult passages. These chapters are indeed important because they undermine the central argument of the investigation report which is, let us remember, that the variability of the repeated counts of basophils were too low according to chance. A summary of the main conclusions is placed at the end of the next chapter.

Reminder on the law of small numbers

The law of small numbers is a statistical law which derives from the binomial law. It governs the counting of objects of any kind. Thus, the counts by unit of time of a radioactivity counter which records the radioactive decay of a substance behaves according to the law of small numbers. If the number of "tops" by unit of time is small enough, then their distribution (number of intervals of time with 0, 1, 2, 3 "tops", etc.) conform to the distribution of Poisson. One of the remarkable consequences of the distribution of Poisson is that the variance s^2 of a series of counts is equal to the mean of these counts: $m = s^2$.

In the case of cell counts, the law of small numbers also applies with some conditions. One considers in this case that the surface of counting is constituted by a large number of elementary surfaces (condition 1: n is a large number) and that the probability of presence of a cell on each of these elementary surfaces is low (condition 2: p is small). The third condition is that the counts must be independent from each other. Thanks to the relation $m = s^2$, the comparison of the variance and the mean allows estimating the homogeneity and the "correctness" of the counts.

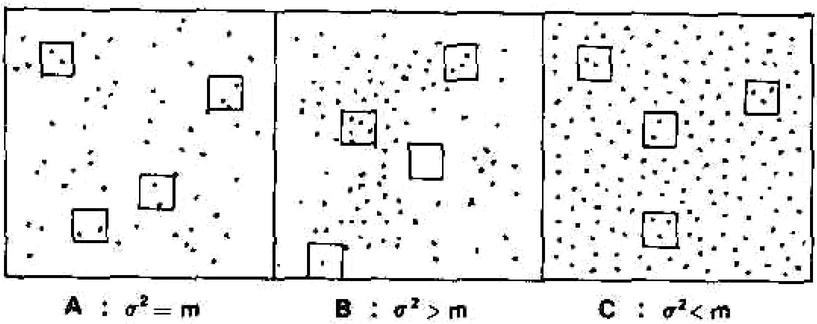


Figure 11.1. This figure illustrates the relationships between the variance and the mean according to the law of small numbers. This law governs the countings of particles such as blood cells, bacteria, etc. In the case of an enumeration, the law of small numbers applies when the probability of an event (presence of a cell) is low and when the number of possible locations (elementary surfaces) for this event is high (p small, n high). Finally, the various counts must be independent. In this case it can be demonstrated that the variance is equal to the mean as in the situation *A* (see text for the situations *B* and *C*).

(Reproduced from S. Frontier, Méthode statistique, Masson, 1980)

In the "real" life, these conditions are not always satisfied. Three cases can appear which are depicted in Figure 11.1. On the left of this figure (case A), the law of small numbers is verified and the scattering of the counts fits $m = s^2$. In case B, the scattering of the counts is more important than predicted by the law of small numbers. In this case the law is not verified because the "particles" that are counted tend to attract each other and to form aggregates ($s^2 > m$).

In the case C, on the contrary, the particles that are counted tend to arrange in a more regular way than chance would allow and consequently the variance of samples is lower than the mean $s^2 < m$). This arrangement is met for example when particles tend to repel each other and to consequently equalize the distances separating them.

What criticized the investigators?

The demonstration of the investigators is summarized by the figure below. It represents the "standardized" distribution of the mean difference of the counts made in duplicate from the results in the notebooks of E. Davenas. According to the investigators the scattering of the counts is narrower than the scattering which one would expect according to the law of small numbers. In other words, the profile of skyscraper ("pooled data") should be closer to the profile of a tumulus ("theoretical best").

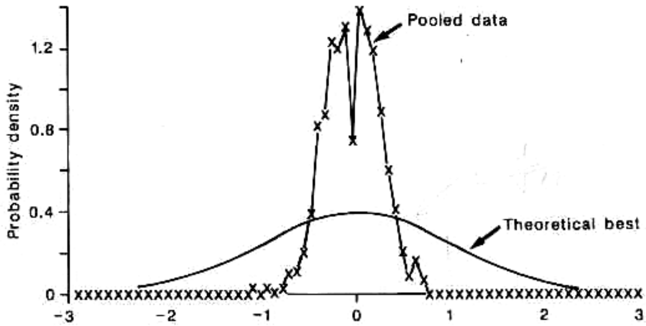


Fig. 4 ***Comparison of measured departures of duplicate normalized readings from their means with the gaussian distribution expected.***

(*Nature* 1988, 334:289)

Figure 11.2. The aim of this figure, which was repeatedly reproduced, was to demonstrate that the counts of basophils of the experiments reported in the laboratory notebooks were biased. But, as described in the text, on one hand, the normalized variable was calculated in an erroneous manner leading to "dramatize" the narrowness of the distribution and, on the other hand, experimental data published in 1981 showed that when the density of basophils increased, the modeling by the law of small numbers did not fit the counts in "real life".

This figure had once again the honor to appear in the issue of *Nature* of October 27th, 1988 when J. Maddox published a text of 4 pages intended to put an end to the debate. He commented the figure in these terms:

> "[*The figure*] 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." [1]

On the other hand, according to the reasoning of the investigators, if one proceeds in the same way with the blind counts performed during their expertise, one notices that the law of small numbers is respected with a variance almost equal to one (Figure 11.3).

The conclusion of the authors of the report was simple: the data were biased, consciously or unconsciously. As we will demonstrate, the reality is far from being such an obvious fact. First of all, the investigators made an error – a mathematical error – by applying without precaution a formula of statistics.

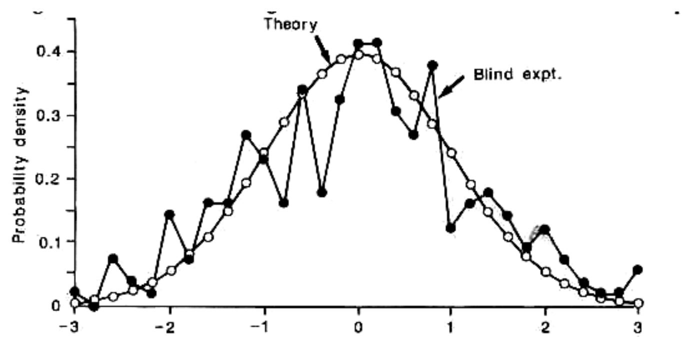


Fig. 5 ***Same as Fig. 4 except that data derive from duplicated readings within the blind experiments only.***

(*Nature* 1988, 334:289)

Figure 11.3. This Figure in the investigation report was the counterpart of Figure 11.2. The variations within every pair of counts of basophils in blind experiments were shown. A distribution with a standard deviation close to 1 (unit) was obtained (abscissa at half-peak) thus indicating, according to the investigators, that in the blind conditions of the investigation the counts of basophils fitted, as expected for cell counts, the law of small numbers. Put into perspective with Figure 11.2, this figure would be thus the proof of an experimental bias for the results of Figure 11.2. However, a distribution of according to the law of small numbers (Poisson's law) with a variance close to 1 (unit) should be obtained only in ideal conditions, without added statistical noise. Furthermore, the formula for the "reduced variable" used for this figure was calculated with an wrong "expected standard deviation" (see text).

What was the expected sampling error?

Contrary to what one would be entitled to demand from a scientific investigation report, the report of *Nature* gave very few explanations on the methods and did not present tables of the data included in the analysis. Let us remind that this analysis was not peer-reviewed. Nevertheless, it was clearly indicated that the calculation of the distribution of both curves described above was performed in the following way:

> "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 unit (1)." [2]

Actually this method of calculation – with the results obtained during the investigation – gives a distribution with a standard deviation close to 1, which is compatible with the law of small numbers (if we suppose that there are no disturbances other than the statistical fluctuations). From W. Stewart's original data, we selected all the counts which were performed in duplicate through blind experiments during the investigation. These counts are in fact from experiment F (cf. chapter 9), each experimental point having been counted twice by each of the two experimenters. These 134 counts of basophils (67 pairs of counts) are reproduced in the appendix so that the interested readers can make their own analysis.

By using the method described in the report, we obtain the distribution in Figure 11.4.

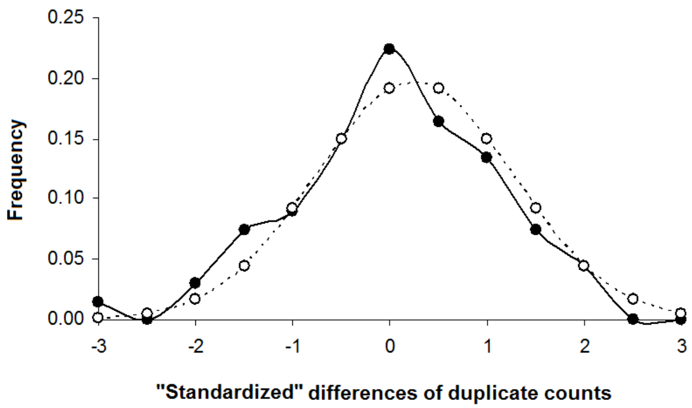


Figure 11.4. With the counts of basophils of experiment F (Figure 11.3) made in duplicate during the investigation of *Nature* (cf. Chapter 8), we calculate, as W. Stewart did, the standardized difference of the couples of counts using the method of the latter. We find the result presented in the investigation report of *Nature* (reproduced in Figure 11.3). The curve in dotted line represents the theoretical distribution of the reduced centered variable (that is with mean = 0 and standard deviation = 1). The abscissa of every point is the upper margin of the considered interval.

In accordance with what the investigators noticed, the standard deviation of the distribution is thus close to 1 (by calculation we find exactly 1.09) what would be actually compatible with a distribution in compliance with the law of small numbers (we can estimate it graphically as the abscissa corresponds to half of the height of the peak).

Moreover, it is rather surprising to find a standard deviation close to 1 because we saw that the cell counts of experiment F were very erratic and that the counting of basophils had been completed only due to J. Maddox and W. Stewart's insistence, precisely for "statistical analysis". It would not be surprising – on the contrary – to obtain an observed variance wider than the expected variance because of a disturbance due to the addition of "statistical noise". Paradoxically, knowing the experimental conditions of experiment F, one can now consider that the results of W. Stewart were "too beautiful" to be true! But, since these results fitted the conclusion that the investigators had predefined, they did not push the analysis farther.

In fact, the formula used to calculate the standardized variable was *false*!

It seems that in their haste, the investigators forgot some rules of statistics.

What formula was it necessary to use?

The formula applied by W. Stewart as indicated in the text of the report (see above) on every pair of counts of basophils (x, y) to calculate the "standardized" variable is the following one:

$$\frac{x-(\frac{x+y}{2})}{\sqrt{\frac{x+y}{2}}} \qquad (1)$$

One substracts from every count x the mean of the two values x and y and one divides by "the expected sampling error", that is, always according to W. Stewart, the square root of the mean.

However – contrary to the statement of the investigators – the expected sampling error (standard deviation) *is not the square root of the mean of the two counts*. Indeed x, on one hand, and the mean of x and y, on the other hand, are two random variables which are *not independent*. It seems that the investigators applied without precaution the classic formula:

$$\frac{x-\mu}{\sigma_x}$$

This formula allows standardizing the distribution of the values of a random variable X of theoretical mean μ and of theoretical standard deviation σ_x. In the present case, we have to deal with the *difference of two random variables*. Let us resume the formula (1). The numerator can be simplified as follows:

$$x-(\frac{x+y}{2}) = 1/2x - 1/2y$$

Because x and y are two independent random variables, we can now estimate the expected standard deviation from this linear combination. Indeed, the variance of the linear combination $aX + bY$ of two independent random variables X and Y is:

$$\sigma^2_{(aX+bY)} = a^2\sigma^2_X + b^2\sigma^2_Y$$

The expected standard-deviation is therefore:

$$\sigma = \sqrt{1/4\sigma^2_X + 1/4\sigma^2_Y}$$

Indeed in this case a = 1/2 and b = -1/2 and since with the law of small numbers the expected variance is equal to the mean:

$$\sigma = \sqrt{1/4x + 1/4y}$$

We obtain thus the value of the normalized variable:

$$\frac{x-y}{\sqrt{x+y}} \qquad (2)$$

In fact, the correct approach consisted in studying the distribution of the differences of the couples (x, y) with an expected variance equal to $x + y$ (because in this case a = 1 and b = 1). The formula (2) is then immediately obtained.

Moreover, one easily calculates that the method used by W. Stewart underestimates the standard deviations of the standardized variable. Indeed, the formula (1) used by W. Stewart can be simplified as:

$$\frac{x-y}{\sqrt{2(x+y)}}$$

With the correct method, the variance of the standardized variable is *twofold higher*; in other words, the standard deviation is 1.4 times higher than the correct value. If one needs to be convinced of the reality of this error, one can redo the same calculations of the standardized variable on a series of pairs randomly obtained according to the law of small counts. Thus, the results obtained with 1000 pairs randomly generated are the following ones:

1) Method of W. Stewart: variance = 0.50 (i.e. standard deviation = 0.71)

2) Correct method: variance = 1.01 (i.e. standard deviation = 1.00)

We notice that, as expected by the calculation above, the calculation done by W. Stewart gives a variance half of the variance calculated with the correct method (and thus a standard deviation 0.71 fold the correct value).

The first consequence is that the standard deviation of the duplicate counts in notebooks of E. Davenas is not too narrow as J. Maddox and W. Stewart hammered over and over again (it is 1.4 times wider). The second consequence is that the value of the standard deviation of the figure supposed to show the conformity of the distribution to the law of small numbers for blind experiments is not close to 1 but to 1.4 (the exact calculation gives 1.48; this calculation can be verified from the results of the experiment F given in appendix). Nevertheless, thanks to their "results", the investigators asserted in their report:

> "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." [3]

One remembers that the "counters of basophils" had drawn the attention of W. Stewart and J. Maddox on the very large differences of cell densities from one chamber of the hemocytometer to another one (cf. chapter 9). This was not surprising given the "method" that W. Stewart had imposed to achieve his goals. The latter indeed pipeted and pipeted again on numerous occasions the low volumes of cell suspensions in spite of our warnings in front of these modifications of the technique. W. Stewart had decided to proceed in this way to get enough duplicate samples for the two experimenters. The investigators carefully avoided reporting this crucial problem for the credibility of their demonstration.

Consequently, thanks to a false formula which minimized the standard deviation (it multiplied it by 0.71) and thanks to poor experimental conditions which increased the dispersal of the measures (standard deviation = 1.48), one compensating the other, the investigators were lucky enough to get a standard deviation close to 1! (exactly equal to $1.53 \times 0.71 = 1.09$) [4]. This visible good "modeling" of the results with the law of small counts strengthened the character of uninfringeable mathematical law which could not be challenged. The investigators could assert thanks to this "result": "the two observers were counting as accurately as could be expected"!

A few years after the investigation, in 1992, M. Schiff (already met in Chapter 3) studied the laboratory notebooks of of E. Davenas to redo the same calculations as W. Stewart. He noticed this:

> "From the laboratory notebooks, I made what Stewart should have made: to analyze the dispersion of blind counts in controls. The dispersion rarely reached values as high as those produced during the investigation of Maddox or were seldom as small as the values exhibited by Stewart." [5]

Furthermore, there is another aspect concerning the dispersion of the counts, but not a mathematical argument, that the investigators did not take into account.

The article of Nancy of 1981

In 1981, well before the affair, an article of H. Gérard *et al* was published concerning technical improvements of the test of basophil degranulation.[6] This article proposed a simple method using centrifugation to obtain basophil-rich blood cells. The researchers of Nancy made the following observation: the law of small counts was verified when the number of basophiles was low, but *was not verified when the cell concentration increased.* Indeed, they noticed that the standard deviation of the measure decreased compared with the expected value when the number of basophils increased thanks to the enrichment in cells. The relationship between cell density and standard deviation is shown in Figure 11.5. Thus, with a mean count of basophils at 75, the standard deviation was only 4.5, which is a standard deviation approximately *half* the value calculated with the law of small counts.

This article was well known of J. Benveniste and his collaborators and these results confirmed their own observations. Therefore the fact that the law of low counts was not respected in all cases for basophil counts did not shock them and this notion had been incorporated in their daily practice.

It is probably why J. Benveniste considered the arguments on dispersion as "theoretical" and did not attach to them – wrongly – so much importance as the investigators. It is moreover surprising to see *a posteriori* that in his answers, J. Benveniste evaded this question. To theoretical arguments, he answered by pragmatic arguments:

> "The central argument of the report bears on sampling errors and statistics of which we are so aware that we performed numerous control experiments. They show similar standard deviations and variances in 24/28 comparisons of blind (4 series, 90 samples, without the Israeli experiments) versus open (7 series, 183 samples) control wells.
>
> Did the "experts" understand that the real controls are water or anti-IgG most often paired with anti-IgE? […] Other allergy tests correlate with degranulation (reference in article), so why is it that our statistics fit for 40 to 70 per cent degranulation at regular ligand concentration and not for the same high dilution."[7]

The empirical results of H. Gérard *et al*, which were obtained not in the ideal world of the mathematics but in "real life" are obviously very interesting for our demonstration. They confirm the idea that the law of small counts is not adequate to model the dispersion of basophils when they are counted in a hemocytometer under a microscope.

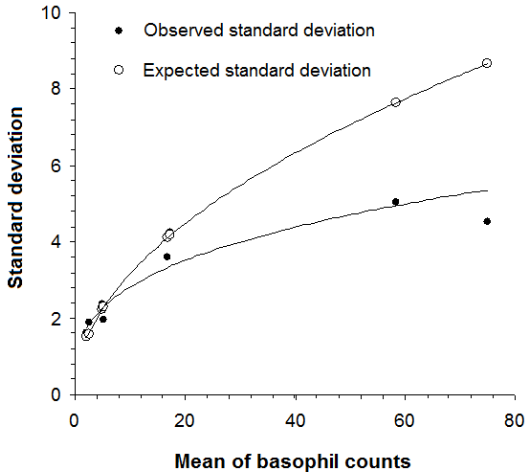


Figure 11.5. This figure, performed with the results of the article of H. Gérard *et al* (1981), indicates how vary – in real experimental conditions – the standard deviation according to the number of basophils counted in samples (black circles). Twenty samples were counted to determine each mean with its standard deviation. This result is thus described in the article: “In reality, multiple counts made on various bands of the hemocytometer with samples variously enriched with basophils show that the distribution is gaussian with a standard deviation lower than the square root of the mean, especially for high enrichments”. Indeed, if the standard deviation of the counts was in accordance with the law of small numbers, it should be approximately equal to the square root of the mean number of basophils counted on the various samples (white circles).

From the Massif Central to the Alpes

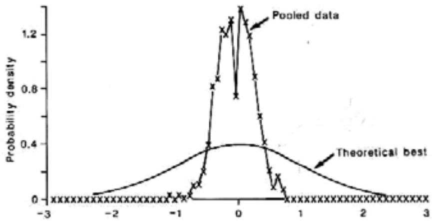
Let us examine again the figure of the investigation report of *Nature* that was supposed to demonstrate a bias due to the experimenter because the distribution of mean difference concerning the duplicate counts was too narrow. Using a software we can generate virtual “counts of basophils” by taking into account both the results of the article of H. Gérard *et al* and the calculation error of W. Stewart evidenced in the previous chapter. We suppose that we “count” in duplicate 1000 wells containing basophils with a mean of 75 and a standard deviation of 4.5 (these values are obtained from the article of H. Gérard et al; see Figure 11.5). We apply the formula that had been used in the report to calculate the “standardized” distribution of the duplicate counts.

As depicted in Figure 11.6, the distribution obtained by taking into account both the dispersion of basophils in the reality of the laboratory and the erroneous calculation of W. Stewart has an aspect which is very similar to the figure in the report of *Nature*. Therefore, the thin aspect of the distribution has nothing surprising and is thus not related to any data "manipulation". If somebody is to blame, it is rather among the investigators.

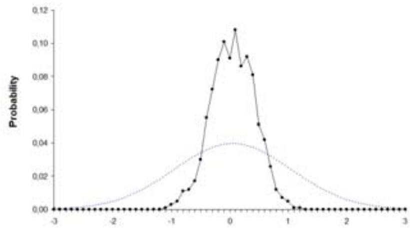
We thus notice that the "central argument" of the sampling error was an idol with feet of clay. Having apparently the solidity of a theorem of mathematics, it nevertheless suffered from two crippling defects. On one hand, an error in the use of the statistics led to more "dramatic" results, supporting the expectations of the investigators. On the other hand, being new to this sector of research, the investigators had not taken into account the pragmatic knowledge of the researchers who considered this "anomaly", which the investigators have blown up out of proportion, as unexceptional. *A posteriori*, the description by W. Stewart telling how he sketched the "narrow" curve by analyzing the results from the laboratory notebooks of E. Davenas is particularly interesting:

> "From the evening, I analyzed with the computer the data of the laboratory notebooks. I introduced data, and I made arrangements to draw a graphic curve to compare them with the optimal results which we could achieve, according to a mathematical modeling. At the end of the second day, it was obvious that the agreement of the data was far too precise. It was not possible that the data fit so well." [8]

We see perfectly how the key for reading of W. Stewart worked: the results "must" fit a predefined mathematical model. At no time, the doubt crossed the mind of W. Stewart as for the legitimacy of the model, its accuracy and its limits.[9]



The graph in the report of *Nature* 1988 (334:289 et 335:762).



Modelling taking into account both the results of the article of Gérard *et al* (1981) and the erroneous calculation of the investigators.

Figure 11.6. The left figure was twice published in *Nature* and was frequently reproduced in the press. Calculated by W. Stewart from the laboratory notebooks of E. Davenas, this figure was supposed to be the demonstration that repeated counts (duplicate counts) were submitted to a bias of the experimenter. The flattened curve (*theoretical best*) corresponds to the standardized distribution that one should observe if the counts fitted the Poisson distribution (law of small numbers) which classically governs this type of enumeration. The result reported by W. Stewart (*pooled data*) was narrower. According to him, this was the proof that there was a bias related to the experimenter. In other words, the results were too good to be true. The reality was not however so simple. On one hand, W. Stewart used an erroneous formula (see text). On the other hand, he did not take into account the knowledge of the researchers who practised this technique and had noticed and published that the variance of the counts of basophils was lower than expected according to the law of small numbers. By taking into account these results (and with the incorrect formula), the modeling of a series of "counts of basophils" generated by a computer (right figure) gives a result that is very close to the result obtained by W. Stewart (left figure). Therefore the counts of basophils extracted from the laboratory notebooks of E. Davenas had nothing exceptional and they could not be suspected of a bias (either unconscious or voluntary). These calculations and this modeling can be easily done again by the reader (see text).

Notes of end of chapter

---

[1] J. Maddox. Waves caused by extreme dilutions. *Nature*, October 27th, 1988, p. 762.

[2] J. Maddox, W. Stewart and J. Randi. "High dilution" experiments a delusion. *Nature* July 28th, 1988, p. 290.

[3] *Ibid.* p. 289.

[4] It is possible that W. Stewart included in his calculations some counts made in duplicate in the experiments D, E and G. If we include these counts the conclusions are the same and we find $1.50 \times 0.71 = 1.06$ for standard deviation.

[5] M. Schiff. Un cas de censure dans la science. L'affaire de la mémoire de l'eau, p. 238 (translation of the French text). The same idea can be found p. 143 of the English version of the book [*Schiff, M. (1998). The Memory of Water: Homoeopathy and the Battle of Ideas in the New Science, London, Thorsons Publishers.*]

[6] H. Gérard, B. Legras, D.A. Moneret-Vautrin. Le test de dégranulation des basophiles humains (TDBH). Intérêt d'un leucoconcentration et du calcul statistique appliqué au taux de dégranulation [*The human basophils degranulation test (HBDT). Leucoconcentration and statistical calculation applied to the degranulation rate*]. *Pathologie Biologie* 1981 ; 29 : 137-142.

[7] J. Benveniste. Dr Benveniste replies. *Nature*, July 28th, 1988, p. 291.

[8] P. Alfonsi. Au nom de la science, p. 92.

[9] On October 8th, 2014, there was a symposium at the Unesco in Paris entitled "*Biology in the Light of Theoretical Physics: New Frontiers in Medicine"*. On this occasion, the mathematician Cédric Villani, who received the Fields Medal in 2010, did a talk entitled "*Memory, oblivion and reproducibility: an outside view on a never solved controversy"* in which he reported some thoughts on the case of the "memory of water". Having read the present book as a source of documentation, he stated about the statistical approach of *Nature*'s investigators: "Stewart forgets experimental data according to which basophils at high concentration are not scattered according to the law of small numbers, so that usual statistical calculations should be modified. To top it off, it appears that Stewart did an elementary error in his calculation of the variance. More than 15 years later after the disputed article, Francis Beauvais recalculates by taking into account these two effects and shows that statistics that were considered by *Nature* as a mathematical proof of error are in fact fully compatible with successful experiments. It is finally a fascinating textbook case about poor use of statistics that could be analysed for lessons of epistemology or statitics".