Agreement Between Theory and Experiment

Sometimes theory and experiment are both correct but do not agree with each other; sometimes a wrong theory agrees with experiment. One must therefore be careful not to jump to conclusions.

Amikam Aharoni

A theory is usually expected to explain existing experimental results and to predict new results, while an experiment is usually expected to check the validity of existing theories and to gather data for modifying them. This approach is normally presented to students as foolproof as if it were one of the basic laws of "good" science. In practice these goals are achieved in some cases, but sometimes the comparison between a theory and an experiment can be very misleading. Here I am going to discuss these unusual cases, to warn against the possible pitfalls. They may or may not be rare, but in any event it is important to bear in mind that they do exist.

One can certainly find examples in any subfield of physics, but I limit myself to cases taken from the literature on magnetism, simply because I happen to know something about this particular field.

A good experiment

What is a good experiment? It is easiest to give an example from the older literature, one that has already been proved right and has passed the test of time. Consider the experiment now known as the de Haas–van Alphen effect. Textbooks tell us that Wander Johannes de Haas and Pieter Marinus van Alphen found an oscillatory behavior in the magnetic susceptibility of bismuth as a function of the applied magnetic field. But those books usually give examples from *later* experiments, where one can see many cycles of these oscillations. In the original experiment[1] however, the crystals were not as good, the magnetic field was not as high and the temperature was not as low as in later experiments. Figure 1 shows what de Haas and van Alphen actually measured. Clearly they actually had very little to support their conclusion that "the susceptibility of bismuth at [liquid] hydrogen temperature is found to be a periodic function of the field."

Moreover, even the way figure 1 presents the original data —with lines that guide the eye — does not convey the courage it took to call them periodic functions. To emphasize the point, figure 2 contains one set of data from figure 1 plotted without the connecting line, to show only the data that de Haas and van Alphen had. A periodic function is certainly not obvious at first sight.

The point is that in those days people had *faith* in their data, even without any theoretical support. The theory in this case, as in many other discoveries in physics, came later. In retrospect we know that the identification of figure 2 as a periodic function was justified. Therefore, the obvious conclusion is that once an experiment is checked for accuracy and systematic errors, the data should be published, even if there is no theory. Experience shows that if the experiment is good, the theory will come later.

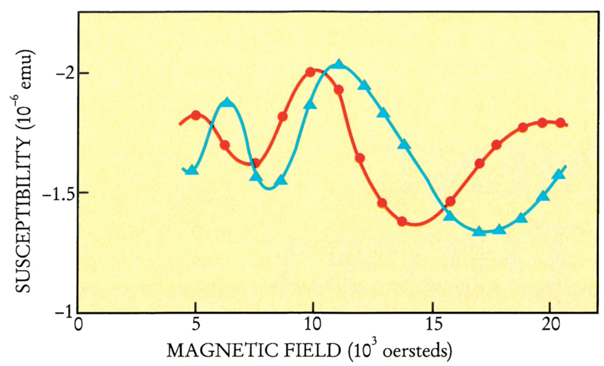
Still, this truth, which used to be taken as self-evident, has somehow been lost in recent years. Philip W. Anderson has urged the publication of experimental papers without a theory. Noting that both the Kondo effect and high-T_c superconductivity started as unexplained experiments, he observed: "Unless such studies were backed by enormous prestige in the form of the right author and institution, most referees were in the habit of rejecting them." (See PHYSICS TODAY, September 1990, page 9.) This point was emphasized by Roman Schrittwieser, who stated that when a new experiment disagrees "with the hitherto applied and generally accepted theory," the referee will "inevitably . . . criticize the *lack of a theory*!" (PHYSICS TODAY, February 1992, page 128).

This recent trend among referees is very dangerous to the development of physics, which has always been regarded as a basically experimental science. Thus, for example, when I was a young student, my physics teacher would repeat that it is wrong to say: The apple falls from the tree *because* it is pulled by Earth's gravitational field." It is wrong because Earth's gravitational field is a theory, while the falling of the apple is an experimental fact, which can be measured and verified. The correct way for a physicist to phrase the above statement is: We hypothesize the existence of a gravitational field because we observe the apple falling from the tree." The difference between these two statements is the basis for the whole philosophy of physics.

A bad experiment

The opposite of a good experiment is a bad experiment, but there are many degrees of badness. The extreme cases involve intentional manipulation of the experimental data and can simply be described as fraud, or at least as self-deception bordering on fraud. Irving Langmuir called such cases "pathological science" (in the October 1989 PHYSICS TODAY,

AMIKAM AHARONI *is the Richard Kronstein Professor of theoretical magnetism in the department of electronics at the Weizmann Institute of Science, in Rehovoth, Israel. His e-mail address is a.aharoni@ieee.org.*



ORIGINAL DATA for the susceptibility of bismuth at 14.2 K as a function of the applied magnetic field. The circles represent data taken with the field perpendicular to the binary axis; the triangles, data taken with the field parallel to that axis. (Adapted from ref. 1.) FIGURE 1

page 36). The resulting comments on this problem, in the March, April and December 1990 issues of PHYSICS TODAY make very interesting reading, because these things do happen. But here I am going to ignore them. When I refer to a bad experiment, I mean only an *honest* mistake. These cases occur more often than is usually believed, and in my mind they are bad enough to deserve our attention.

Charles Guillaud's experiment on the magnetic properties of MnBi crystallites is an example of a good experiment that could have easily been bad. Figure 3 shows his data on three magnetic properties, plotted as functions of the maximum applied field used for the measurement.[2] Note in particular the curve of the coercivity H_c, which shows a tendency to saturate at about 250–300 oersteds at a maximum applied field of 3000–4000 Oe, before it starts increasing again. Now, engineers have a rule of thumb that a maximum applied field of four times the coercivity is sufficient for taking such data. This rule works for many materials, but as figure 3 shows, it does not work for these crystallites of MnBi. If Guillaud had stopped his measurements at an applied field of around 3000 Oe (which is 10 times—not just 4 times—the measured coercivity), he would have obtained an experimental value of 300 Oe. Because he did proceed further, the experimental value of the coercivity became 5000 Oe. We thus see how easy it is to obtain the wrong order of magnitude by overlooking a seemingly unimportant detail of an experiment.

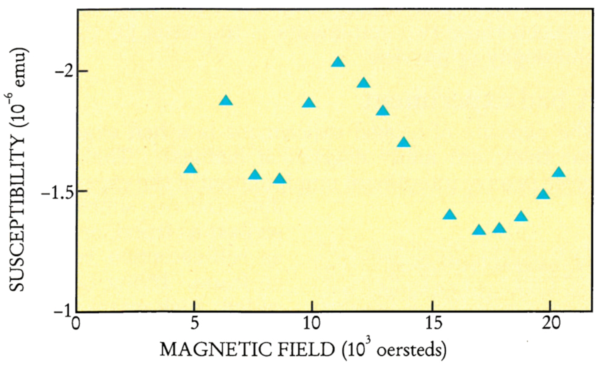
Wrong agreement

Coincidences can and do happen, and a wrong theory may agree with experiment by mere coincidence. Therefore, agreement with experiment does not always prove that a theory is correct, as the following example illustrates.

In the late 1950s and early 1960s investigators did many experiments on the magnetic properties of thin films made of an iron–nickel alloy called permalloy. They observed three types of walls between antiparallel domains and named them Bloch walls, Néel walls and cross-tie walls. The definitions of these walls and the energies associated with them are not important for our discussion; it is sufficient to know that there were experimental definitions and that the observers could tell the difference between these walls. In those days everybody took it for granted that both the Néel wall and the Bloch wall had a basically one-dimensional structure and that their energies (figure 4) were more or less as estimated by Louis Néel in 1955. Figure 4 is from a 1960 paper in which Martin Prutton[3] gave his theory of the energy of the cross-tie wall in permalloy films and compared this energy with those of the other two walls. Many other estimates (reviewed[4] in 1971) of the energies of these walls gave very similar results.

It is not easy to measure the wall energy, although there were some attempts in that direction.[4] One can argue, however, that nature will choose the state of lowest energy, and therefore figure 4 implies that Néel walls should be observed below a certain film thickness, Bloch walls above a different thickness and cross-tie walls in between. In this respect the predictions of figure 4 for the turnover from Néel walls to cross-tie walls and from cross-tie walls to Bloch walls were both in quite good agreement with experiment.[3] The crude approximations used in calculating the cross-tie wall energy seemed, therefore, well justified. This conclusion was repeated in Prutton's book,[5] along with figure 4. It remained unchallenged for almost a decade, because of the undeniable agreement between Prutton's estimation of the cross-tie wall energy and the various experiments.

The situation changed in 1969, when Anton LaBonte computed the Bloch wall energy in permalloy films, *without* the assumption of one-dimensionality, which had been taken for granted.[6] His energy values were half those of previous theories. My Ritz model for a Néel wall also



ONE OF THE TWO SETS of data in figure 1, plotted without the connecting line. FIGURE 2 reduced the accepted values for the energy of that wall, although the difference was not as dramatic as that of LaBonte.[7] In light of these new calculations, Prutton's estimate of the energy of a cross-tie wall, plotted in figure 4, suddenly became much too high. To fit the experimental turnover points between the different walls, the value for the cross-tie wall in figure 4 must be reduced by at least a factor of 3 to 4. It may even be more than that, because we now know that the best computed energies for Bloch[6] and Néel[7] walls are still minimized under certain constraints, which makes them upper bounds to the true minima.

We thus see how a theoretical value that was wrong by a factor of 3 or 4 still agreed quite well with experiment. Clearly the agreement was just a coincidence, which can happen in principle. But the important feature is that it *did* happen in this case. Moreover, this agreement, which existed for a decade, did not stop being an agreement because of a change in the experimental data, which remained the same between 1960 and 1970. Prutton's theoretical value of a cross-tie wall energy ceased to agree with experiment because *another* theory changed. It is very important to remember this point, because a theoretical value is rarely compared with a *directly* measured value. More often than not, there are hidden assumptions in the analysis of the experimental data, and these assumptions can turn out to be wrong. In this case the wrong assumption was that the Bloch wall is one-dimensional. It is thus always necessary to question points that are taken for granted before drawing conclusions from the agreement between a theory and an experiment.

Wrong disagreement

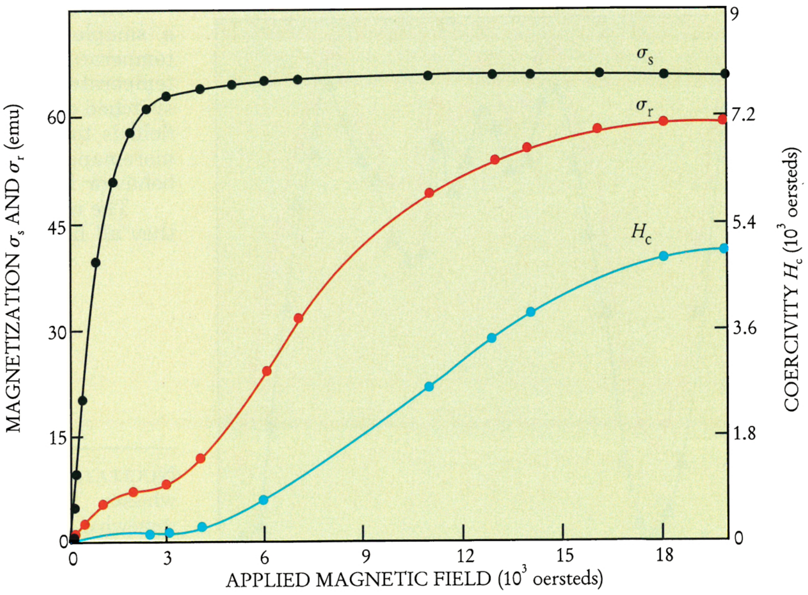
A theory may be correct and still disagree with experiment. Before giving a real example, consider a hypothetical case, based again on Prutton's calculation of the energy of a cross-tie wall in permalloy films. Suppose that by some miracle somebody in the 1960s came up with a theoretical estimate that we know now to be correct. This estimate would have been much below Prutton's, shown in figure 4; and in those days (when the other two curves were considered correct) it would not have agreed with experiment. The intercepts with those other two curves would have been way off the experimental values.

This example did not happen, but it could have happened, in principle. There is no *a priori* reason to assume that a disagreement between theory and experiment cannot result from some mistake in a *different* theory that is simply taken for granted. In this hypothetical case the wrong assumption again would have been that the Bloch wall is one-dimensional. The important point is that it is possible for a correct theory to disagree with experiment.

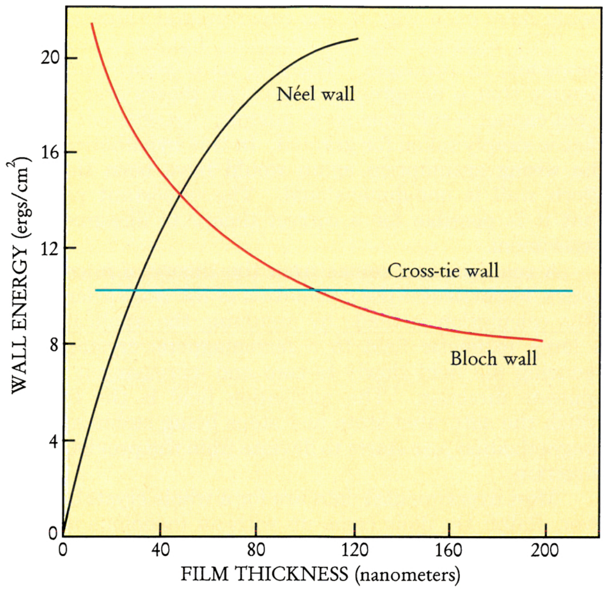
A real example comes from the theory of micromagnetics, in particular the calculation of the nucleation field. This calculation is made by assuming the application of a large magnetic field to a ferromagnetic crystal, then reducing the field slowly to zero and increasing it in the opposite direction. The nucleation field is the field at which the saturated state just stops being stable and the magnetization just starts to deviate from the previous field direction.

Most experimental data are for a whole crystal. However, Ralph De Blois came up with an ingenious method for measuring this nucleation field in a small part of a very long iron crystal, or whisker.[8] Figure 5 shows typical results based on a minor modification[9] of the original De Blois experiment. (The two curves in this figure represent slightly different experimental techniques, which do not make a difference for the present discussion.) One sees that the nucleation field varies considerably from one point to the other along the whisker.

The theoretical nucleation field for a very long iron crystal is 560 Oe. It is seen from figure 5 that this theoretical value is pretty close to the experimental values at some points along the whisker, but much lower values are observed at other points. Moreover, in some parts of the whisker, not shown in the figure, the experimental value was a fraction of an oersted—three orders of magnitude below the theoretical value. And at least one point with a nucleation field of 1 Oe or less was observed in each of the whiskers studied, thus giving a very large range of experimental values. However, it was shown that rather large surface imperfections are present near regions with low nucleation field values. In fact, a good correlation was reported between the volume of surface defects and the reduction in the nucleation field at places



MAGNETIZATION AND COERCIVITY as functions of the maximum applied magnetic field in "large" crystallites of MnBi. Top curve: saturation magnetization. Middle curve: remanent magnetization. Bottom curve: coercivity. (Adapted from ref. 2.) FIGURE 3



THEORETICAL ENERGY OF DOMAIN WALLS in permalloy films as a function of the film thickness. The crossovers from a Bloch to a cross-tie wall and from a cross-tie to a Néel wall agreed with experiment for 9 years and were taken as a proof of the validity of the cross-tie wall calculation. The plotted energy of the cross-tie wall was later found to be too large by a factor of at least 3 to 4. (Adapted from ref. 3.) FIGURE 4

on the whisker.[9]

The conclusion was that the nucleation theory, which assumes a smooth surface, does not agree with experiment only where surface defects are present. The theory must obviously be modified to include the effect of a rough surface before it can be applied to a realistic material, which we still cannot do properly. However, besides the assumption of a perfectly smooth crystal, there is nothing wrong with this theory, because it agrees with experiment where the surface is smooth. It should be noted that once a magnetization reversal is nucleated, it very easily propagates over the whole crystal. Therefore, if one measurement is done for the crystal as a whole, it measures the nucleation at the worst point. This is like a chain breaking at a force that is just sufficient to break its weakest link, even if all the other links are very much stronger. And because it is impossible to have a perfectly smooth whisker, measurements of a whole whisker always lead to a nucleation field of less than 1 Oe; values as small as 0.1 Oe or 0.01 Oe are even encountered. The difference between theory and experiment is several orders of magnitude, but the theory is still basically correct. In practice, this theory is used only for small particles, for which the surface roughness is less important and for which it does agree with experiment.[10]

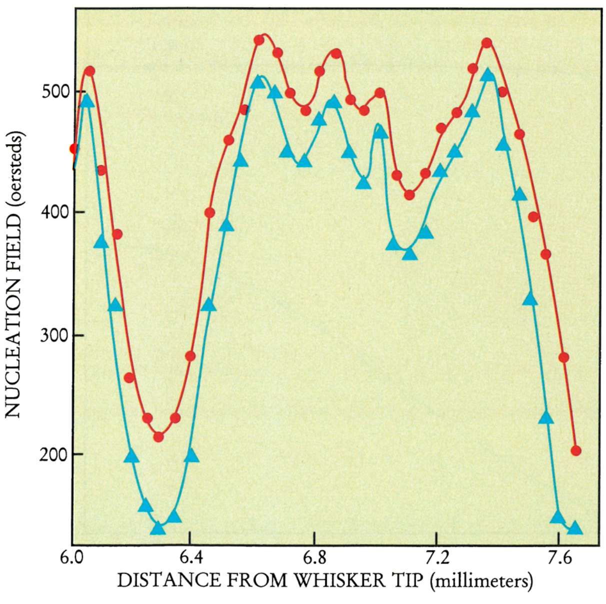
This argument was thus understood and agreed upon for many years. Then in a very long paper, Marc A. Pinto stated[11] that "the [theoretical] nucleation field may be as large as 500 Oe while the experimental value in the same conditions is below 0.1 Oe." Pinto said that I "discussed" the problem, but without going into details, he declared only that "this remains to be explained." An exchange ensued with no change of mind on either side.[11–13] Pinto's conclusion was that the whole theory of micromagnetics has to be discarded. I believe it still has some vitality in it.

Bad theories

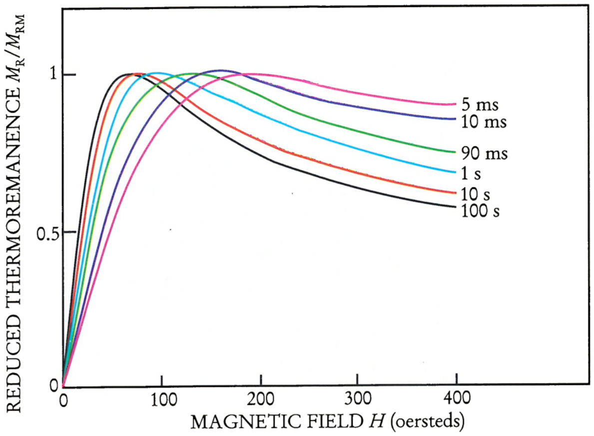
Some theories can be shown to be bad even without comparing them with experiment, because they use impossible arguments or calculate a small term while neglecting a larger one or make some other mistakes. This type of bad theory often agrees with experiment, not because of some coincidence but because the experimental results are built into the theory's assumptions. One example is the peak in the thermoremanence curve.

The thermoremanence curve is measured by taking a sample of magnetic material to a sufficiently high temperature, then cooling it slowly to a low measuring temperature in a dc magnetic field H, which is finally switched off. The remanence (the magnetization after the field is turned off) thus obtained is defined as the thermoremanence; figure 6 shows a typical example of its behavior for a particular material[14] as a function of H.

The important feature of the curves in figure 6 is that they all have a maximum, which does not seem to make



NUCLEATION FIELD as a function of position along an iron whisker. The circles and triangles represent data from somewhat different measuring techniques. The theory agrees well with the highest points, but not with the rest of the data. (Adapted from ref. 9.) FIGURE 5



THERMOREMANENCE in the spin glass $Er_{0.4}Sr_{0.6}S$ at 1.32 K. The time shown on each curve is the time between switching off the field H and measuring the remanence M_R. The latter is normalized to be 1 for the maximum remanence M_{RM} in each of the curves. (Adapted from ref. 14.) **FIGURE 6**

sense. A maximum means that the remanence at a higher field, say around 400 Oe in this particular figure, is smaller than the remanence at some smaller field. Hence a certain small field can arrange more spins in its direction than can a larger field. It is not easy to think of a physical mechanism by which a smaller field can have a bigger effect than a larger one. And yet figure 6 was very typical. Dozens of measurements on various compositions of spin glasses at various temperatures showed a similar peak in the curve. And this peak was observed only in spin glasses, never in a powder of ferromagnetic particles.

There was even a theory, or rather a computer simulation, that showed such a peak.[15] The experimenters quoted it, even though it was not likely that any of them could understand it, because the theory was not written in a way that clarified the physical assumption. It can best be described by the immortal words that the English physicist Edmund Stoner used to criticize a different theory: "It is not unfair to say that obscurities in the presentation do not seem to arise wholly from the inherent complexities of the problem."[16] In other words, it was one of those theories that is just not meant to be understood. In the case of the thermoremanence curve, the theory[15] actually had the peak built into it, in the assumption that some nonphysical properties are possible only in spin glasses. But the assumption was well hidden. Researchers should have realized that the computer simulation was not necessarily correct just because it gave an observed peak. It was also necessary that the simulation be based on a model that has physical meaning.

The late Peter Wohlfarth and I tried to explain this anomalous peak by an experimental artifact.[17] We argued that switching off a magnetic field is not a step function. Even before the measurement starts, there is a finite time during which the magnetization decays. If it takes a longer time to switch off a larger field, then more magnetization can decay before the effect of the larger field can be measured, so that the result may be smaller in the larger field. Our model had a physical meaning, but it also had the obvious drawback that the same mechanism must also apply to ferromagnets, for which no peak had ever been observed.

Nobody accepted our suggested mechanism until the very same peak was measured in magnetite particles.[18] It turned out that the peak had never been observed in ferromagnetic particles only because nobody had measured them at the right temperature range. Of course, this experiment does not in itself prove that our suggested mechanism is correct, and so it remains only a possible mechanism. However, the experiment does prove that any theory that claims that the peak is unique to spin glasses, because of their special properties, is undoubtedly wrong.

The only way to avoid bad theories is to have referees insist that the authors specify clearly the *physical* assumption of the theory. For example, if A is neglected with respect to B, it cannot be taken for granted without saying so, nor is it sufficient to mention it somewhere between equations 56 and 57. Such an assumption belongs in the introduction, if not in the abstract. Because this rule has already been so nicely phrased by Stoner (in 1950), I will conclude by quoting him:

> It is important to bear in mind that the validity of a mathematical (or physico-mathematical) argument in itself cannot be confirmed merely by an agreement with experiment of approximate relations which may have been derived; and that the value of a theory as interpretive of observable phenomena cannot be properly assessed until the essential details in the argument from the premises to the conclusions have been clearly presented.[16]

I based this article on a series of lectures that I gave in various places as an IEEE Magnetics Society distinguished lecturer for 1993.

References

1. W. J. de Haas, P. M. van Alphen, Commun. Phys. Lab. Univ. of Leiden, report 212a (1930).
2. C. Guillaud, J. Phys. Radium **12**, 492 (1951).
3. M. Prutton, Philos. Mag. **5**, 625 (1960).
4. A. Aharoni, J. de Phys. (Paris) **32**, suppl. C-1, 966 (1971).
5. M. Prutton, *Thin Ferromagnetic Films*, Butterworth, London (1964), figure 3.18, p. 52.
6. A. E. LaBonte, J. Appl. Phys. **40**, 2450 (1969).
7. A. Aharoni, J. Appl. Phys. **41**, 186 (1970).
8. R. W. De Blois, C. P. Bean, J. Appl. Phys. **30**, 225S (1959).
9. A. Aharoni, E. Neeman, Phys. Lett. **6**, 241 (1963).
10. A. Aharoni, Rev. Mod. Phys. **34**, 227 (1962).
11. M. A. Pinto, Phys. Rev. B **38**, 6824 (1988).
12. A. Aharoni, Phys. Rev. B **43**, 8670 (1991).
13. M. A. Pinto, Phys. Rev. B **43**, 8671 (1991).
14. J. Ferré, J. Rajchenbach, H. Maletta, J. Appl. Phys. **52**, 1697 (1981).
15. W. Kinzel, Phys. Rev. B **19**, 4595 (1979).
16. E. C. Stoner, Rep. Prog. Phys. **13**, 83 (1950), see p. 140.
17. A. Aharoni, E. P. Wohlfarth, J. Appl. Phys. **55**, 1664 (1984).
18. M. el-Hilo, K. O'Grady, IEEE Trans. Magn. **26**, 1807 (1990). ■