Proceedings of the 1997 Winter Simulation Conference
ed. S. Andradóttir, K. J. Healy, D. H. Withers, and B. L. Nelson

DOCTORAL COLLOQUIUM KEYNOTE ADDRESS

CONDUCT, MISCONDUCT, AND CARGO CULT SCIENCE

James R. Wilson

Department of Industrial Engineering
North Carolina State University
Raleigh, North Carolina 27695, U.S.A.

In the South Seas there is a cargo cult of people. During the war they saw airplanes land with lots of good materials, and they want the same thing to happen now. So they've arranged to make things like runways, to put fires along the sides of the runways, to make a wooden hut for a man to sit in, with two wooden pieces on his head like headphones and bars of bamboo sticking out like antennas—he's the controller—and they wait for the airplanes to land. They're doing everything right. The form is perfect. It looks exactly the way it looked before. But it doesn't work. No airplanes land. So I call these things cargo cult science, because they follow all the apparent precepts and forms of scientific investigation, but they're missing something essential, because the planes don't land.

Now it behooves me, of course, to tell you what they're missing. ... It's a kind of scientific integrity, a principle of scientific thought that corresponds to a kind of utter honesty—a kind of leaning over backwards. For example, if you're doing an experiment, you should report everything that you think might make it invalid—not only what you think is right about it: other causes that could possibly explain your results; and things you thought of that you've eliminated by some other experiment, and how they worked—to make sure the other fellow can tell they have been eliminated.

... In summary, the idea is to try to give *all* of the information to help others to judge the value of your contribution; not just the information that leads to judgment in one particular direction or another.

—Richard P. Feynman, *"Surely You're Joking, Mr. Feynman!"* (1985)

ABSTRACT

I will elaborate some principles of ethical conduct in science that correspond to Richard Feynman's well-known precepts of "utter honesty" and "leaning over backwards" in all aspects of scientific work. These principles have recently been called into question by certain individuals who allege that such rules are based on a misunderstanding of "how science actually works" and are therefore potentially "damaging to the scientific enterprise." In addition to examining critically the general basis for these allegations, I will discuss the particular relevance of Feynman's ideals to the field of computer simulation; and I will emphasize the need for meticulous validation of simulation models together with exact reproducibility and unimpeachable analysis of experiments performed with those models. Finally I will discuss the ethical dilemmas inherent in the peer review system, and I will offer some concrete suggestions for improving the process of refereeing primary journal articles.

1. INTRODUCTION

Much has been written recently about what constitutes scientific misconduct, and public esteem for science has been damaged by high-profile episodes such as the "cold fusion case" at the University of Utah (Huizenga 1993) and the "David Baltimore case" at MIT (Elliott and Stern 1997). Against this backdrop I will examine several claims about principles of ethical conduct in science that were made by James

Woodward and David Goodstein of the California Institute of Technology in an article entitled "Conduct, Misconduct and the Structure of Science," which appeared in the September 1996 issue of the *American Scientist*. The gist of the principles in question is summarized in the quotation by Richard Feynman given above. I will argue that these principles are especially relevant to the field of computer simulation, and I will elaborate my view that Feynman's ideals of "utter honesty" and "leaning over backwards" constitute a mandate for meticulous validation of simulation models together with exact reproducibility and unimpeachable analysis of experiments performed with those models. Several key references are highlighted in this discussion—in particular, see the pamphlets entitled *On Being a Scientist* (1995) and *Honor in Science* (1986). Interested individuals are invited to examine the relevant literature and to judge for themselves the validity of the arguments given here.

2. "THE SCIENCE OF THINGS THAT AREN'T SO"

In addition to performing Nobel Prize–winning research, the American physicist Irving Langmuir explored extensively a subject he called "pathological science," defining this as "the science of things that aren't so." Although he never published his investigations on this subject, he presented a colloquium on pathological science at General Electric's Knolls Atomic Power Laboratory on December 18, 1953. Subsequently Robert N. Hall, one of Langmuir's former colleagues at General Electric, transcribed and edited a recording of Langmuir's presentation so that it could be published in the October 1989 issue of *Physics Today*. Langmuir and Hall (1989) should be required reading for everyone who pursues a career in scientific research.

This article is a fascinating account of famous cases of self-deception by scientists working in a broad diversity of disciplines. Perhaps the most remarkable of these cases concerns the discovery of N rays by the French physicist René Blondlot in 1903. This exotic form of radiation was claimed to penetrate inches of aluminum while being stopped by thin foils of iron. When N rays impinged on an object, Blondlot claimed a slight increase in the brightness of the object; but he admitted that great experimental skill was needed to detect the effect of these rays.

During the period from 1903 to 1906, over 300 papers were published on N rays by 100 scientists and medical doctors around the world (Nye 1980). When the American physicist Robert W. Wood learned about the discovery of N rays, he went to France to observe Blondlot's experimental procedure. At that time Blondlot was using a spectroscope fitted with an aluminum prism to measure the refractive indices of N rays. Although Blondlot's experiments were performed in a darkened room, a small red (darkroom) lantern enabled Blondlot to see a graduated scale for measuring to three significant figures the position of a vertical thread coated with luminous paint. The thread was supposed to brighten as it crossed the invisible lines of the N-ray spectrum. According to Langmuir and Hall (1989), Wood asked Blondlot the following question:

> . . . from just the optics of the thing, with slits 2 mm wide, how can you get a beam so fine that you can detect its position to within a tenth of a millimeter?

Blondlot is reported to have given this reply:

> That's one of the fascinating things about N rays. They don't follow the ordinary laws of science . . . You have to consider these things by themselves. They are very interesting but you have to discover the laws that govern them.

His suspicions aroused at this point, Wood used the cover of the darkened room to remove the prism and put it in his pocket. Wood then asked Blondlot to repeat some of his measurements. With the critical component of the experimental apparatus missing, Blondlot obtained exactly the same results. In a letter that was published in *Nature*, Wood (1904) exposed Blondlot's experiments on N rays as a case of self-deception. Although Wood's letter killed research on N rays outside France, it is interesting to note that the French Academy of Sciences chose Blondlot to receive the 1904 Le Conte Prize—even though the other leading candidate was Pierre Curie, who together with Marie Curie and Henri Becquerel had shared the 1903 Nobel Prize in physics for pioneering work on radioactivity.

Langmuir and Hall (1989) also discuss a number of other anomalous phenomena, and they analyze the main symptoms of pathological science (or cargo cult science, to use Feynman's more colorful expression). These symptoms are summarized in Table 1. The case of N rays exhibits all of these symptoms. It is important to bear these symptoms in mind when considering the validity of certain claims made by Woodward and Goodstein (1996) about ethical conduct in science. Numerous cases of pathological science involving pseudoscientific cranks are discussed in the book *Fads and Fallacies in the Name of Science* by Martin Gardner (1957). Some famous cases

Table 1: Langmuir's Symptoms of Pathological Science

1. The maximum effect that is observed is produced by a causative agent of barely detectable intensity, and the magnitude of the effect is substantially independent of the intensity of the cause.
2. The effect is of a magnitude that remains close to the limit of detectability or, many measurements are necessary because of the very low statistical significance of the results.
3. There are claims of great accuracy.
4. Fantastic theories contrary to experience are suggested.
5. Criticisms are met by *ad hoc* excuses thought up on the spur of the moment.
6. The ratio of supporters to critics rises up to somewhere near 50% and then falls gradually to oblivion.

of self-deception by legitimate scientists are detailed on pages 107–125 of the book *Betrayers of the Truth* by William Broad and Nicholas Wade (1982).

3. THE LOGICAL STRUCTURE OF SCIENCE

3.1 Baconian Inductivism vs. Data Selection

As a basis for their discussion of how science actually works, Woodward and Goodstein examine critically the theories of the scientific method that are due to Francis Bacon ([1620] 1994) and Karl Popper (1972). Baconian inductivism prescribes that scientific investigation should begin with the careful recording of observations; and as far as possible, these observations should be uninfluenced by any theoretical preconceptions. When a sufficiently large body of such observations has been accumulated, the scientist uses the process of induction to generalize from these observations a hypothesis or theory that describes the systematic effects seen in the data.

On the contrary, Woodward and Goodstein assert that "Historians, philosophers, and those scientists who care are virtually unanimous in rejecting Baconian inductivism as a general characterization of good scientific method." Woodward and Goodstein argue that it is impractical to record all one observes and that some selectivity is required. They make the following statement:

> But decisions about what is relevant inevitably will be influenced heavily by background assumptions, and these ... are often highly theoretical in character. The vocabulary we use to describe the results of measurements, and even the instruments we use to make the measurements, are highly dependent on theory. This point is sometimes expressed by saying that all observation in science is "theory-laden" and that a "theoretically neutral" language for recording observations is impossible.

I claim that in the context of computer simulation experiments, this statement is simply untrue. By using portable simulation software, we can achieve exact reproducibility of simulation experiments across computer platforms—that is, the same results can be obtained whether the simulation model is executed on a notebook computer with a 16-bit operating system or on a supercomputer with a 64-bit operating system. Moreover, the accumulation of relevant performance measures within the simulation model can be precisely specified in a way that is completely independent of any theory under investigation. Thus we can attain Feynman's ideal of "a kind of utter honesty" in which every simulation analyst has available the same information with which to evaluate the performance of proposed theoretical or methodological contributions to the field. In my view, it is impossible to overstate the fundamental importance of this advantage of simulated experimentation; and we are deeply indebted to the developers and vendors of simulation software who have taken the trouble and expense to provide us with the tools necessary to achieve the reproducibility that is an essential feature of all legitimate scientific studies.

According to Woodward and Goodstein, Baconian inductivism leads to the potentially erroneous and harmful conclusion that data selection and overinterpretation of data are forms of scientific misconduct, while a less restrictive view of how science actually works would lead to a different set of conclusions. In many prominent cases of pathological science, the root of the problem was data selection ("cooking") that may have been subconscious but was nonetheless grossly misleading. In addition to the case of Blondlot's nonexistent N rays, Langmuir and Hall (1989) and Broad and Wade (1982) detail several other noteworthy cases of such cooking and overinterpretation of experimental data in the fields of archaeology, astronomy, geology, parapsychology, physics, and psychology. I claim that whatever the theoretical deficiencies of Baconian inductivism may be, they have no bearing on the field of computer simulation; moreover, there are sound practical reasons for insisting that researchers in all fields should avoid selection or overinterpretation of data that has even the appearance of pathological science.

3.2 Validating vs. "Cooking" Simulation Models

Because simulationists work far more closely with the end users of their technology than specialists in many other scientific disciplines, we are sometimes exposed to greater pressure from clients or sponsors to fudge or "cook" our models to yield anticipated or desired results. With the advent of powerful special- and general-purpose simulation environments including extensive animation capabilities, such model-cooking is far easier for simulationists to carry out than it is for, say, atmospheric physicists.

In addition to intentional model-cooking, there is the danger of unintentional self-deception resulting from faulty output analysis. In many of the cases of self-deception documented in Langmuir and Hall (1989) and Broad and Wade (1982), the most notable common feature was the experimenter's attempt to detect visually an extremely faint signal in situations where auxiliary clues enabled the experimenter to know for each trial observation whether or not the signal was supposed to be present. For example in the N-ray experiments described previously, Blondlot could see the scale measuring the current position of the thread coated with luminous paint. With each change in the thread's position, Blondlot knew if he was supposed to see a brightening of the thread—and thus he was able to deceive himself into "seeing" effects that other experimenters could not reproduce. In the context of simulation experiments, animation can be one of the primary visual means for self-deception. Equally dangerous is faulty output analysis based on visual inspection of correlograms, histograms, confidence intervals, etc., computed from an inadequate volume of simulation-generated data. With all of these simulation tools, there is the ever-present danger of seeing things that simply do not exist or of not seeing things that do exist.

To guard against cooking a simulation model or its outputs, simulationists should place much greater emphasis on meaningful, honest validation of their models as accurate representations of the corresponding target systems. To reemphasize the role of validation in the field of computer simulation, we need fundamental advances in both the practice and theory of model validation. So far as I know, the simulation literature contains very little documentation of real-world applications in which a simulation model was carefully validated. A comprehensive methodology for validating simulation models is detailed in Knepell and Arangno (1993) and Sargent (1996), but it not clear that many practitioners and researchers have given due consideration to either the implementation or the extension of this methodology. I believe that we need to pay much greater attention to simulation model validation in teaching and research as well as in practical applications.

3.3 Popperian Falsificationism

Next we turn to the falsificationist ideas of Karl Popper. According to this theory of the scientific method, we test a hypothesis by deducing from it a prediction that can be tested in an experiment. If the prediction fails to hold in the experiment, then the associated hypothesis is said to be falsified and must be rejected. Thus Popperian falsificationism requires a scientist to hold a hypothesis tentatively, to explore and highlight the ways in which the hypothesis might break down, to uncover and scrutinize evidence contrary to the hypothesis rather than discarding or suppressing such evidence, and in general to avoid exaggeration or overstatement of the evidence supporting the hypothesis. Perhaps the most forceful statement of this view of science was given by Richard Feynman in the quotation at the beginning of this article.

According to Woodward and Goodstein, there are also serious deficiencies in Popperian falsificationism as a general theory of good scientific method:

> One of the most important of these is sometimes called the Duhem-Quine problem. We claimed above that testing a hypothesis H involved deriving from it some observational consequence O. But in most realistic cases such observational consequences will not be derivable from H alone, but only from H in conjunction with a great many other assumptions A (auxiliary assumptions, as philosophers sometimes call them). ... It is possible that H is true and that the reason that O is false is that A is false.
>
> ...It may be true, as Popper claims, that we cannot conclusively verify a hypothesis, but we cannot conclusively falsify it either.

The most distinctive feature of computer simulation experiments is that the simulationist has complete control over the experimental conditions via (a) the random number streams driving the simulation model's stochastic input processes, and (b) the deterministic inputs governing model operation. Thus in simulated experimentation it is possible to isolate the effects of auxiliary assumptions, so that the Duhem-Quine problem can be effectively resolved. However as several colleagues have pointed out, often practitioners fail to evaluate the effects of auxiliary assumptions in large-scale simulation projects. This failure may be due to the lack of a well-documented, widely recognized methodology for addressing the Duhem-Quine problem in the context of simulation studies. Future simulation research should focus on the development of such methodology together with a comprehensive investigation of the connections between methods for solving the Duhem-Quine problem and methods for validating a simulation model.

Beyond their theoretical objections to Popperian falsificationism, Woodward and Goodstein claim that this approach has serious practical disadvantages:

> Suppose a novel theory predicts some previously unobserved effect, and an experiment is undertaken to detect it. The experiment requires the construction of new instruments, perhaps operating at the very edge of what is technically possible, and the use of a novel experimental design, which will be infected with various unsuspected and difficult-to-detect sources of error. As historical studies have shown, in this kind of situation there will be a strong tendency on the part of many experimentalists to conclude that these problems have been overcome if and when the experiment produces results that the theory predicted. Such behavior certainly exhibits anti-Popperian dogmatism and theoretical "bias," but it may be the best way to discover a difficult-to-detect signal. Here again, it would be unwise to have codes of scientific conduct or systems of incentives that discourage such behavior.

The scenario of Woodward and Goodstein is a remarkably accurate description of the experimental setting in which occurred all of the cases of pathological science detailed by Langmuir and Hall (1989) and Broad and Wade (1982). Moreover, this scenario describes the notorious cold fusion experiments of Martin Fleischmann and B. Stanley Pons as documented in the book *Cold Fusion: The Scientific Fiasco of the Century* by John R. Huizenga (1993). It seems clear that in such a scenario, the scientist's foremost concern should be to avoid lapsing into self-deception and pathological science.

4. THE SOCIAL STRUCTURE OF SCIENCE

Woodward and Goodstein claim that ultimately inductivism and falsificationism are inadequate as theories of science because they fail to account for the psychology of individual scientists and the social structure of science. First Woodward and Goodstein consider the role of social interactions in scientific investigation:

> Suppose a scientist who has invested a great deal of time and effort in developing a theory is faced with a decision about whether to continue to hold onto it given some body of evidence. . . . Suppose that our scientist has a rival who has invested time and resources in developing an alternative theory. If additional resources, credit and other rewards will flow to the winner, perhaps we can reasonably expect that the rival will act as a severe Popperian critic of the theory, and vice versa. As long as others in the community will perform this function, failure to behave like a good Popperian need not be regarded as a violation of some canon of method.

Turning next to the psychology of individual scientists, Woodward and Goodstein explore the difficulty of sustaining the necessary long-term commitment of time and resources to a hypothesis without mentally exaggerating the supporting evidence and downplaying the contrary evidence—especially in the early stages of a project when belief in the hypothesis may be extremely fragile:

> All things considered, it is extremely hard for most people to adopt a consistently Popperian attitude toward their own ideas.
>
> Given these realistic observations about the psychology of scientists, an implicit code of conduct that encourages scientists to be a bit dogmatic and permits a certain measure of rhetorical exaggeration regarding the merits of their work, and that does not require an exhaustive discussion of its deficiencies, may be perfectly sensible. . . . In fact part of the intellectual responsibility of a scientist is to provide the best possible case for important ideas, leaving it to others to publicize their defects and limitations.

In contrast to this point of view, Peter Medawar, the winner of the 1960 Nobel Prize in medicine for his work on tissue transplantation, made the following statement in his book *Advice to a Young Scientist* (Medawar 1979, p. 39):

> *I cannot give any scientist of any age better advice than this: the intensity of the conviction that a hypothesis is true has no bearing on whether it is true or not.* The importance of the strength of our conviction is only to provide a proportionately strong incentive to find out if the hypothesis will stand up to critical evaluation.

(The emphasis in the quoted statement is Medawar's.) Like Langmuir and Hall (1989), Medawar's *Advice to a Young Scientist* should be required reading for individuals at all stages in their scientific careers.

Over the past twenty years, I have accumulated considerable experience in mediating extremely acrimonious disputes between researchers acting as "severe Popperian critics" of each other's work. Much of this hard-won experience was gained during the nine years that I served as a departmental editor and former departmental editor of the journal *Management Science*. To avoid reopening wounds which have not had much time to heal, I will not go into the particulars of any of these cases; but I feel compelled to draw some general conclusions based on these cases.

In every one of the disputes that I mediated, the trouble started with extensive claims about the general applicability of some simulation-based methodology; and then failing to validate these claims independently, reviewers and other researchers proceeded to write up and disseminate their conclusions. This in turn generated a heated counterreaction, usually involving claims of technical incompetence or theft of ideas or both. Early in my career I served as the "special prosecutor" in several of these cases. Later on I moved up to become the "judge," and in the end I was often forced to play the role of the "jury" as well. In every one of these cases, ultimately the truth emerged (as it must, of course)—but the process of sorting things out involved the expenditure of massive amounts of time and energy on the part of many dedicated individuals in the simulation community, not to mention the numerous professional and personal relationships that were severely damaged along the way. In summary, I claim that when individual researchers violate Feynman's precepts of "utter honesty" and "leaning over backwards," the cost to the scientific enterprise of policing these individuals rapidly becomes exorbitant.

5. SCIENCE AS CRAFT

Woodward and Goodstein question the general validity of the following principle:

> Scientists must report what they have done so fully that any other scientist can reproduce the experiment or calculation.

They claim that science has a large "skill" or "craft" component, and that

> Conducting an experiment in a way that produces reliable results is not a matter of following algorithmic rules that specify exactly what is to be done at each step.

This may be true of some areas in the biological sciences and other experimental sciences in which the behavior of living organisms or the functioning of complicated instrumentation may not be well understood, but this does *not* apply to computer simulation experiments. We can and must insist on exact reproducibility of simulation experiments; and this should, in fact, be a matter of following precisely stated, fully documented algorithms.

There is of course a large "craft" component in building and using simulation models. Different individuals presented with the same system to be modeled will neither build identical simulations nor apply those models in precisely the same way, just as different researchers in any other scientific discipline will neither build the same experimental apparatus nor carry out exactly the same experimental protocol to study a given effect. Nevertheless in these situations different simulationists should be able to reproduce each other's results in order to judge the significance and limitations of the conclusions based on the experiments in question. More generally, there is a large "craft" component in doing simulation research just as there is a large "craft" component in doing other types of scientific research—but this state of affairs does not mitigate the need for reproducibility of the main experiments associated with such research.

6. PEERS AND PUBLICATION

6.1 Is the Scientific Paper a Fraud?

Woodward and Goodstein cite Peter Medawar's (1991) paper entitled "Is the Scientific Paper a Fraud?" to argue that because most archival papers in the scientific literature do not accurately portray the way scientific research is actually done, these papers fail to measure up to Feynman's ideal of "leaning over backwards." It is certainly true that primary journal articles in the scientific literature do not document all of the mistakes, dead ends, and backtracking that are an inevitable part of virtually every successful scientific investigation. Medawar (1982, p. 92) himself admitted that

> I reckon that for all the use it has been to science about four-fifths of my time has been wasted, and I believe this to be the common lot of people who are not merely playing follow-my-leader in research.

In my view, the fundamental issue here is that there simply is not enough space in all the scientific journals to document the way that science is actually done; moreover no one has the time to absorb all the final results even in a relatively narrow area of specialization, much less to read the associated background material. Nowadays many high school students are sufficiently sophisticated to realize that primary journal articles are vehicles for efficiently communicating significant discoveries rather than for documenting the processes by which those discoveries were made. Moreover, this issue is rapidly becoming moot because of current trends toward complementing the printed version of a primary journal article with comprehensive supporting documentation (such as appendices containing lengthy proofs or detailed descriptions of experimental protocols) archived on a World Wide Web server that is maintained by the journal's sponsoring organization.

6.2 Problems with the Peer Review System

Finally Woodward and Goodstein examine the peer review system for evaluation of research proposals and primary journal articles, concluding that the conflict of interest inherent in asking competitors to evaluate each other's work has inflicted genuine distress on the system. In my own experience, by far the most common form of misconduct by peer reviewers has nothing to do with conflicts of interest; instead the problem is simple dereliction of duty by reviewers who cannot be bothered to read and evaluate carefully the work of other researchers. Although this remark applies to evaluation of research proposals as well as refereeing of primary journal articles, I am most concerned with problems in refereeing. In my judgment, the problem of nonperformance by referees has reached epidemic proportions, and I believe it is urgently necessary for the scientific community to address this scandalous state of affairs.

In preparing these remarks I solicited comments from numerous colleagues not only in the simulation community but also in the "hard" scientific disciplines, and I have been startled by the vehemence of their agreement with my evaluation of the current state of the refereeing system. Based on numerous conversations with colleagues in biology, electrical engineering, industrial engineering, mathematics, and statistics, I have a sense that problems with refereeing are much worse in these fields than in the simulation community. Perhaps the most egregious failure of the refereeing system in recent years was the publication of the initial paper on cold fusion by Fleischmann and Pons (1989a). This paper was published in the *Journal of Electroanalytical Chemistry* in just four weeks; and a long list of errata soon followed (Fleischmann and Pons 1989b)—including the name of M. Hawkins, a coauthor who was somehow omitted from the original paper. A detailed account of this infamous episode can be found on pp. 218–220 of Huizenga (1993).

6.3 Refereeing Remedies

The two main reasons for breakdowns in the operation of the refereeing system are (a) misconceptions by referees about the job they are supposed to do, and (b) lack of incentives for doing a good job of refereeing. As Gleser (1986) points out, many referees think that a manuscript must be checked line by line for errors; and seeing that this will be extremely time-consuming, they continually put off the task. On the contrary, the referee's main responsibility is to serve the editor as an "expert witness" in answering certain key questions about the manuscript—*and most of these questions can be answered under the assumption that the manuscript is error-free.* These key questions are given in Table 2 and are elaborated in Forscher (1965), Gleser (1986), and Macrina (1995, pp. 84–89) along with general guidelines for refereeing that should be required reading for every research worker in the field of computer simulation.

If a paper passes the initial screening that consists of answering questions 1–8 in Table 2, then it is necessary to undertake the verification of technical correctness required to answer questions 9 and 10. If competent referees had scrutinized the initial paper on cold fusion by Fleischmann and Pons (1989a) with the objective of answering questions 9 and 10 in Table 2, then the fatal flaws in this work would have been uncovered immediately. In my view it is imperative that we protect the simulation literature against the long-lasting stigma that results from permitting the publication of technically incorrect work. If everyone in the simulation community followed the guidelines in Table 2 for preparing referee's reports, then I believe our problems with peer review would largely disappear.

Additional tips on effective refereeing are given by Waser, Price, and Grosberg (1992). A set of questions similar to those given in Table 2 can be found on the home page of the *ACM Transactions on Modeling and Computer Simulation* by using the URL

Table 2: Key Questions to be Answered in a Referee's Report

1. Are the problems discussed in the paper of substantial interest? Would solutions of these problems materially advance knowledge of theory, methods, or applications?
2. Does the author either solve these problems or else make a contribution toward a solution that improves substantially upon previous work?
3. Are the methods of solution new? Can the proposed solution methods be used to solve other problems of interest?
4. Does the exposition of the paper help to clarify our understanding of this area of research or application? Does the paper hold our interest and make us want to give the paper the careful reading that we give to important papers in our area of specialization?
5. Are the topic and nature of this paper appropriate for this journal? Are the abstract and introduction accessible to a general reader of this journal? Is the rest of the paper accessible to a readily identified group of readers of this journal?
6. Are the clarity and readability of the manuscript acceptable? Is the writing grammatically correct?
7. Does the manuscript contain an adequate set of references? Is adequate credit given to prior work in the field upon which the present paper is built?
8. Is the material appropriately organized into an effective mix of text, figures and tables? Are data given in tables better presented in figures or in the text?
9. Is the work technically correct? Are the main conclusions justified by the experimental data and by logically valid arguments? Are the theorems stated and proved correctly given the assumptions? In practical applications of the theoretical results, do the authors check the validity of the underlying assumptions?
10. Are there gaps in the discussion of the experimental methods or results? If there are such gaps, can the closing of these gaps be considered (i) essential, (ii) desirable, or (iii) interesting? Are the experimental methods described in sufficient detail so that other investigators can reproduce the experiments?

`http://www.acm.org/pubs/tomacs/review/review.html`.

There remains the question of adequate incentives for good refereeing. In reviewing preliminary versions of these remarks, several individuals complained about general lack of editorial feedback on (a) the strengths and weaknesses of their reviews, and (b) the issues identified in other referees' reports on the same paper. As a routine professional courtesy, editors should include such feedback with their letters of appreciation to referees. Moreover, editors should strive to ensure that individuals who provide prompt and thorough refereeing will receive comparable service when those individuals submit their own papers for review. Ultimately refereeing is one of the professional responsibilities that each of us must fulfill to ensure the vitality of our chosen field, but doing this job well should be a source of pride and satisfaction commensurate with that of our other professional contributions to the field.

7. CONCLUSION

To close these remarks, I come back to the opening quotation by Richard Feynman. In essence my central thesis is simply this: as scientists we should all strive to live up to the standards of professional conduct so memorably articulated by Feynman. Sophisticated (or merely sophistic) rationalizations of anything short of this standard serve no constructive purpose and should be avoided. In a time when public esteem for science has been damaged by high-profile cases of scientific misconduct, we in the simulation community have a unique opportunity to lead the way in achieving Feynman's ideals not only in the design and execution of our experimental procedures but also in our collective response to the challenges of responsible, professional peer review.

ACKNOWLEDGMENTS

Although they may not have found these remarks to be completely congenial, I thank David Goodstein and James Woodward for their comments on this article. I also thank the following individuals for insightful suggestions concerning this article: R. H. Bernhard, L. F. Dickey, S. E. Elmaghraby, and S. D. Roberts (North Carolina State Univ.); F. B. Armstrong and B. J. Hurley (ABB Power T&D Co.); C. Badgett (U.S. Navy Joint Warfare

Analysis Center); K. W. Bauer (Air Force Institute of Technology); R. C. H. Cheng (Univ. of Kent at Canterbury); M. M. Dessouky (Univ. of Southern California); P. L'Ecuyer (Univ. de Montréal); D. Goldsman (Georgia Institute of Technology); P. Heidelberger (IBM T. J. Watson Research Center); M. Irizarry (Univ. of Puerto Rico); R. W. Klein (Regenstrief Institute for Health Care); R. E. Nance (Virginia Polytechnic Institute and State Univ.); B. L. Nelson (Northwestern Univ.); A. A. B. Pritsker (Pritsker Corp. and Purdue Univ.); R. G. Sargent (Syracuse Univ.); B. W. Schmeiser (Purdue Univ.); T. J. Schriber (Univ. of Michigan); R. W. Seifert (Stanford Univ.); A. F. Seila (Univ. of Georgia); P. M. Stanfield (ABCO Automation, Inc. and North Carolina Agricultural and Technical State Univ.); J. J. Swain (Univ. of Alabama–Huntsville); and M. A. F. Wagner (Boeing Information Services). The quotation by Richard Feynman appearing at the beginning of this article is reproduced with permission from W. W. Norton & Company.

REFERENCES

Bacon, Francis. [1620] 1994. *The novum organum; with other parts of "The great instauration."* Chicago: Open Court.

Broad, William, and Nicholas Wade. 1982. *Betrayers of the truth*. New York: Simon and Schuster.

Elliott, Deni, and Judy E. Stern, eds. 1997. *Research ethics: A reader*. Hanover, New Hampshire: University Press of New England, for the Institute for the Study of Applied and Professional Ethics at Dartmouth College.

Feynman, Richard P. 1985. *"Surely you're joking, Mr. Feynman!": Adventures of a curious character*. New York: W. W. Norton & Co.

Fleischmann, Martin, and Stanley Pons. 1989a. Electrochemically induced nuclear fusion of deuterium. *Journal of Electroanalytical Chemistry* 261 (2A): 301–308.

Fleischmann, Martin, and Stanley Pons. 1989b. Errata. *Journal of Electroanalytical Chemistry* 263: 187–188.

Forscher, Bernard K. 1965. Rules for referees. *Science* 150:319–321.

Gardner, Martin. 1957. *Fads and fallacies in the name of science*. New York: Dover Publications.

Gleser, Leon J. 1986. Some notes on refereeing. *The American Statistician* 40 (4): 310–312.

Honor in science. 1986. 2d ed. New Haven, Connecticut: Sigma Xi, The Scientific Research Society.

Huizenga, John R. 1993. *Cold fusion: The scientific fiasco of the century*. New York: Oxford University Press.

Knepell, Peter L., and Deborah C. Arangno. 1993. *Simulation validation: A confidence assessment methodology*. Los Alamitos, California: IEEE Computer Society Press.

Langmuir, Irving, and Robert N. Hall. 1989. Pathological science. *Physics Today* 42 (10): 36–48.

Macrina, Francis L. 1995. *Scientific integrity: An introductory text with cases*. Washington, D.C.: ASM Press.

Medawar, Peter B. 1979. *Advice to a young scientist*. New York: BasicBooks.

Medawar, Peter B. 1982. *Pluto's republic*. Oxford: Oxford University Press.

Medawar, Peter B. 1991. Is the scientific paper a fraud? In *The threat and the glory: Reflections on science and scientists*, ed. David Pyke, 228–233. Oxford: Oxford University Press.

Nye, Mary Jo. 1980. N-rays: An episode in the history and psychology of science. *Historical Studies in the Physical Sciences* 11 (1): 127–156.

On being a scientist: Responsible conduct in research. 1995. 2d ed. Washington, D.C.: National Academy Press.

Popper, Karl R. 1972. *The logic of scientific discovery*. 3d ed. London: Hutchinson.

Sargent, Robert G. 1996. Verifying and validating simulation models. In *Proceedings of the 1996 Winter Simulation Conference*, ed. J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain, 55–64. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Waser, Nickolas M., Mary V. Price, and Richard K. Grosberg. 1992. Writing an effective manuscript review. *BioScience* 42 (8): 621–623.

Wood, Robert W. 1904. The *n*-rays. *Nature* 70 (1822): 530–531.

Woodward, James, and David Goodstein. 1996. Conduct, misconduct and the structure of science. *American Scientist* 84 (5): 479–490.

AUTHOR BIOGRAPHY

JAMES R. WILSON is Professor and Director of Graduate Programs in the Department of Industrial Engineering at North Carolina State University. He was *Proceedings* Editor for WSC '86, Associate Program Chair for WSC '91, and Program Chair for WSC '92. Currently he serves as a corepresentative of the INFORMS College on Simulation to the WSC Board of Directors. He is a member of ASA, ACM, IIE, and INFORMS.