Essays of an Information Scientist, Vol:4, p.660-665, 1979-80 Current Contents, #44, p.5-10, November 3, 1980

Current Comments

Multiple Independent Discovery & Creativity in Science

Number 44 November 3, 1980

Many scientists have told me how disappointing it is to have their research anticipated by other independent investigators. To have your ideas preempted in print just as you are about to publish is especially frustrating. A few studies suggest that anticipated research is an occupational hazard for the *majority* of active researchers. Indeed, some investigators report being anticipated several times in their careers.

Building on general trends of anticipation within various fields he first observed in 1965,[1] University of Wisconsin sociologist Warren Hagstrom surveyed 1,718 US academic research scientists in 1974.[2] Of these, 46.2% believed they were anticipated once or twice, and 16.4% more thought they were anticipated three or more times. In 1971, Jerry Gaston, Southern Illinois University sociologist, surveyed 203 British high energy physicists.[3] He found that 38% believed their results had been anticipated once, and another 26% more than once. These figures agree rather well with the simple theoretical models of multiple discovery that will be discussed in this essay.

Anticipation or preemption of research results is a subset of a wider phenomenon referred to by historians and sociologists as "multiple independent discovery." Robert K. Merton, Columbia University sociologist of science, defines what he calls "multiples" as instances in which similar discoveries are made by scientists working independently of each other.[4] Merton says, "Sometimes the discoveries are simultaneous or almost so; sometimes a scientist will make a new discovery which, unknown to him, somebody else has made years before."[4] One of the most commonly cited examples is the independent formulation of the calculus by Newton and Leibniz, which has been definitively described by A. Rupert Hall.[5] Another is the theory of the evolution of the species, independently advanced by Charles Darwin and Alfred Russel Wallace.

But multiple independent discovery is not limited to only a few historical instances involving the giants of scientific research. On the contrary, Merton believes that multiple discoveries, rather than unique ones, represent the *common* pattern in science.[6]

Interest in the widespread phenomenon of multiples in science has rekindled a philosophical debate on the process of discovery and creativity in science. One side of the debate has its roots in the "great man" or "genius" theory of history, proposed in the 19th century.[7] Traditionally opposed to this theory is the "social determinist" or "*zeitgeist*" argument.[8] More recently, a "chance" theory based on probabilistic models has tried to account for the occurrence of multiples.[9]

Jacob Schmookler, University of Minnesota economist, says the great man theory explained discovery and invention primarily in terms of the individuals who made them.[7] (p.189) Dean Simonton, social psychologist at the University of California at Davis, adds that the individual scientists and inventors were believed to possess extraordinary abilities, backgrounds, and personalities which allowed them, rather than their "ordinary" colleagues, to make important discoveries.[8] Simonton points out that the genius or great man theory of creativity can't adequately explain the *simultaneous* appearance of independent discoveries. However, he says it does explain one special case of multiples—"premature" discoveries that were "rediscovered" by independent investigators after a period of years. I recently discussed the delayed recognition of premature scientific discoveries in a separate essay.[10]

The social determinists acknowledge that mental ability plays an obvious role in discovery and invention, but they believe that cultural factors are far more important than individual personalities. The distinguished anthropologist A. L. Kroeber was an early proponent of the social determinist argument. In a 1917 article,[11] he reasoned that there are many people with high intelligence in a large population at any given time. Thus, if a particular inventor had died as an infant, there is a good chance that the invention would still have been conceived by someone else. The fact that certain ideas or inventions occur at the same time to different people proved, to Kroeber at least, that they "seem to have been destined to come about precisely when they did"[11] because of cultural factors.

A few years later in 1922, the sociologists William Ogburn and Dorothy Thomas, Columbia University, compiled a list of 148 independent discoveries and inventions made between 1420 and 1901.[12] They observed that the rate of such repeated innovations increased with time. They concluded that cultural evolution must account for the increased frequency of multiple independent invention because inherited mental ability could not have significantly increased in 500 years. In fact, Merton observes, Ogburn and Thomas believed "that the innovations became virtually inevitable as certain kinds of knowledge accumulated in the cultural heritage and as social developments directed the attention of investigators to particular problems."[6]

Simonton has labeled the social determinist argument the *zeitgeist* theory of creativity.[8] He says the social determinists claim that "the sociocultural system as a whole, embodied as the spirit of the times [*zeitgeist*], is ultimately responsible for any given technoscientific advance."[8] While it is a plausible explanation of simultaneous independent discoveries, Simonton says that premature discoveries "must prove an embarrassment to *zeitgeist* theory."[8]

When Merton redirected the attention of sociologists and scientists to the importance of multiple discoveries in 1952,[13] he recommended a moderate position between the opposing theories of scientific creativity. Instead of perpetuating "the false disjunction between the heroic theory centered on men of genius and the sociological theory centered on the social determination of scientific discovery,"[6] Merton proposed an alternative theory that combined elements of both. He acknowledges the central position of the accumulated knowledge base in culture, but denies that discovery and invention are *inevitable* as a result.[6] Merton points out that the main ingredients of some discoveries were present in culture for many years before they were finally made.

Also, he attributes a distinctive role to genius but denies that the "great man's" mentality gives him insights imperceptible to the ordinary man.[6] Merton redefined genius in sociological rather than psychological terms. The genius doesn't see something to which the ordinary man is blind—rather, he sees *more*, and more quickly. As a result, "The genius will have made many scientific discoveries altogether. . . . This means that each scientist of genius will have contributed the functional equivalent to the advancement of science of what a considerable number of other scientists will have contributed in the aggregate."[6]

The great man theory predicts that the genius conceives of things far in advance of "the crowd" by virtue of his unique capability, and his ideas therefore stand alone. Merton's expanded sociological theory of genius predicts that the genius will be involved in *more* multiple discoveries than the ordinary scientist because of a great number of discoveries altogether. Merton cites many cases of the "multiplicity of multiples" among the great scientific thinkers.[6] Interestingly, Simonton showed that more eminent scientists are "still more likely to participate in multiples even after controlling for individual differences in the number of notable contributions."[8]

Several authors point out that Merton's modified social determinist position, which denies both the inevitability of scientific discovery and a unique role for individual genius, is actually compatible with a "chance" model of multiple discovery.[7,8,14] Derek Price, Yale University historian of science, suggested a "ripe apple" model using a Poisson statistical distribution. He asks, "If there are 1000 apples in a tree, and 1000 blindfolded men reach up at random to pick an apple, what is the chance of a man's getting one to himself, or finding himself grasping as well the hand of another picker, or even more than one?"[14] The statistically expected distribution of multiples was then compared with an actual distribution among 264 multiples compiled by Merton and Elinor Barber in 1961.[6] As can be seen in Table 1, the distributions agree to some extent, especially between doublet, triplet, and quadruplet discoveries. However, Price cautions that the results shouldn't be given too much weight since they are based on only approximate parameters.

However, Price believes this admittedly very simple and approximate model makes interesting predictions that are near the truth. For example, he says, "One can calculate from Table 1 that about 37% of the apple pickers will make uncontested discoveries but the remaining 63% will be engaged in multiple discovery. It is worth noting that the distribution of the discoveries is different from that of the discoverer. Some 58% of the discoveries are unique and the remaining 42% are shared in the multiple process."[15]

Table 1: Poisson distribution and simultaneous discovery (see **Price D J D.** *Little science, big science.* New York: Columbia University Press, 1963. p. 67).

Number of simultaneous discoveries	Merton-Barber data cases	1000 apples and men cases
0	—	[368]*
1	—	368
2	179	184
3	51	61
4	17	15
5	6	3
6 or more	11	1

* i.e., there are 368 cases in which an apple is *not* picked. The remaining 632 apples are picked, singly or multiply, by the thousand pickers for an average of 1.6—that is, 1.6 discoverers/discovery.

Simonton also subjected the Merton-Barber list of 264 multiples to a statistical analysis using a Poisson distribution.[9] Instead of arbitrarily defining the number of "apples and pickers," Simonton set mean values ranging from 0.8 to 1.6 to see which would most closely agree with the Merton-Barber distribution. He found that a mean value of 1.4 agreed most closely with the Merton-Barber figures. With this mean, 41% of the total discoveries would be multiples, and 35% would be singletons.

Simonton applied the same methods to the Ogburn-Thomas list of 148 multiples. The optimal mean in this case turned out to be 1.2. This value is close enough to the Merton-Barber mean of 1.4 for Simonton to conclude that "the true [mean] is probably somewhat more than unity and definitely less than two. Hence, so far techno-scientific contributions hardly appear inevitable."[9] Simonton also reconstructed the *probable* frequencies of multiple discoveries from the Ogburn-Thomas list. He found that about 30% of the probable discoveries are "undiscoveries"—apples that were on the tree but were passed over by the pickers, to borrow Price's imagery. This supports Merton's statement that discoveries are not necessarily inevitable.[6]

When Simonton compared the probable frequency of multiple discovery to the actual number of cases reported by Ogburn and Thomas, he found that low grade multiples are less likely to be reported than high grade multiples. Low grade multiples involve only two or three independent researchers while high grade multiples involve more than three. Simonton says that the number of multiples involving six independent discoverers probably amounts to *ten* times the number of reported cases, while those involving two discoverers were underreported by a factor of 44!

This supports Merton's claim that the number of reported multiples is much lower than the number of actual multiples. He gives several reasons why multiples are underreported.[6] For example, the announcement of a discovery or invention forestalls simultaneous research efforts that could have produced similar results. Researchers may abandon their projects and not report their independent but identical results once they've been anticipated in print. Also, many singletons prove in later years to be rediscoveries of prior research that either wasn't published or wasn't recognized at the time.

Lastly, fewer scientists now engage in priority disputes when multiple discoveries are made: uncontested multiples are simply less often noticed. Merton cites the following figures to demonstrate this point: before 1700, 92% of the multiple discoveries were contested, in the 18th century 72%, 59% in the last half of the 19th century, and finally only 33% in the first half of the 20th century.[6]

In light of his statistical analysis, Simonton argues that the distribution of independent discoveries doesn't support the idea that discoveries are inevitable, as the old social determinists argued. Also, the genius theory is not supported by the results. While Simonton doesn't deny that social factors can influence the content and intensity of scientific creativity, he concludes that "chance seems to have usurped the dominating position in explicating the appearance of multiples in the history of science and technology."[9] This conclusion is compatible with Price's ripe apple model and Schmookler's simple probabilistic argument.[7]

In a later study, Simonton examined the *zeitgeist*, genius, and chance theories of multiple discovery in relation to four separate "critical tests" to determine which is most plausible.[8] The four tests were conducted on two samples of multiples, one listing 199 multiples and the other 579. Simonton found that "the position that best meets all critical tests is the chance theory. . . .Despite the apparent explanatory superiority of the chance theory, however, the other two theories are not completely overturned."[8] In effect, all three theories contribute to an explanation of multiple discoveries and scientific creativity.

Merton's expanded sociological model approaches a "unified theory" of multiple discovery by combining elements of both the *zeitgeist* and genius perspectives, and it is compatible with the chance model. If we affirm only the chance model, we have to consider multiples as mere *coincidences* whose occurrence is uniform across all scientific disciplines. But Merton believes these "coincidental" multiples are not trivial scientific phenomena—rather, they are significant research areas for sociologists and historians of science.

As Merton correctly points out, "The sheer fact that multiple discoveries are made by scientists working independently of one another testifies to the . . . fact that, though remote in space, they are responding to much the same social and intellectual forces that impinge upon them all."[4] It would be worthwhile to uncover those common social and intellectual forces, particularly at a time when so much stress is put on the specialization of research fields. By studying the occurrence and frequency of multiple discoveries, perhaps we'll be able to keep sight of the forest of science through the increasingly narrow specialty trees.

One of the earliest projects we undertook by use of the *Science Citation Index® (SCI®)* was the detection of unwitting duplications of research. I have reported many times on the examples we discovered that might have been uncovered through bibliographic coupling, that is, finding papers which cited one or more papers in common. It would be a massive computer effort but it is an experiment we ought to try one day. If two independent authors cite four or more authors in common it is not unusual, but such coupling may indicate they are working on the same idea. If editors had immediate access to such files how many inadvertent published duplications might be avoided? Perhaps the increased use of word processors and microcomputers to prepare manuscripts will mean that authors' bibliographies can be matched against our master files, not only to insure bibliographic accuracy, but also to identify degrees of bibliographic coupling or even earlier premature discoveries.

ISI®'s co-citation analysis of scientific literature is another way multiple independent discoveries may be identified. Co-citation is the number of times two earlier published papers are cited together in the more recent literature. As I pointed out in my essay on opiate receptor research, co-citation analysis revealed the multiple discovery of opiate receptors in rat brain.[16] Another study based on ISI's data base using the co-citation model identified the simultaneous discovery of a precursor molecule for collagen called "procollagen."[17] Still another co-citation study of ISI's data uncovered the joint discovery of reverse transcription.[18] I've discussed the details of ISI's co-citation analysis procedures in a separate essay, which also explained the difference between co-citation and bibliographic coupling.[19]

In the area of technology, the US Patent Office (and other patent offices around the world) is a potentially vast reservoir of information on duplicate and multiple discovery. One would expect that patents show the same sort of Poisson distribution of multiple discoveries as we have discussed for science. Ogburn and Thomas observed in 1922 that almost half of the patent applications in the US were denied.[12] Many of these may have been denied because they duplicated already granted patents. However, Schmookler points out that patents may be denied because only one of myriad "claims" are already protected, and identical claims are practically non-existent.[7] (p. 191-2)

But this question is difficult to decide because some claims are denied on grounds that aren't revealed once the patent is approved and printed. However, one can examine the "file wrappers" in the US Patent Office library in Washington. I often did this when I first experimented with Patent Citation Indexes,[20] based on the idea of attorney Arthur Seidel.[21] I'll have more to say about this when the subject of patents is reviewed in a future essay.

* * * *

My thanks to Alfred Welljams-Dorof for his help in the preparation of this essay.

REFERENCES

1. **Hagstrom W O.** *The scientific community.* New York: Basic Books, 1965. 304 p.
2. --------------------. Competition in science. *Amer. Sociol. Rev.* 39:1-18, 1974.
3. **Gaston J.** Secretiveness and competition for priority of discovery in physics. *Minerva* 9:472-92, 1971.
4. **Merton R K.** Resistance to the systematic study of multiple discoveries in science. *Eur. J. Sociol.* 4:237-82, 1963.
   (Reprinted in: **Merton R K.** *The sociology of science. Theoretical and empirical investigations.* Chicago: University of Chicago Press, 1973. p. 371-82.)
5. **Hall A R.** *Philosophers at war.* New York: Cambridge University Press, 1980. 338 p.
6. **Merton R K.** Singletons and multiples in scientific discovery: a chapter in the sociology of science. *Proc. Amer. Phil. Soc.* 105:470-86, 1961.
   (Reprinted in: **Merton R K.** *The sociology of science. Theoretical and empirical investigations.* Chicago: University of Chicago Press, 1973. p. 343-70.)
7. **Schmookler J.** *Invention and economic growth.* Cambridge, MA: Harvard University Press, 1966. 332 p.
8. **Simonton D K.** Multiple discovery and invention: zeitgeist, genius, or chance? *J. Personal. Soc. Psychol.* 37:1603-16, 1979.
9. --------------------. Independent discovery in science and technology: a closer look at the Poisson distribution. *Soc. Stud. Sci.* 8:521-32, 1978.
10. **Garfield E.** Delayed recognition or premature discovery—why? *Current Contents* (21):5-10, 26 May 1980.
11. **Kroeber A L.** The superorganic. *Amer. Anthropol.* 19:163-213, 1917.
12. **Ogburn W F & Thomas D.** Are inventions inevitable? A note on social evolution. *Polit. Sci. Quart.* 37:83-98, 1922.
13. **Merton R K.** Foreword. **Barber B.** *Science and the social order.* New York: Free Press, 1952. p. xi-xxiii.
    (Reprinted in: **Merton R K.** *The sociology of science. Theoretical and empirical investigations.* Chicago: University of Chicago Press, 1973. p. 210-20.)
14. **Price D J D.** *Little science, big science.* New York: Columbia University Press, 1963. p. 67.
15. ---------------. Telephone communication. 25 September 1980.
16. **Garfield E.** Controversies over opiate receptor research typify problems facing awards committees. *Current Contents* (20):5-19, 14 May 1979.
17. **Small H G.** A co-citation model of a scientific specialty: a longitudinal study of collagen research. *Soc. Stud. Sci.* 7:139-66, 1977.
18. **Griffith B C, Small H G, Stonehill J A & Dey S.** The structure of scientific literatures II: toward a macro- and microstructure for science. *Sci. Stud.* 4:339-65, 1974.
19. **Garfield E.** ABCs of cluster mapping. Part 1. Most active fields in the life sciences in 1978. *Current Contents* (40):5-12, 6 October 1980.
20. -------------. Breaking the subject index barrier; a citation index for chemical patents. *J. Pat. Off. Soc.* 39:583-95, 1957.
21. **Seidel A H.** Citation system for patent office. *J. Pat. Off. Soc.* 31:554, 1949.