

T/6, wh

See SAM/329

Communication as the root of scientific progress
Lecture presented by Joshua Lederberg
Sixth International Conference
International Federation of Science Editors
Woods Hole, Mass. October 16, 1991.

Thank you John [Kendrew]. Your tacit question about what I am doing here I share, but I've done lots of mischief together with Ken Warren, some of it here -- it's usually come out to a good end. Yes, I am very interested in scientific information. I don't do very much editorial work these days; I'm back working in the laboratory after a lapse of twelve years and that has kept me very busy trying to reacquaint myself with the literature of my own field. So I will offer you the perspective of a scientific reader. Now some people tell me that's a vanishing species! For anyone to say that, even with some sense of irony, is an atrocity. One of my main functions with my own laboratory group is that I try to be its principal reader. If something goes on in the world outside and none of us has heard about it for two or three weeks, I'm the one who feels responsible. I want to be alert to events that might have a very important bearing on the way we think about our own research, our planning, of the data coming in, of the sources of error.

Let me begin with a few truisms, just to be sure that we are operating on a common ground of reverence for the publication process. Publication is, to start with, just that! *Public*-ation. It converts private to public knowledge, in the service of registering a private claim of original authorship -- in science, of discovery. Above all, the act of publication is an inscription under oath, a testimony. It is accepted as valid until firm evidence to the contrary; and there is an extremely high standard of accountability for what is published under a given person's name. Just look at the daily headlines. It is the essential ingredient to make scientific work responsible in the sense that one cannot readily retreat from assertions that have been signed, delivered to the printer and made available to thousands.

These publically asserted claims also play an extremely important role in the allocation of resources, the ability of different scientists to survive in the competition with other legitimate claims for expenditures, for support of laboratories, for positions at the institutions, for space in the journals, for the attraction of students and collaborators. All these rest on those claims, the evidence for which in the end is in the public record. Both author and audience benefit from the successful assertion of those claims: especially credibility, that one doesn't have to spend an inordinate amount of time reexamining every detail of an individual's output if that person has established credibility through prior publication and exposure.

Publication also results in a repository, constructing the tradition of science. Up to this point it can hardly be anonymous in order to perform the functions that I have just indicated. But as time goes by, we have the reassimilation of the content of scientific work and as it settles in and survives the criticism that it should have had at its early stages of the process, it becomes the common tradition, the unquestioned shared wisdom -- often becoming anonymous by obliteration.

The literature is also a forum. It's a gladiatorial arena for competing claims, resolving

discrepancies in data or interpretation. There used to be oral duels, and we revel in stories like Pasteur's confrontation with Pouchet that finally put spontaneous generation to rest in 1864. Today, our battles are more often fought out in print, which is indeed appropriate because the testimony then becomes available to the universe, not simply to the immediate onlookers. Despite the opportunity for very broad dissemination, there is the paradox, nevertheless, that broadcast restricts individuals' access to feedback. The publication system, at least in principle, should allow a dialectic to appear in more symmetrical terms where anyone with something purposeful to say has a way to get into the system.

If the literature is a forum, it is also a rumen, a place for the digestion and assimilation of the variety of inputs where scientific claims go through a period of seasoning, modification, modulation. Even the truths look different five or ten years later regardless of explicit criticisms. We can expect a process of reinterpretation, a post-historical reexamination of the meaning of their terms. And now I only need to remind you of the term "imprimatur" (a wonderful metaphor): the imprinted witness that, an article having appeared in a refereed journal, it had survived a critical process, a conspiracy if you like, of the editors and the publishers and the referees -- that something has appeared which is worthy of the shared interest and precious attention of the community.

May I tell you what I do as a reading scientist today? Reading the scientific literature has been my primary vocation for fifty years. Books play a diminishing role. Today they are mostly for targeted reference. In the scientific domain, we rarely have the leisure today to read a book from cover to cover. A few biographies command attention. I just finished a proof copy of Carl Djerassi's life story: "The Pill, Pygmy Chimps and Degas' Horse"; another of that genre was Francois Jacob's revelation of the development of his scientific work: "The Statue Within. These are obviously not very contributory to the details on how to do my next experiment, but they tell me a lot about the scientific personality, providing object lessons and models for emulation. Rarely, I do see a work that compels total ingestion -- for example "Physiology of the Bacterial Cell" by Neidhardt, Ingraham and Schaechter. This is such a magnificent synthesis at a fairly elementary level of exposition that I really marveled at the deliberation and distillation that went into the telling. Wonderful books like that are rare. In printed form they surely will be the survivors of any electronic revolution. At an intermediary level of indispensibility as books in print format are the Annual Reviews. They are reference works for whatever you have to look up; but they also give a chance to browse through an enormous literature with some coherence. Compare an Annual Reviews of Genetics with current issues of the journal Genetics. Even if I had the time to read every article, I wouldn't have the background to be able to place each one of them in the appropriate context of what comes through. And I regard this as my home discipline! People will spend varying amounts of their time and energy as well in trying to understand what is going on in science beyond the window of their own specific work in their research and teaching.

There are about a dozen journals that I subscribe to and maybe seven or eight of them that I do scan from cover to cover: Nature, Science, Proceedings of the National Academy -- those are the very general ones. The Journal of Bacteriology, Microbiological Reviews, Genetics, Biochemistry. I pick up a hot paper now and then from The Scientist, and look at The Sciences, New Scientist, American Scientist, and Scientific American for general

scientific culture. That is a textual sampling, not immersion. You couldn't read every article in a critical and detailed fashion in just the ones I have listed within the number of hours that there are in the week. What you can do within a couple of hours a day is to scan that range of material and try to pick out those things that might be of interest. To follow the structure of argument just in one's own specialty, you must go to the detail of trying to check the numbers on the graphs and see if they match the authors assertions -- an arduous task.

We are well served by those kinds of journals in terms of maintaining a general currency about what is going on in the field, and they match very well the energy and interest and intellectual acuity that our scientific readers are able to put into the process. I see no occasion for those to be altered. Most scientists are very grateful for them: what thousands of scientists will share as common currency, to carry in their briefcase and read on the airplanes and the commuter rides, with all the convenience of the present print format.

My main problem is how do you reacquire, retrace that intellectual traffic. What do you do with all of your marginal notes and, how do you synthesize a coherent system of what you've read? Well to try to deal with this on a current basis, I have Gene Garfield's wonderful products. I get the weekly Current Contents on diskette with all of its embellishments. I eagerly await the five or six diskettes that have to be loaded, every week, and sometimes impatient about how long it takes to load them and get going with that week's literature. My stored profiles work out reasonably well, but have to be embellished from time to time. You discover new keys, other notations that authors insist on in changing fads and idiosyncrasies of language. I can warrant that my profiles recover on a current basis about 90% of what I have read or would want to read. God help me if I lose my notes on the rest!

Then, how keep up with what's closest to my immediate specialty? Acquiring a couple or three papers a day is not hard. And even with a fairly detailed critical examination down to checking the points on the graphs and so on, reading them as they come in is entirely doable. My problem is the arithmetic of accumulation. After a decade, I've got about 10,000 papers that I have got to keep track of: the texts and my marginal notes and so on. And here my system is absolutely broken down! A technological fix is on the way: document scanners that can store page images and digitize scripts on searchable media. One or a few CD-ROM's will take care of the storage. But what a lot of bother for information, yes full text, that I should be able to acquire electronically in the first place. The more so for specialty journals and references to be searched on demand.

Within a given specialty there are usually one or two journals that specialists must see. There may be only a couple of hundred people who have such a level of interest that they will look at every article. There are the journals of broad appeal, and then a very flat distribution of the other sources. For my part, an additional 30 articles a month -- perhaps half of them come from about fifteen journals; you can probably extrapolate with Bradford's Law to the rest. 90% will come from about thirty-five, and then there is a gradual asymptote out to the vanishing returns from the total coverage that the system is going to offer. Every now and then an article does pop up from an obscure place whence you had no systematic way to recover it; but in retrospect it was really quite important.

So each of us faces the task of selective retrieval from a cosmic domain of stuff that

every other eager beaver in the world has been busily putting into the repository. Our present technology enables an approximation with reasonable confidence. Keeping track of what you have accumulated on pieces of paper is the frustration. That's not your bedside reading, well served by the print on paper version. The next step, to integrate that into your own private library of useful knowledge, is simply not achievable with last year's technology.

The fact is that scientific literature inherently has grown beyond the scope of any hundred people to have understood it and gone into it in some depth; it is built in to the growth of knowledge that past improvements in communication and storage aren't going to alter. What are the consequences? For one thing, the problematics of assessing the literature reinforces all the other drives to specialization. The ambitions of scientists have changed, to focus on ever narrower targets. It's just too much hard work to master an interdisciplinary area on top of all the other institutional obstacles. Never mind the intellectual conceptual problems. Never mind the problem of getting funding or moral and fiscal support, just to get hold of the necessary expertise and information! But that impediment is in principle remediable. At the same time, are we drowning in information, inundated by the numbers of journals? You know, when you come to any specific issue, when there is some important special fact that you would like to know all about, the shoe is very often on the other foot. My usual experience in asking a new question: the odds are that the exquisite detail needed to take the next step has just never been done. So here, far from being drowned, I have a great deficit of specific and detailed knowledge of exactly what happens in such and such a system with such and such reagents, and so forth. Our systems for acquisition of that kind of material is not perfect, but it is getting a lot better; and with devices like key word searching, like related articles, like full abstract searching, which is just about what the technology does offer today, I can feel reasonably confident that explicit matters of factual detail, whether somebody has done that particular experiment can be retrieved, but often only with a lot of effort. Much more difficult, has anybody else had a good idea that would be pertinent to my search? Those keys are so much more difficult to catalogue. Often it takes great creative act to recognize that a concept developed in one context really is pertinent to another. So there will never be a guarantee that those can all be acquired. But there is at least the hope of finding it in that literature, and it is very important hope to try to preserve.

There are different adaptations to that flood, and more and more we do see, what I can only describe as a scandal, that scientific literature is not always taken seriously any more. In polls of scientists many will say that the primary source of their information about scientific work in their field is not with the published literature. It's by word of mouth, it's by telephone networks, by attendance at meetings, and so on. People have got to do what they've got to do. But I find those kinds of sources so unreliable! I feel very uncomfortable when the only place that I have heard something is by word of mouth. If I can't pin it down, if I can't hold its source accountable by saying that was in a published item, I can't look at it in detail, ruminate about it, think through what second order reactions I would have, I don't know whether my colleagues share that. They may feel that they don't have any alternative except to pick up what's on the rumor mill, but I think great mistakes can be promulgated in that fashion.

The telephone is a wonderful instrument, but when I try to use that to get information, people who have what I am looking for are all pretty busy. I hate to impose on their time,

and if I do, there is usually a round of telephone tag of three or four attempts to catch somebody before you actually do and try to get hold of them for the information. If it's a reference, I am delighted. If it's an attribution, it cannot be pinned down more definitely than you know, this is what I think. I don't feel like I have made a great advance over what I have had before. Not taking literature seriously reinforces the trend that libraries in desperation are cancelling subscriptions on journals that they don't see being very much read locally. And it doesn't make any sense to have a local copy of a serial where perhaps one in a hundred titles will ever be examined by anybody in that institution. For some of these journals, de facto, are approaching the point where they might as well only print one copy, send it to the National Library of Medicine or some other repository, and let it redistribute reprints by interlibrary loan. The economics obviously is insupportable. The fundamental problem is trying to foist an inappropriate level of number of vehicles on an outmoded mechanism for the purpose of dissemination. So that would fall of its own weight. You can see what I'm leading to: go from 1000 to 1 to zero print copies.

Meantime, the libraries are in a great dilemma trying to figure out exactly what to do. They get a fight from the faculty -- what a librarian hears when they want to drop a journal, you would think that every professor was reading every issue of every journal in the library! For the operational procedures by which libraries can make sensible decisions about acquisition priorities, they could get any number of technical aids on that point*, but it does put them in a very tough spot. Besides the budgetary crunch, the libraries are also running out of space. The older stuff is deteriorating anyhow! Maybe ink on paper was not a totally bad idea, for that reason alone, provided one clean copy remains available; unfortunately things don't always work out that way.

One direction things could take if we don't reform the system is that invisible colleges will take over as the principal but unreliable routes of communication. Archival copies of material will eventually be sent in to the repository, but there will be a limbo of material that doesn't know if it is going to go to hell or heaven for four or five years, while it is still cooking and unaccountably available, on a basis far from equitable. So, in due course there has to be a wholehearted exploitation of the new technologies and I don't have to plead for it. It's happening because electronic networks are becoming more and more available de facto to people working in a variety of fields. A couple of dozen of them now operate with a routine exchange of preprints (sic). The central problem facing the journal has been a radical change in the economics and technology of printing, without an adequate recognition of the essential value-added in the journal process. From Gutenberg's time until mechanized and computerized composition, that was providing the capital and the entrepreneurship and the organization to facilitate a process whereby an expensive and precious printed article was the product. It was characterized by rather high capital investment in the initial composition of any material. Once it was composed, there was a rather low variable cost for further dissemination. We had a market system for determining what was worthy of that degree of capital investment. Well today the capital investment on the printing technology is almost zero. The important value-added is the editorial process including issues of selection, then of editorial work and improvement. And that very precious imprimatur. When something comes out in a journal of high repute (to make a circular argument), that's a journal worth my time and worth my attention. If it is just thrown up in the air without having undergone that kind of editorial review, it will not have been refined in terms of both the presentation, and perhaps

even substance of the argument, and it won't have the imprimatur of other people, whose judgment I trust, that it's worth reading and can be relied upon for accountability. Whether the article then gets into *print* is almost an irrelevancy at this point. Any of a variety of media of communication could follow on that editorial process. What we need to see more than has happened so far is the marriage -- of that editorial role on the one hand, with a production role that uses the electronic technologies rather than the print, on the other. And that's where the spontaneous bulletin boards don't quite make it. They quickly get filled up with obscenities, literal and otherwise for lack of that sort of control. I don't mind the obscenities as long as I don't have to plow through them, but I'd like a truth-in-advertising framework that tells me, as I say, what's worth reading. I'd like to know that x,y or z editorial committee had been established as a guide for what is worth capturing the priority of my attention.

I think it will be the societies that provide the most likely framework for the organization of those functions. It won't make any money to start with. But that the economics and the technology will converge with the social necessities for this kind of improvement. Technically, we don't need much more than what we now have. There are a few problems about transmitting graphics and formatting manuscripts. Some standards have to be established, and some minor fixes especially on the graphics. But we are basically right there. Machines with gigabyte storage and ever smaller 25-meg processors are very routine today. You will find them by the hundreds in the laboratories and the libraries and so on, with a doubling of capabilities per unit costs every couple of years. So in ten years today's "super-computer" will be available certainly in every institution, and to a large degree in every laboratory. Communication links won't grow quite as fast as that, but if you consider the bandwidth of a package of CD ROMs, you have a variety of technologies for all the communication we need. So those are not limiting factors either. They are not very expensive. The machinery, the social framework, the decisions involved, the wetware, the distribution channels, the marketing and so on, really are all that stand in the way. There are not the same kinds of profit incentives that drive paper publishing; so I think the not-for-profit institutions will start taking over. Perhaps the for-profit publishing houses will provide the essential technical services because they can have the economy of scale, the organization, the hardware and so on, and then contract that out to the societies for providing the other elements of the equation.

That partnership could be a very productive one for the entire scientific community. One feature of that kind of a system to which we have only a crude approximation today, is feedback, dialectic. It shouldn't take a federal case for reactions to a paper to be elicited from the scientific community and not just on the rumor network, but some place where everybody else can see it. This is the bulletin board system of commentary and would complement what the fixed board of editors would have to say. If there is a good dialectical system and the critical community has an opportunity to express its views, even *ex-post facto*: that's how the scientific process works at its best. Here the economics and the technology for dialectic give a great edge on the electronic systems over the printed ones, if for no other reason than how to get propinquity? I mean, if an article has been printed and then a little later on, I write a critical reaction to it, (even in the rare case that the journal accepts that sort of commentary and further dialogue,) they do not adjoin one another on the shelves. It's a nuisance trying to find them. Let's say I write something six months ago; Gene Garfield wrote a blistering

critique sometime after that. How are the two of them going to be brought together? That kind of reshuffling of the units is very hard with printed paper. It's trivial, of course, to do it with electronic media via the networks of linkage of material and commentary. That potential for reaggregation stands just after mechanized search and tempo of availability as the greatest advantage that these new kinds of media can offer.

Let me make one further comment about global access: something very dear to Ken's heart, and to my own. There was a remark in my letter of invitation: "You may feel like you are in a flood, but people in the Third World are in a real drought. They never get the journals that you complain of getting too many of". And so forth. The economics of sharing will shift dramatically with these media. For trivial marginal costs you can provide 100 CDs a year which would far exceed the total volume of publication that they could ever hope to get in any other way. There is no other way in the world that we can reduplicate all the paper libraries that are we now have as a privileged treasure.

Another feature about globality that electronic systems will offer are built-in translation aids. I am not talking about the Nirvana of automated perfect smooth translation. Most of us here have a smattering of two or three foreign languages; a few of you are great linguists. But when I am reading an article in German, which I am fairly fluent in, wouldn't I love to have a built-in dictionary to help out when I run into a phrase that I didn't understand? I'll take the risks of that crude translation. It may come out with some of the ridiculous puns that you all know about. Again this becomes trivially easy in terms of its marginal cost, and will greatly extend the global accessibility of literature to a wide variety of people whose command of the current international standard English may not be perfect. So these are some of the arguments for the reforms that I hope you share with me and I would like to see brought about. Thank you very much.

* For The Rockefeller University library, I used the Science Citation Index to prepare an index to the frequency with which different journals were cited in their published papers. That would be a bad algorithm for acquisition of books and review journals.