CAN NOBEL PRIZE WINNERS BE PREDICTED?

Eugene Garfield and Morton V. Malin

Paper presented at 135th Annual Meeting,
American Association for the Advancement of Science,
Dallas, Texas – December 26-31, 1968

In a recent issue of the *New Scientist*, the editor asked, "What are Nobel Prizes for?"[1] Scientists and laymen alike have a right to ask this question. In 1967, Harriet Zuckerman convincingly reported in the *Scientific American*[2] how the Nobel Prizes have become a social force which shapes the lives of many individuals and organizations and exerts a strong influence on the future development of science.

Aside from the fact that the prizes do not adequately cover the present breadth of fundamental, and certainly not the applied sciences, there appears to be little relevance to some direct benefit to mankind as, e.g. a vaccine or a useable artificial organ. Furthermore, the relatively affluent team of research scientists has replaced the lone, often poverty-stricken scientist so common fifty years ago. The scientist today has been accurately described as an entrepreneur.[3] And, the publication of The Double Helix[4] will prevent the most naïve from thinking all science is for purely altruistic motives.

For these and other reasons, therefore, one of the most invidious characteristics of the Nobel Prize is not that so many prizes are obsolete so to speak, that is, for work done thirty years after the original work on which the prize is based; but rather that the Nobel Prize winners themselves are a significant factor in compiling the list of candidates.

The fact remains that selection of Nobelists is highly subjective.

The problem of selecting prize winners is similar to the problem of allocating limited resources for support of science – who shall get the support, who shall get the rewards?

It is not my purpose here to introduce a system to replace any that exists now for selecting prize winners. My main purpose is to report some interesting data. I do suggest that, if Nobel and other prizes are worth giving, then the sponsors of those prizes should also support some research which might lead to more objective methods – unless we prefer that awards for intellectual achievement, like the spoils system, be based on political considerations.

There is a common method of measuring scientific output, that is known as publications counting. I used this method several years ago to study[5] the output of chemists who synthesize new chemical compounds. The number of papers a man publishes does indeed provide some measure of his or her activity. It may be scientific or political or administrative. Regardless, it is a measure of activity. However prolific a scientist may be in publishing papers or as regular as he or she may be in attending meetings, there is no guarantee that the messages communicated through scholarly papers have had an ***impact*** on the scientific community.

Impact can be felt in many ways. Few will question the importance of a good teacher. Sir Hans Krebs[6] has recently shown that excellence breeds excellence in science. Tracing the careers of student descendents of great teachers is one method of observing impact.

At the Institute for Scientific Information, we have found that impact is a sociometric phenomenon. Impact appears to be measurable or identifiable, both quantitatively or qualitatively. Impact can be determined by utilizing information inherent in bibliographic citations – the linear symbols which tell us about the interconnecting links of scientific events. Through citation networks[7] it is possible to observe certain historical and sociological processes at work which are, in fact, the implicit, if not the explicit, concern of Nobel Prize Committees.

Let me turn sharply from this generalized theme to the specific point of this short paper. Can Nobel Prize winners be predicted?

The question has, of course, been posed somewhat facetiously. What does one mean by "predict"? There are surely over one million trained scientists in the world. About half of them are cited each year in the *Science Citation Index*. Obviously some are cited more than others.

But who are the cream of the crop? Who are the top one percent? There are about 10,000 scientists, selected by a purely mechanical process, whose published papers or books are cited fifty or more times per year. (The job of reducing 1,000,000 names to 10,000 is, incidentally, no trivial accomplishment.)

It is undeniably true that there is a given probability that some important and deserving names will not be selected for a variety of reasons. However, one must consider that almost all Nobel Prize winners have **regularly** appeared on this list and that, in recent years, more Nobelists appear on the list. Moreover, as one raises the stakes – that is, narrows the list to the top 2,500 or less than one-half percent of those cited – over 90% of the Nobelists are still anticipated. In fact, over 75% are found in the super elite group of the 1,000 most frequently cited scientists – less than 1 out of every 1,000 scientists in the world.

To predict is to talk about probabilities. Any bookmaker or Wall Street broker or insurance actuary could construct a table of odds that would prove reasonably accurate; and the accuracy would improve with each passing year – based on *SCI* citation data. If he then combined these citation data with a reasonable amount of analysis of past performance and other characteristics, such as peer judgments, the odds can be significantly improved.

Factors to be considered include the length of time which has elapsed since the individual scientist's most frequently cited paper, the field in which he works, and the type of work involved. For example, one finds among the most frequently cited (in addition to Nobelists) those scientists who have developed widely used methodologies. If O.H. Lowry has not received a Nobel award, he must deserve some special recognition for having published the most cited paper in the scientific literature.[8] I am not qualified to judge his work; but the Lowry method of protein determination must be hard to improve upon – else why is it so universally used? Surely any ambitious biochemist should consider the economic and scientific importance of improving upon such a widely used

method. Whether this is of greater or lesser scientific significance than the radio carbon dating methodology of W.F. Libby[9] is also a question I am not qualified to answer. He has received the prize. His book was his most-cited work.

It so happens that Lowry, a man I have never met, unlike other authors of methodology papers, also appears in the top 1% of cited authors even if one discounts his classic 1951 paper. This is not always the case, for authors of highly cited papers. Incidentally, Lowry's paper was cited no less than 3,000 times in 1967. The second most cited paper, by E.S. Reynolds,[10] was "only" cited 1,000 times. Subsequently it was cited over 22,570 times.

Another variable the bookmakers would have to consider in using citation counts is the population of papers in a given field. Biochemists and biologists are much more prolific than physicists. The current physics literature is, at present, less than 40,000 papers per year. Bio-scientists are not only more numerous than physicists; but their work is closely associated with the medical and agricultural literature, which is quite large as compared to physics, nuclear science, and engineering. When we compiled the *Genetics Citation Index* several years ago[11] we found that the proportion of citations for the average geneticist was much higher than for the average scientist. It is, therefore, incredible to ponder the significance of a physicist like Gellman, whose work was cited about 950 times last year – an extremely large number for a physicist.

Having determined that Nobel Prize winners almost invariably fall into the top one percent of highly cited scientists, it is important to find out why. One reason is that Nobelists publish quite frequently. Only the rare Nobel Prize winner has published only a few papers. In addition, Nobelists are not only prolific communicators; what they publish is eventually – and quite consistently – cited over long periods of time both **before and after** the Prize. Further research may reveal that the Nobel Prize itself prompts more research in the fields represented by the awards. Awards may further increase the impact factors. This may have been the case for the Russian physicist, Lev Landau. Zuckerman claims that the Nobel Prize did reduce the amount of research published by Nobelists after the awards due to the social and other pressures that result. However, others apparently are stimulated to expand their work.

It is also characteristic of Nobelists that one or more of their papers will also be found on a separate list of the most frequently cited papers. This is not insignificant as these papers are, in fact, important crossroads on the historical map of science (what I have described as the historical network). One can draw those maps using citation data[12] and computer methods of displaying the information[13]. Using these methods, one can more readily observe the development of scientific fields. Since that time see www.histcite.org

Unlike certain "flash-in-the-pan" papers which crop up each year, Nobel Prize winning work is increasingly cited as the years pass. Unfortunately we do not yet have *SCI* data for the 1950s or for earlier decades – a situation which makes it difficult to observe these phenomena for the Nobelists who did their important work twenty to thirty years ago. [Since this paper was presented in 1968, ISI has gone on to produce the *SCI* from 1900 to 1964 – the year *SCI* was launched.]

However, the data on M.W. Nirenberg, a 1968 winner, can illustrate my point. His 1961 paper in *the Proc. Nat. Acad. Sci. U.S.* [14] was cited over 100 times in each year studied, 1964 through 1968.

His more recent paper published in *Science*[15] in 1964 was cited 55 times in 1965, 70 times in 1966, and 110 times in 1967.

Similarly, Murray Gellman, mentioned previously, the physicist who has not yet received a Nobel Prize and who is a "shoo-in", characteristically has published several highly cited papers. His 1960 paper in *Nuovo Cimento* is increasingly cited each year.[16] His 1962 paper in the *Physical Review*[17] was cited over 150 times on average from 1964 to 1968, and his more recent 1964 paper in *Physics Letters*[18] has been cited 125 times in each of the past three years.

Unless the **implicit** ground rules for the selection of Nobel Prize winners change, it is inconceivable that any scientist could achieve the award unless he has published significant works that have made a wide impact both in the so-called invisible colleges and the scientific community generally. He will have published one or more really landmark papers which stimulate many others to conduct experiments or investigations which are related to or are derivatives from these landmark papers. Whether they like him or not, these people will cite his work (contrary to common opinion) and they will cite him frequently. In most cases, personal contact will prevent omission of pertinent references in most papers even if the refereeing process fails in other cases. In spite of the known fact that friends may cite friends and that some language barriers still exist to slow down communication, the really "breakthrough" paper does, in fact, usually break through.

To be a Nobel Prize winner, one must not only be able to think and to perform good science, one must also be able to sell one's ideas; that is, to communicate. To cite the case of Gregor Mendel as a reason for ignoring the method implied here is absurd for several reasons. Scientific communication and research itself has undergone great changes since the time of Mendel.[19] But, even if there are a few "Mendels" around who are being ignored, it is probably their own fault. In any case, there are only a few hundred people who can with the Nobel Prize in the next few decades; but there are thousands who qualify not only on the basis of citation counts, but also because there is a great deal more good work in process. Only the worst possible communicator is condemned to oblivion. In science, as in business, persistence as well as originality is required.

Since it is probably an academic exercise to speculate on who will win the Nobel Prize per se, it is important to consider how we can make use of the data available through citation counting. While most individuals do not aspire to a Nobel Prize, they can aspire to join that exclusive club called the National Academy of Sciences. During the past five years, approximately 200 new members have been elected. Of this number, about two-thirds appear on ISI's list of the 10,000 most frequently cited scientists. Of the remaining third, most are administrators and others who, just prior to or at the time of their election, had not been publishing in scientific journals or conducting research as e.g. the presidents of certain universities. A few were economists who published in journals not presently covered in our studies. Furthermore, it might also be said of this one-third that their elections had been unnecessarily delayed as
is often the case with Nobel Prize winners. Like a small percentage of Nobelists, these Academy members published works which are not in current vogue.

Significantly, of the foreign associates elected to the NAS, over three-fourths appear on ISI's list. At this point, it is worth mentioning an artifact that could certainly affect the interpretation of our findings. At worst it can only increase the size of the list of 5,000 by ten to twenty percent; however most significantly, it might later change the relative rankings of those on the list.

The *Science Citation Index* was designed primarily for information retrieval purposes. Given a specific published paper, one can find all subsequent papers that cite it by simply knowing the first author and year of the paper in question. If one is interested in genetic regulatory mechanisms in the synthesis of proteins, it is only necessary to know that Jacob published this paper in 1961.[20] In the *SCI*, hundreds of papers on this subject can be found by this simple piece of information. Most biochemists realize, of course, that the co-author of this paper was J. Monod, who also happens to be the first author of another highly cited paper "on the nature of allosteric transitions".[21] That 1965 paper was cited over 200 times in 1966 alone! It is still highly cited. Of course, both Jacob and Monod shared a Nobel Prize in 1965.

Nevertheless, unlike this case, it is well known to that many scientists, especially after they have achieved a degree of eminence whether through a departmental chairmanship or some other award, do not list themselves as the first author. This merely creates a logistic problem, but in no way affects the basic validity of the method discussed here. To determine the total citation count an impact for any given individual, one must obtain his or her bibliography and then one can easily determine how often each paper has been cited.

In spite of this artifact, we have found that the majority of eminent scientists do eventually list themselves as first authors enough times to insure that they appear in the top one percent. We hope in the near future to complete a study which takes into account multiple authorship to determine which scientists have not been accounted for when using the first author method.

It should be obvious that this problem does not affect our list of most-cited papers. We intend to publish this list of super-classics with a complete co-author index and subject index. This list will not only be of great value to science historians and sociologists – it will also be the grist for a dictionary of primordial citations[22] which will enable students and librarians to find key citations for the most frequently cited works. Of the 2 million different works that are cited each year in the *Science Citation Index*, about one percent or 20,000 will be selected. Each year there will be a list of the most-cited papers, giving their relative ranks. These include papers that are cited six or more times each year.

We can, of course, eliminate self-citations if necessary. By studying the changing ranks of these papers each quarter and year, it will be possible to identify interesting pockets of scientific activity. Contrary to general belief, citation data can reveal such activity very quickly. Although the papers that are most cited in each yearly *SCI* were published two years previously, fields of rapid activity do show up in the quarterly *SCI* cumulations. Publication lags do not appear to affect this process.

The members of the invisible colleges can and do cite hot papers which they have seen in press or in pre-print form. By the time they are publishing their own work, the cited work is finally published and the inherent time lag in both papers has been erased. Furthermore, the advent of quickie journals has further reduced time lags.

The ability to identify the select group of innovative individuals who will have made a major impact on science has important social and economic consequences.

Administrators in academic, industrial, and government organizations have frequently indicated the need for a research management tool – a prediction model – for identifying this group.

The allocation of increasingly scarce intellectual and financial resources for supporting research could be managed more efficiently with such a prediction tool. The ability to spot creative people much earlier in their careers than at present will permit a more equitable distribution of the limited resources available. Therefore, research on methods of predicting or measuring creativity is not just an academic exercise.

Methods currently used by science administrators and others for identifying "creative" individuals essentially consist of obtaining the subjective opinions of peers. Prizes, grants, fellowships, or other forms of recognition are usually based upon a qualitative review by a committee or cohort. In the case of the Nobel Prize, these committees consist of eminent scientists, including previous Nobelists.

What these subjective evaluations are attempting to establish is a measure of the impact of various candidates in their fields. In other words, the role of an award committee is to measure the role and significance of individuals in the history of science.

The procedures presently available are not only slow and costly, but necessarily tedious. They have become more so because of the exponential growth of scientific research and the increasing number of scientists. This leads to the use of such "objective" methods of counting the number of papers published (Publish or Perish) because truly objective methods have not been developed.

Using the *Science Citation Index*, as I have described, it has been possible to rank the top 1 or 2% of scientists who should always be considered for Nobel and other prizes. The data we have collected at ISI covers over five years of the *Science Citation Index* and includes over eleven million reference citations. Almost without exception, Nobel Prize winners have total citation counts approximately fifty times higher than the average scientist. Further analysis showed that unusually significant creative work, typified by Nobel Prize winners, is also associated with a continuity factor. Not only do they have unusually high impact factors, but the impact is observable over a long, continuous period. In addition, prize winners usually have published one or more key papers which have extraordinarily high and continuous impact.

In spite of small errors that may be introduced by known citation practices and by artifacts of the *SCI* itself, the file is now so large that the results are valid within the limits of acceptable statistical error.

The use of citation data appears to isolate the creative, innovative, or above-average group of individuals from the massive total population of scientists. Significantly, the procedure involved is essentially **automatic** and objective. If the method is used judiciously in **combination** with peer judgments, the process of making awards or giving other forms of recognition becomes more manageable and equitable. The prediction model can be used, within certain limits of confidence, by various committees and agencies to identify potential candidates from a total so-called "cohort" population.

REFERENCES:

1. Lost Reference

2. Zuckerman H. Sociology of Nobel Prizes, *Scientific American* 217 (5): 25& 1967.

3. Garfield E. and Malin M. The Entrepreneur as a Doctoral Candidate. Presentation at American Chemical Society National Meeting. Philadelphia, PA. April 6-11, 1975. http://www.garfield.library.upenn.edu/paper/entrepreneurdoctoralcandidate1975.pdf

4. Watson JD. *The Double Helix : A Personal Account of the Discovery of the Structure of DNA.* Touchstone (Simon & Schuster) NY. Copyright 1996.

5. Garfield, E., Revesz, G. S., and Batzig, J. H. "The Synthetic Chemical Literature from 1960 to 1969 ," *Nature*, 242 (5396) p.307-309, 1973.
Reprinted in:*Current Contents*, May 9, 1973; EIS V.1, P.443
http://www.garfield.library.upenn.edu/essays/V1p443y1962-73.pdf

6. Krebs HA. Making of a Scientist, *Nature* 215(5109): 1441-1445

7. Garfield, E., I. H. Sher, and R. J. Torpie. "The Use of Citation Data in Writing the History of Science." Published by The Institute for Scientific Information, December 1964.
Report of research for Air Force Office of Scientific Research under contract F49(638)-1256. .
http://www.garfield.library.upenn.edu/papers/useofcitdatawritinghistofsci.pdf

8. Lowry OH, Rosebrough NJ, Farr AL & Randall RJ. Protein measurement with the Folin phenol reagent. *J. Biol. Chemistry*, 193: 265-75, 1951.

9. Libby WF. *Radiocarbon Dating*. 1955. 2nd edition. University of Chicago Press.

10. Reynolds ES. Use of Lead Citrate at high PH as an electron-opaque stain in electron microscopy. *Journal of Cell Biology*, 17(1): 208& , 1963.

11. Garfield E. Sher I. Genetics Citation Index - Experimental Citation Indexes to Genetics with Special Emphasis on Human Genetics. Reprinted in *Essays of an Information Scientist*, Vol:7, p.515-522, 1984.
http://www.garfield.library.upenn.edu/essays/v7p515y1984.pdf

12. Garfield, E., Revesz, G. S., and Batzig, J. H. "The Synthetic Chemical Literature from 1960 to 1969 ," *Nature*, 242 (5396) p.307-309, 1973.
Reprinted in:*Current Contents*, May 9, 1973; EIS V.1, P.443
http://www.garfield.library.upenn.edu/essays/V1p443y1962-73.pdf

13. Garfield, E., I. H. Sher, and R. J. Torpie. "The Use of Citation Data in Writing the History of Science." Published by The Institute for Scientific Information, December 1964.
Report of research for Air Force Office of Scientific Research under contract F49(638)-1256. .
http://www.garfield.library.upenn.edu/papers/useofcitdatawritinghistofsci.pdf
Also see : http://www.histcite.com/ and www.garfield.library.upenn.edu/histcite

14. Matthaei JH and Nirenberg MW. Characateristics and stabilization of DNAASE- sensitive protein synthesis in e-coli extracts. *Proceedings of the National Academy of Sciences of the United States of America*. 47(10): 1580& 1961.

15. Nirenberg MW. And Leder P. RNA codewords + protein synthesis – effect of trinucleotides upon binding of SRNA to ribosomes. *Science* 145(363): 1399& 1964.

16. Gellman M. and Levy M. The axial vector current in beta-decay. *Nuovo Cimento* 16(4): 705-726, 1960.

17. Gellman M. Symmetries of Baryons and Mesons. *Physical Review* 125(3):1067& 1962.

18. Gellman M. A schematic model of Baryons and Mesons. *Physics Letters* 8(3):214-215,1964.

19. Garfield E. Would Mendel's work have been ignored if the Science Citation Index was available 100 years ago? *Essays of an Information Scientist*, Vol:1, p.69-70, 1962-73. Reprinted from *Current Contents*, #2, January 14, 1970. Full text available at :
http://www.garfield.library.upenn.edu/essays/V1p069y1962-73.pdf

20. Jacob F. and Monod J. Genetic regulatory mechanisms in synthesis of proteins, *Journal of Molecular Biology* 3(3): 318& 1961.

21. Monod J. Wyman J. Changeux JP. On nature of allosteric transitions – a plausible model. *Journal of Molecular Biology* 12(1):88 &, 1965

22. Garfield E. Citation indexing, historio-biography and the sociology of science. *Proceedings of the third international congress of medical librarianship* (Amsterdam: Excerpta Medica, p.187-204, 1970). Reprinted in *Essays of an Information Scientist*, Vol:1, p.158-174, 1962-73 and *Current Contents*, #6, April 14, 1971
http://www.garfield.library.upenn.edu/essays/V1p158y1962-73.pdf