ELSEVIER

Research Policy 24 (1995) 507–519

research policy

Inventive productivity

Francis Narin *, Anthony Breitzman

CHI Research, Inc., 10 White Horse Pike, Haddon Heights, NJ 08035, USA

Final version received December 1993

Abstract

The 1926 observation by Lotka that the number of highly productive scientists was a relatively small fraction of all scientists has been extended to patented technology. Specifically, Lotka observed that for every 100 scientists who produce one paper there are approximately $100/2^2$, or 25, who produce two papers, $100/3^2$, or 11, who produce three, etc., with only one scientist in the set who will produce ten papers. An investigation of the number of patents per inventor was carried out for four companies, two American and two Japanese, in the area of semiconductors. For all four cases a Lotka-like distribution was found, with a relatively small number of highly productive inventors with their names on ten or more patents, and a large number of inventors with their names on only one, and a general factor of ten difference in productivity between the most- and the least-productive inventors.

1. Introduction

In his book *Little Science, Big Science*, the late Derek de Solla Price reviews the early interest in measuring the distribution of quality or eminence among individuals (de Solla Price, 1963). Price starts with a discussion of the work of Francis Galton, who was concerned with estimating the rarity of outstanding men, particularly those in science. Galton used a number of informal literary criteria for measuring eminence, such as inclusion in biographical compilations or in selected columns of obituary notices. Later studies by others were based on other criteria of importance such as listings in noteworthy bibliographies. All of these early studies concluded that eminence is very highly concentrated within a small fraction of a population.

In 1926, this high concentration of productivity was crystallized for scientific bibliometrics by A.J. Lotka, of the Metropolitan Life Insurance Company, as an inverse square law of productivity (Lotka, 1926). In his landmark paper, Lotka states that

> ... it would be of interest to determine, if possible, the part which men of different calibre contribute to the progress of science.

Lotka used entries from the *Decennial Index of Chemical Abstracts, 1907–1916*, against which appeared 1, 2, 3, ..., entities covering the letters A and B of the alphabet, both separately and together. He also looked at other listings.

The result of Lotka's investigation was an inverse square law of productivity, by which the

* Corresponding author. Tel. (609) 546-0600; fax (609) 546-9633.

SSDI 0048-7333(94)00780-B

number of people producing n papers is proportional to $1/n^2$. This says that for every 100 authors who produce one paper in a given period of time, there are approximately $100/2^2$, or 25, who produce two papers, $100/3^2$, or 11, who produce three, and so on. Fig. 1, taken from Lotka's paper, illustrates this point.

For a long time, this concept seemed to be largely forgotten. In 1957, W. Shockley, the co-winner of the 1956 Nobel Prize in physics, considered scientific productivity from an institutional viewpoint by analyzing the statistics of individual productivity in research laboratories (Shockley, 1957). His overall conclusion was that

> ...in any large and reasonably homogenous laboratory ...there are great variations in the output of publications between one individual and another.

He also concluded that

> ...the more or less normal distribution of the logarithm of the rate of publication is characteristic of the statistics of the scientific creative process. Perhaps the most important feature of this conclusion is that the rate of publication increases approximately exponentially from individual to individual, taken in order of increasing rate, and that the differences in rates between low and high producers are very large.

Fig. 2, taken from Shockley's paper, shows that the publication productivity of the Brookhaven National Laboratory staff was essentially log-normal over a wide range. Shockley then speculated on some of the possible reasons for the exponential characteristics of productivity, and suggested that one reason might be that the number of ideas that an individual can simultaneously be aware of is dependent on the permutations and combinations of m, and increase very rapidly with increasing m.

He also proposed a different way of rationalizing the productivity difference by suggesting that the many factors involved in publishing a scientific paper might be multiplicative. He then said that if to some approximation the probability that a worker will produce a paper in a given period of time is a product of a number of factors, each of which is normally distributed in a population, then the product of those would be approximately log normal.

In a later paper, Price used a probabilistic model to speculate on the reason for this, saying, in essence, that the probability of somebody publishing a paper increases with the number of papers he publishes, so that you get a relatively steep distribution (de Solla Price, 1976).

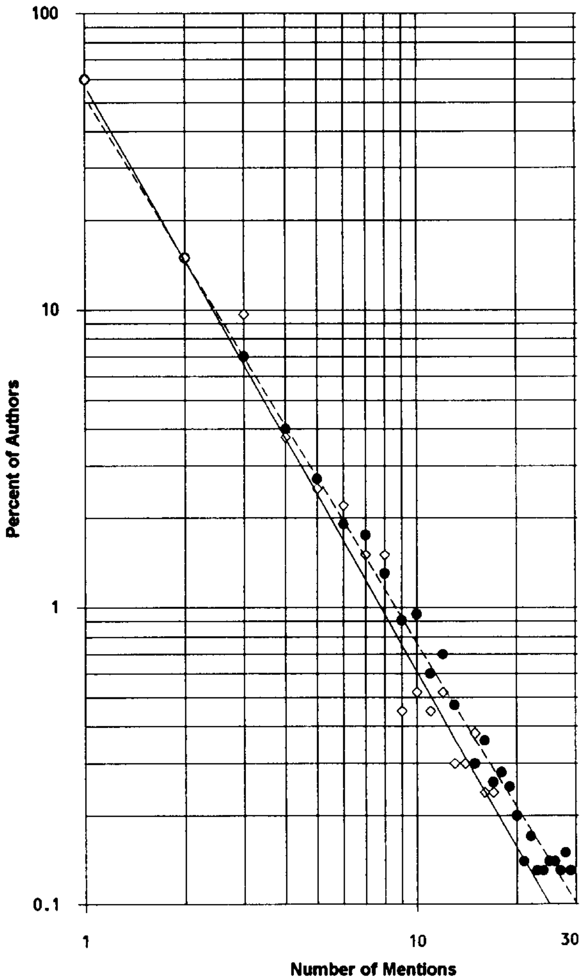

Fig. 1. Publication frequency for authors (replotted from Lotka, 1926). Logarithmic frequency diagram showing number of authors mentioned once, twice, etc., in Auerbache's tables (points indicated by diamonds) and in Chemical Abstracts, letters A and B (points indicated by circles).

Table 1
Concentration of productivity in key researchers. Patent productivity of a world-renowned major manufacturer

Inventor	Number of US patents (15 years)
No. 1, key person	19
No. 2, key support	12
No. 3, originator (retired)	7
Nos. 4 and 5, contributors	6
No. 6, contributor	4
8 other inventors	2
26 other inventors	1

Note: The patents of inventor No. 1 are also highly cited, a further indication of his key importance.

There have since been many other papers in this general area, all of which have shown that the distribution of scientific productivity is highly concentrated (Seglen, 1992). This is similar to the well-known 80–20 rule of thumb that 80% of the work in any organization is done by 20% of the people, and so on.

One other possible reason for this and many of the other highly skewed distributions observed in science is related to the normal distribution of intelligence. If one assumes that publishing scientists are in the upper tail of the normal distribution, then, in fact, in that tail the population drops very dramatically, almost logarithmically, and certainly would approximate the observed rarity of extremely productive individuals and the relative abundance of individuals with a smaller number of scientific contributions.

2. Origin of this paper

In a proprietary study done more than a year ago for a client, CHI was investigating the productivity of inventors within one very well defined area of technology. In this area, our client has a world-renowned position. In looking at the list of inventors and patents, it is quite apparent, as shown in Table 1, that a relatively small number

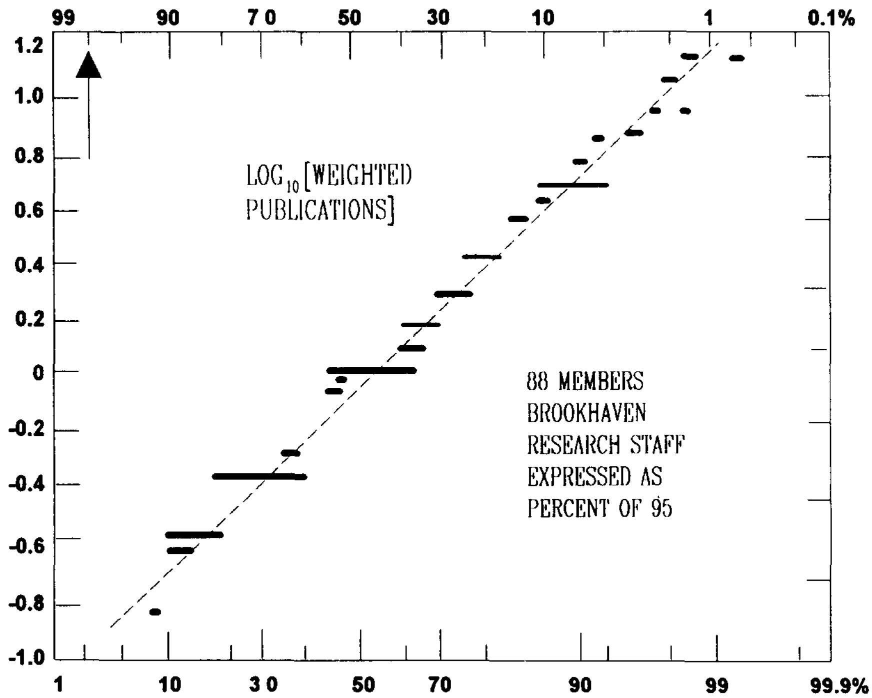

Fig. 2. Comparative distribution of logarithm of 'weighted' rate of publication (replotted from Shockley, 1957). Cumulative distribution of logarithm of 'weighted' rate of publication at Brookhaven National Laboratory plotted on probability paper.

of inventors accounted for a very large fraction of the patents. Specifically, three inventors accounted for more than 42% of the patents. Of these three inventors, the third on the list invented the technology and is now retired. The first man on the list is now the key inventor, and his patents are not only prolific but very highly cited, and many are joint with inventor No. 2.

In analyzing this data, however, we were faced with a problem: whether to do what is commonly called a 'whole count' in bibliometrics and attribute a full patent to each inventor on a patent, or a fractional count and give each inventor partial credit, i.e. if three inventors are on a patent, credit each with one-third of a patent. In looking at our client's data, it turned out that it really made no difference from a distributional viewpoint. Either way the top three inventors would still be the top three inventors, and you would still have a very steep distribution and, as will be shown in Section 4 of this paper, that observation holds in the area of semiconductors.

Because the work for our client was proprietary, we could not discuss it publicly except in very general terms. Therefore, we decided to look at this question in an area in which we have been doing some research, but that is not directly connected to any client. We chose the area of semiconductors, and investigated eight years of patenting for four companies, Xerox, AT&T, Fuji Electric, and Matsushita Electric, using the methodology described in the next section.

3. Methodology

3.1. Company and patent selection

Four companies active in semiconductor technology were selected for the study, two US companies (AT&T and Xerox) and two Japanese companies (Matsushita Electric and Fuji Electric). One reason for including the Japanese companies was to see if the rather group-oriented philosophy of Japanese companies would reflect itself in a less steep distribution – that is, to see if outstanding individuals would be less frequent in Japan. As will be seen in the next section, this is not the case.

A 'filter' was then built using the US Patent Office Classifications (USPOC) to identify patents related to semiconductors and semiconductor applications. The filter was used to capture all US patents in semiconductor technology issued from 1984 to 1991. In the 8-year period, AT&T had 2040 patents, Matsushita had 161, Fuji had 156, and Xerox had 141.

3.2. Inventor name unification

A list of inventors was compiled from the patent set for each company. The inventor's names were sorted and then 'unified' for analysis. The unification step, where different spellings of the same name are made consistent, is a crucial step. For example, consider an inventor Yiu-Huen Wong who obtained eight patents, three under the name Yiu-Huen Wong, three under the name Yiu H. Wong, and two under the name Y.H. Wong. If these names were not unified, a highly prolific inventor such as Wong would be viewed by the computer as three average inventors.

After unification, AT&T had 2310 inventors working in semiconductor technology, Matsushita had 309, Fuji had 211, and Xerox had 163.

3.3. Whole patent counts versus fractional patent counts

Since this is a study on productivity among inventors, the question arises: how does one deal with co-inventors? That is, should each co-inventor get credit for a whole patent or only for a fraction of a patent?

Using whole counts, every inventor whose name appears on a patent is given credit for the whole invention regardless of the number of co-inventors. It is then straightforward to determine how prolific each inventor is for a given company, but it should be noted that this method artificially increases the patent count for each company.

One problem with this counting method is that an inventor who obtains a single patent with four co-inventors is considered as prolific as an inventor who produces his invention singlehandedly.

With fractional patent counts, an inventor is given a simple fractional credit for patents he/she has invented with others. For example, suppose an inventor obtains a first patent with one other person, and another patent with two other people. With fractional counts the inventor is given credit for $\frac{1}{2}+\frac{1}{3}=\frac{5}{6}$ of a patent. Of course, it is highly unlikely that each member of a group of co-inventors contributes equally to an invention, but that assumption has to be made.

Lotka did not have this problem. In 1926, single authorship was the norm. However, for the last 20 years or so multiple authorship has become very common, and there has been much debate about which counting method should be used. Rather than continue the debate, we decided to compute the patent data both ways, using whole counts and fractional counts, and then compare the results.

4. Results

4.1. Whole patent counts

The whole patent count data for the four companies is given in Table 2. For example, for AT&T whole patent counts there were 2310 inventors patenting in semiconductor technology, of which 1522 have only one patent. Lotka's law predicts that $1522/2^2$ (381 people) would have two patents; the actual number is 384. Similarly, Lotka's law predicts that $1522/3^2$ (169 people) would have three patents, while the actual count is 185.

A log–log graph comparing AT&T data, Lotka predicted values, and the linear regression trend (best fit data) is given in Fig. 3. In this case, notice that the most prolific inventor is even more productive than Lotka's law predicts. The Lotka exponent (as computed from the linear regression data) is 2.67 instead of Lotka's 2.0. A detailed look at computing the best-fit line and Lotka-computed exponent is given later.

Similar results were obtained for the other companies. In all but one case, Xerox whole counts, the computed Lotka exponent is greater than 2. Even in the Xerox case, where the exponent is only 1.71, a Lotka-like distribution exists, with the group of inventors having only one patent far outnumbering the prolific inventors.

The Xerox inventor productivity stick diagram in Fig. 4 illustrates this. In the diagram, each stick represents an inventor; the height of the stick

Table 2
Whole patent count data, semiconductor patents 1984–1991

AT&T		Fuji Electric		Matsushita Electric		Xerox	
Patent count	Inventors	Patent count	Inventors	Patent count	Inventors	Patent count	Inventors
22	1	14	1	8	1	18	1
18	3	13	1	7	5	13	1
17	1	10	1	6	3	12	1
14	2	9	1	5	7	11	1
13	1	7	2	4	6	9	1
12	1	6	1	3	10	7	2
11	4	5	3	2	52	6	1
10	3	4	7	1	225	5	5
9	4	3	12			4	5
8	18	2	29			3	10
7	20	1	153			2	16
6	35					1	119
5	35						
4	91						
3	185						
2	384						
1	1522						
4134	2310	356	211	479	309	309	163

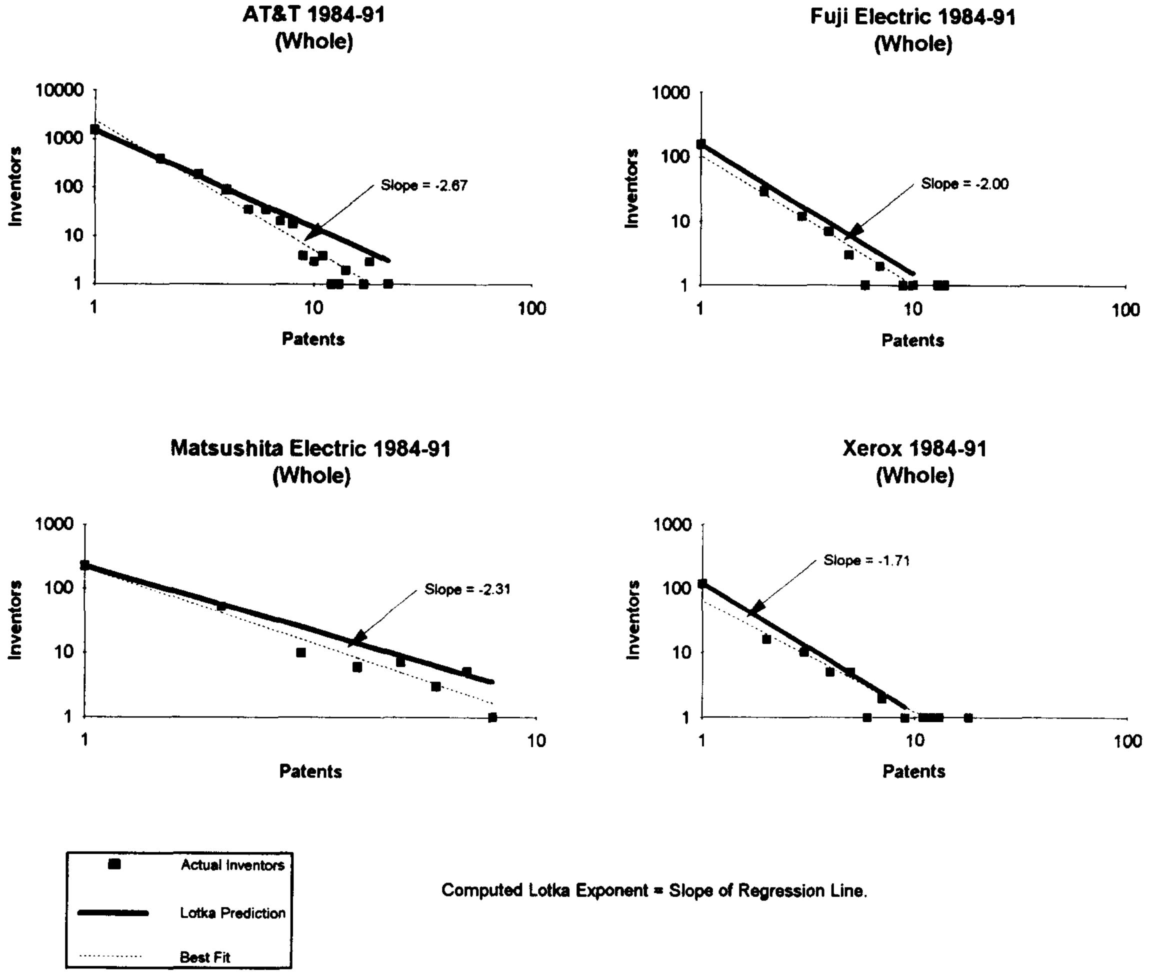

Fig. 3. Whole patent count plots – semiconductors.

represents the number of patents granted to that inventor. The large number (119) of inventors with one patent contrasts sharply with the 13 inventors with ten or more patents.

4.2. Fractional patent counts

The fractional patent count data for the four companies is given in Table 3. Notice that there are no fractions in the patent count column. This is because rounding is necessary in order for Lotka's inverse square law to make sense [1].

[1] If the fractions are not rounded one gets many fragmented groups instead of a few simple clusters at specific numbers of patents.

Simple rounding was used instead of floor rounding (greatest integer rounding) or ceiling rounding. Floor rounding was abandoned because it discards too many legitimate inventors (for example the inventor with $\frac{5}{6}$ of a patent would get credit for zero patents). Ceiling rounding was discarded because it gives an inventor credit for a patent even if he/she had ten or more co-inventors, which defeats the purpose of having two different studies. The number of fractional-count inventors is always lower than reality because inventors with less than half a patent in total are discarded.

Once the patent count numbers are rounded, the computations are done in the same way as with whole counts. For example, for AT&T frac-

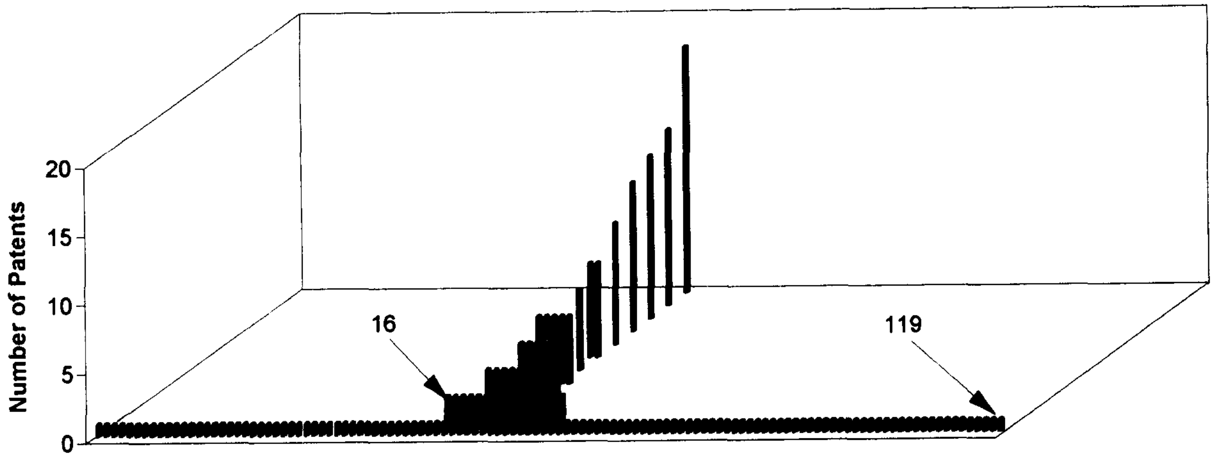

Fig. 4. Xerox whole inventor stick diagram – semiconductor patents 1984–1991. Note: Each stick represents one inventor and all his patents (whole counts).

tional counts, there were 1205 inventors with only one patent in semiconductor technology. Lotka's law predicts that $1205/2^2$ (301 people) would have two patents. The actual count is 213. Similarly, according to Lotka, $1205/3^2$ (134 inventors) should have three patents, while the actual number is 77.

The graph associated with Table 3 is given in Fig. 5. Again notice the Lotka-like distribution, with the most prolific inventor being even more productive than expected. In this case, the Lotka computed exponent is 2.98 instead of 2.0.

Similar results were obtained for the remaining companies. In every case the computed Lotka exponent is greater than 2.0, which means that the highly productive inventors are even more concentrated than Lotka's law predicts. Fig. 5 shows that the computed exponent is closer to Lotka's for the remaining companies.

4.3. Combined results and the computed Lotka exponent

Table 4 contains the data for all of the companies combined from Tables 2 and 3. All four companies may now be viewed on one graph (Fig. 6).

Table 3
Fractional patent count data (rounded) semiconductor patents 1984–1991

At&T		Fuji Electric		Matsushita Electric		Xerox	
Patent count	Inventors	Patent count	Inventors	Patent count	Inventors	Patent count	Inventors
14	1	8	1	7	1	7	1
12	1	7	1	5	1	6	1
11	1	5	4	4	1	5	2
10	1	4	1	3	2	4	1
8	2	3	3	2	11	3	6
7	4	2	8	1	98	2	15
6	6	1	97			1	74
5	13						
4	28						
3	77						
2	213						
1	1205						
2166	1552	161	115	142	114	149	100

The table also gives a column with the trend data for the four companies combined. To obtain this column, the log of the inventors column was taken along with the log of each corresponding patent count. The regression line was then computed and exponentiated to remove the logs. The slope of this line gives the computed Lotka exponent. Recall that the slope of a line is given by: slope $= (y_2 - y_1)/(x_2 - x_1)$. In the case of whole counts we get: $(\log(0.9) - \log(1.52))/\log(22) - \log(18)) = -2.63$, which indicates that for the four companies, the number of people producing N patents is proportional to $1/N^{2.63}$.

Table 4 and Fig. 6 also show similar results for the fractional counts. In this case, the computed exponent is -3.03, which says that for the four companies, the number of people producing N patents is proportional to $1/N^{3.03}$.

It is not obvious why the exponent is higher for fractional counts than it is for whole counts. Nor is it obvious which exponent is more realistic. What is important, though, is that both methods appear to show a power-law relationship similar to Lotka's law.

It is possible that the exponent for all patented technology could be approximately 2.0, as in Lotka's law for scientific papers, or it could be larger, as shown here. A much larger sample of companies would be needed to accurately estimate the exponent, and the restriction to one area of patenting (semiconductors) would have to be researched. However, computing the exponent

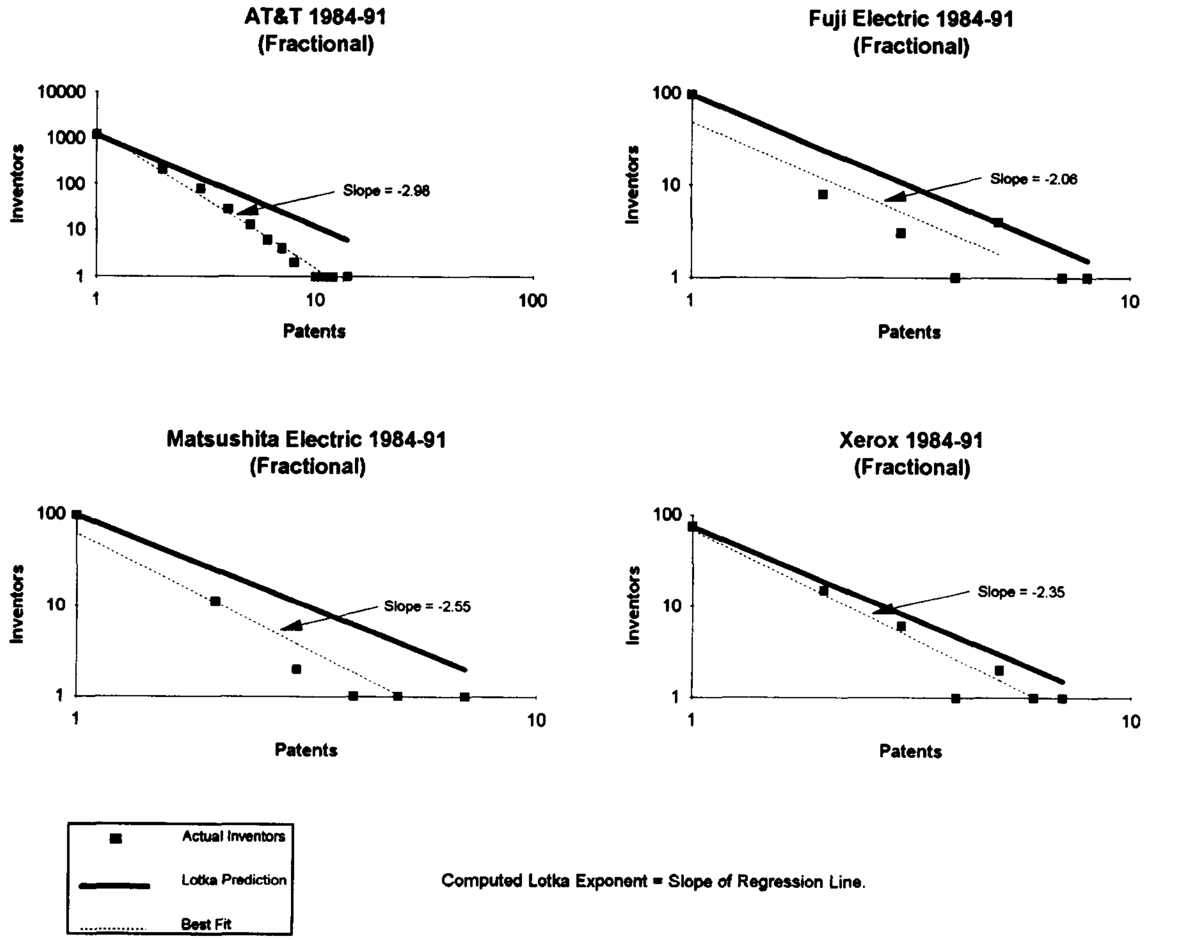

Fig. 5. Fractional patent count plots – semiconductors.

Table 4
All companies combined, 1984–1991

Whole counts				Fractional counts			
Patents	Inventors	Best fit	Prediction	Patents	Inventors	Best fit	Prediction
22	1	0.90	4	14	1	0.64	8
18	4	1.52	6	12	1	1.02	10
17	1	1.76	7	11	1	1.33	12
14	3	2.94	10	10	1	1.78	15
13	3	3.57	12	8	3	3.49	23
12	2	4.40	14	7	7	5.23	30
11	5	5.53	17	6	7	8.34	41
10	4	7.11	20	5	20	14.49	59
9	6	9.37	25	4	31	28.47	92
8	19	12.77	32	3	88	68.01	164
7	29	18.14	41	2	247	232.07	369
6	40	27.19	56	1	1474	1891.81	1474
5	50	43.89	81				
4	109	78.87	126				
3	217	167.91	224				
2	481	487.09	505				
1	2019	3008.17	2019				

exactly was not our goal. We were primarily trying to determine if a Lotka-like relationship holds, i.e. a large number of inventors with one patent and exponentially smaller numbers of people with many patents. That was certainly found to be the case, both for the US and the Japanese companies, and the distributions were found to be even steeper than Lotka's.

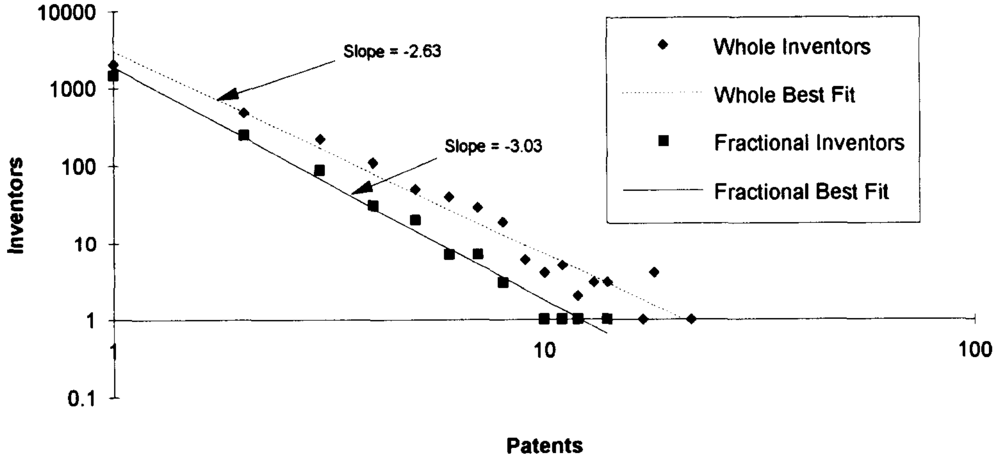

Fig. 6. All companies combined (1984–1991).

4.4. Relative productivity

Another way of showing the result is given in Fig. 7. Fig. 7(a) shows the relative productivity among inventors in the whole patent count study. We define relative productivity as (% patents/% inventors). This graph shows that the top 1% of inventors in a laboratory are 5–10 times as productive as the average inventor in the same laboratory, and that the top 10% of inventors are 3–4 times as productive as the average inventor. Fig. 7(b) for the fractional patent count study shows a

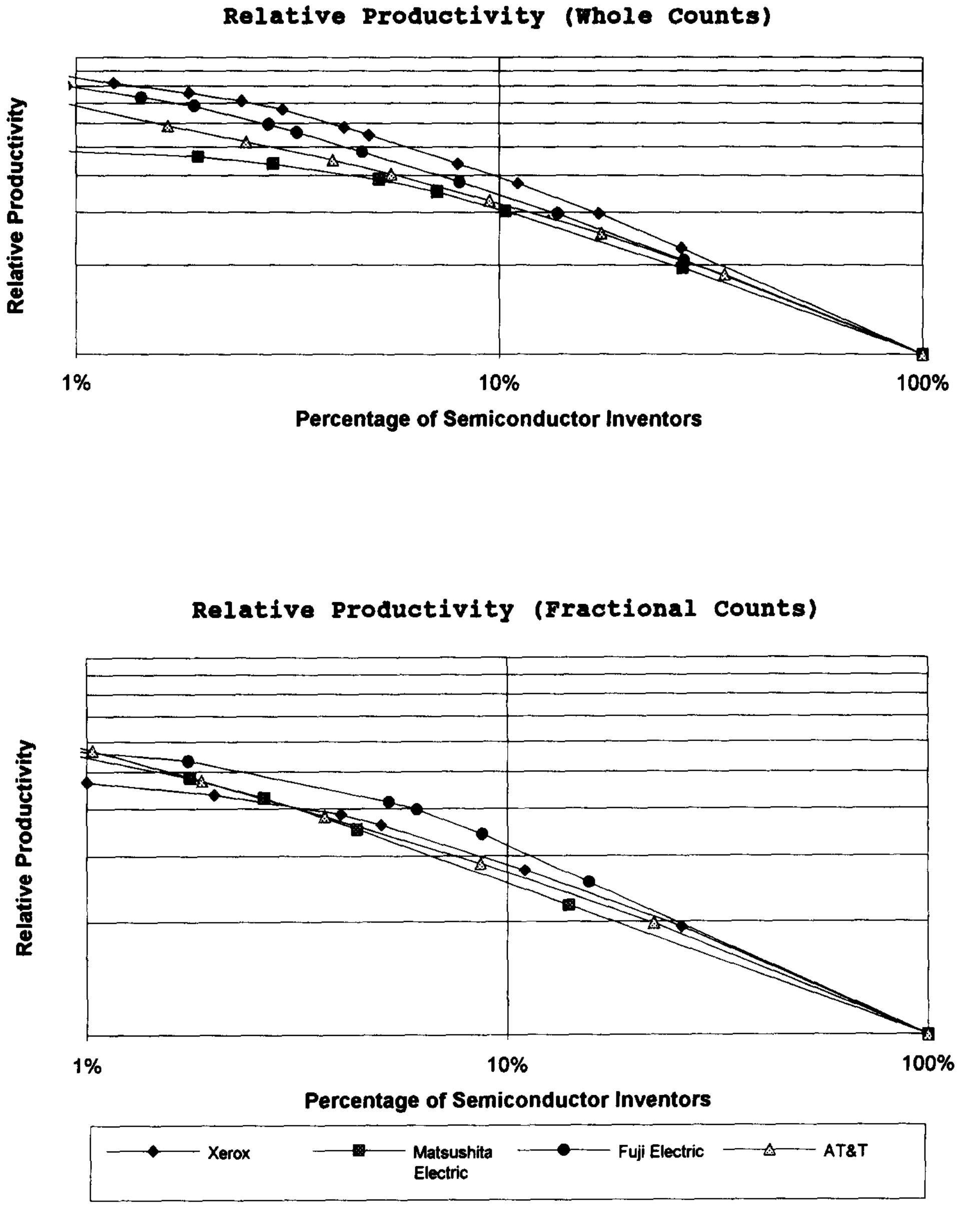

Fig. 7. Relative productivity. (a) Whole counts. (b) Fractional counts.

Table 5
Relative productivity (whole patent counts 1984–1991)

Cumulative patent count	% Total patents	Cumulative inventor count	% Inventors	% Patents/ % Inventors
AT&T				
22	0.53	1	0.04	12.29
76	1.84	4	0.17	10.62
93	2.25	5	0.22	10.39
121	2.93	7	0.30	9.66
134	3.24	8	0.35	9.36
146	3.53	9	0.39	9.06
190	4.60	13	0.56	8.17
220	5.32	16	0.69	7.68
256	6.19	20	0.87	7.15
400	9.68	38	1.65	5.88
540	13.06	58	2.51	5.20
750	18.14	93	4.03	4.51
925	22.36	128	5.54	4.04
1289	31.18	219	9.48	3.29
1844	44.61	404	17.49	2.55
2612	63.18	788	34.11	1.85
4134	100.00	2310	100.00	1.00
Fuji Electric				
14	3.93	1	0.47	8.30
27	7.58	2	0.95	8.00
37	10.39	3	1.42	7.31
46	12.92	4	1.90	6.82
60	16.85	6	2.84	5.93
66	18.54	7	3.32	5.59
81	22.75	10	4.74	4.80
109	30.62	17	8.06	3.80
145	40.73	29	13.74	2.96
203	57.02	58	27.49	2.07
356	100.00	211	100.00	1.00
Matsushita Electric				
8	1.67	1	0.32	5.16
43	8.98	6	1.94	4.62
61	12.73	9	2.91	4.37
96	20.04	16	5.18	3.87
120	25.05	22	7.12	3.52
150	31.32	32	10.36	3.02
254	53.03	84	27.18	1.95
479	100.00	309	100.00	1.00
Xerox				
18	5.83	1	0.61	9.50
31	10.03	2	1.23	8.18
43	13.92	3	1.84	7.56
54	17.48	4	2.45	7.12
63	20.39	5	3.07	6.65
77	24.92	7	4.29	5.80
83	26.86	8	4.91	5.47
108	34.95	13	7.98	4.38
128	41.42	18	11.04	3.75
158	51.13	28	17.18	2.98
190	61.49	44	26.99	2.28
309	1	163	1	1

Table 6
Relative productivity (fractional patent counts 1984–1991)

Cumulative patent count	% Total patents	Cumulative inventor count	% Inventors	% Patents/ % Inventors
AT&T				
14	0.65	1	0.06	10.03
26	1.20	2	0.13	9.31
37	1.71	3	0.19	8.84
47	2.17	4	0.26	8.42
63	2.91	6	0.39	7.52
91	4.20	10	0.64	6.52
127	5.86	16	1.03	5.69
192	8.86	29	1.87	4.74
304	14.04	57	3.67	3.82
535	24.70	134	8.63	2.86
961	44.37	347	22.36	1.98
2166	100.00	1552	100.00	1.00
Fuji Electric				
8	4.97	1	0.87	5.71
15	9.32	2	1.74	5.36
35	21.74	6	5.22	4.17
39	24.22	7	6.09	3.98
48	29.81	10	8.70	3.43
64	39.75	18	15.65	2.54
161	100.00	115	100.00	1.00
Matsushita Electric				
7	4.93	1	0.88	5.62
12	8.45	2	1.75	4.82
16	11.27	3	2.63	4.28
22	15.49	5	4.39	3.53
44	30.99	16	14.04	2.21
142	100.00	114	100.00	1.00
Xerox				
7	4.70	1	1.00	4.70
13	8.72	2	2.00	4.36
23	15.44	4	4.00	3.86
27	18.12	5	5.00	3.62
45	30.20	11	11.00	2.75
75	50.34	26	26.00	1.94
149	100.00	100	100.00	1.00

similar result. The numbers that support Fig. 7 are given in Tables 5 and 6.

5. Conclusion and implications

Although the detailed study has been done for only the one case of semiconductors, there is no reason to believe that the distribution would be any different in any other field of technology, and we certainly hope, in the future, to look at this phenomenon in other areas.

However, given that essentially the same thing was found in the one proprietary area for our client, and that it is well known and widespread throughout scientific paper publications, it seems that this concentration is widespread and general.

The implications of this from a strategic viewpoint are quite important. Perhaps one of the most important implications is that this codifies a great deal of intuitive perception on the part of many laboratory managers. In discussing these findings with the managers of major laboratories, they can almost always point out the one, two, or three individuals who are really driving their laboratory. What this paper says is that this is not an accident; that the key role of a few researchers seems to be a law of nature, and, therefore, a laboratory manager should expect to find such a distribution in his laboratory.

It also seems clear that, in order for a company to stay active and technologically competitive, it must identify, nurture, and work to retain its leading producers. The finding of this paper says that these leading producers are going to be relatively rare in any given laboratory and, therefore, they must be retained if a company is to stay creative and competitive.

Furthermore, in a technically based acquisition it is quite clear that a major consideration is whether the key inventors remain with the acquisition. In our consulting practice we are aware of a number of cases where a company was acquired without the key inventors, and in most of these cases it has been relatively disastrous for the acquiring company. The finding of this paper simply says that you should do a very careful analysis of key people before a technological acquisition, and make sure that the key people are going to stay.

Finally, it is generally important that a company's rewards structure, both monetary rewards and internal recognition, should reflect the fact that the most productive inventors may be 5–10 times as productive as the average inventor.

It certainly appears that technological creativity and productivity, just like scientific creativity and productivity, are very highly concentrated in a population, and in the minds and abilities of a relatively small number of highly talented individuals. Companies should make every effort to retain and nurture these key contributors.

6. References

De Solla Price, D., 1963, Little Science, Big Science (Yale University Press, New Haven, CT).

De Solla Price, D., 1976, A general theory of bibliometric and other cumulative advantage processes, Journal of the American Society for Information Science 27(5) September–October.

Lotka, A.J., 1926, The frequency distribution of scientific productivity, Journal of the Washington Academy of Sciences 16, June, 317–323.

Seglen, P.O., 1992, The skewness of science, Journal of the American Society for Information Science 43(9) 628–638.

Shockley, W., 1957, On the statistics of individual variations of productivity in research laboratories, Proceedings of the IRE, March, 279–290.