

# Scientific productivity and accumulative advantage: a thesis reassessed in the light of international data

Roland Mittermeir,\* *Department of Computer Sciences, University of Texas at Austin*

Karin D. Knorr, *Institute for Advanced Studies, Vienna*

## 1. INTRODUCTION

Ever since Lotka (1926) published his inverse square law of scientific productivity, both the nature and origin of the phenomenon have been investigated in the literature. Although the specific parameters of those functions which indicate what proportion of scientists produces what bulk of scientific products vary with different indices (Davis, 1941; Leavens, 1953; Price, 1963; Britton, 1964; Allison & Stewart, 1974; Knorr *et al.*, 1975), there is general agreement that the distribution of scientific products is highly skewed, irrespective of the measurement chosen or the discipline investigated. Alternative hypotheses for explaining these differences among U.S. scientists were described by Cole & Cole (1973) and Allison & Stewart (1974). A third account of differences in publication rates is implied in a recent argument by Bourdieu (1975), and substantiated by Knorr (1977) and Knorr *et al.* (1978).

The present paper will first summarize the various hypotheses offered in explanation, and then proceed to reassess the plausibility of accumulative advantage based upon data from an international study of scientific productivity conducted in 6 European countries from 1973 to 1977.

### 1.1. *The sacred spark hypothesis*

Those who argue in favour of a sacred spark model believe that differences in productivity result from unequal distribution of intellectual and productive capabilities among various scientists. Although the existence of such differences cannot be denied, we must ask whether they can account for the size of the effect encountered, given the homogenizing effect of the long-term education which scientists undergo. According to Britton (1964), scientific productivity is much more unevenly distributed than, say, income among individuals—a case in which we cannot assume the equalizing effects of a common education. While marginal groups—which scientists are as compared to the rest of the population—often show more extreme differences than less autonomous and elitarian groups, such a grossly unequal distribution as that published by Price (1963), with 6% of the scientists publishing 50% of all papers, demands further investigation.

### 1.2. *The accumulative advantage hypothesis*

The accumulative advantage hypothesis was first advanced by Merton (1968). He found that scientists who are well known receive more recognition for their work than those who are less

\* The paper was prepared while the author was with the Institut für Digitale Anlagen, Technische Universität Wien.

well known, a phenomenon he calls, in analogy to the Gospel of St. Matthew, XXV/29,<sup>1</sup> 'Matthew Effect'. Since the formal control mechanisms of science, such as reviewing and publication processes, are seemingly not affected by differences in status among the authors of papers submitted (Zuckerman & Merton, 1971), explaining the Matthew effect is of special interest. The effect was generalized by Cole & Cole (1973) to hold not only for recognition, but for scientific productivity in general. The process is seen as consisting of two reinforcement loops. First, since scientists differ (slightly) at the beginning of their careers, some will make contributions to their discipline earlier than others. Once recognized, these scientists will be motivated to maintain or increase their recognition through further publication, and will also be influenced by the expectations placed upon them by colleagues. Second, recognition will open up access to such resources as laboratory manpower, project funds, early access to new information and so forth. Apart from being instrumental in any scientific work, these resources are thought to be rewarding in themselves, and to have a further positive effect on a scientist's productivity. Allison & Stewart provide evidence for this thesis by assuming that, if accumulative advantage exists in this sense, the distribution of productivity among scientists should become increasingly dispersed as time goes on. This assumption was tested through the calculation of a Gini coefficient<sup>2</sup> which measures the productivity-dispersion across different age strata for 1947 U.S. scientists in biology, mathematics, chemistry and physics. The data confirmed that the Gini index for publications and citations increases substantially with the age of the scientists in terms of years passed since earning the Ph.D. (Allison & Stewart, 1974).

### 1.3 *An alternative hypothesis*

Indirect criticism of the accumulative advantage hypothesis comes from a theory which considers scientists to be involved in a competitive struggle for the accumulation of scientific credit (Bourdieu, 1975; Knorr, 1977). According to this theory, it depends on the professional position a scientist holds what

<sup>1</sup> For to every one who has will more be given, and he will have abundance; but from him who has not, even what he has will be taken away (The New Oxford Annotated Bible, O.U.P., 1973).

<sup>2</sup> The Gini-index is mostly used in econometrics as a quantification of inequality. It measures the difference of the integral of the Lorenz-curve, showing the empirically found distribution, with the integral over a straight line representing an equal distribution. The quantity thus obtained is then standardized to obtain values between 0 (equal distribution) and 1 (maximal inequality).

counts as scientific credit. It follows, then, that for scientists in different institutions, countries or scientific fields we cannot automatically assume that credit will materialize in the same form. Perhaps even more importantly, we cannot assume that an accumulation of advantages (i.e. resources relevant to a given position) will be mirrored solely by an increase in productivity of publications and citations received. Thus, while theory implies that individual gains rest upon potentially accumulative advantages, this does not necessarily entail the Allison & Stewart assumption that accumulative advantage coincides with an increasingly greater probability of publication or citation over each successive time interval. To the contrary, it suggests that the number of publications or citations during a given period may yield diminishing returns for a scientist once a certain level of achievement has been reached, and that 'recognized' scientists may want to transform their credit into scientific honours, prestigious positions in associations, journal editorships, government consultantships or invited lectureships, or even into monetary advantages. Finally, we should not forget how implausible it is to assume that publication or citation rates would increase indefinitely in the case of scientists, when this is not the case with other population statistics where processes of re-enforcement operate.

## 2. APPROACH AND METHODOLOGY OF THE PRESENT STUDY

With the above hypothesis in mind, we will try to reassess the thesis proposed by Allison and Stewart using data on scientists in six European countries (Austria, Belgium, Finland, Hungary, Poland and Sweden). For each country a sample of 150 to 250 research units, stratified by type of organization and scientific field was taken. The final dataset comprises 1,222 research units and 4,057 scientists, principally from the natural and technological sciences<sup>3</sup> working in academic settings, cooperative institutes and industrial enterprises.<sup>4</sup> Data were collected in 1974 by means of 5 different questionnaires (based on a pretest of 150 research units in three countries). Information from unit heads was obtained by personal interview; individual scientists and the technical and service staff of the unit were given self-administered questionnaires. External evaluators of work done by the unit were addressed either by personal interview or administered questionnaires. The response rate varied from 70% to 85% (depending on the country and

<sup>3</sup> For selection of scientific fields *Unesco Proposed International Standard Nomenclature for Fields of Science and Technology* has been used. The international dataset includes the following disciplines (in parentheses number of respondents): Physics (280); Chemistry (825); life Sciences (708); Earth and Space Sciences (288); Agricultural Sciences (331); Medical Sciences (189); Technological Sciences (1176); Social Sciences (258) and others such as Mathematics, Astronomy, etc. (62).

<sup>4</sup> In addition to universities the category 'academic settings' includes institutes attached to universities, and academies of science. The category of 'cooperative institutes' comprises those research units which belong to institutions that serve wholly or partly a branch of industry, and/or to governmental institutions. In the international dataset most respondents are from academic institutions (2566). The rest are working in cooperative (744) or industrial (657) organizations.

field of study), with no indication of a serious response bias by rank of respondent, field or type of organization.

While Allison and Stewart used the number of publications and the number of citations received as indicators of productivity, we rely on the number of publications and the number of patents as an additional indicator especially relevant for industrial research.<sup>5,6</sup> Using these performance indicators, Gini indices were constructed for eight age strata, as defined by number of years of R & D-experience (a measure similar to years since earning the Ph.D. as used by Allison and Stewart).

## 3. RESULTS AND DISCUSSION

The results initially obtained when calculating the Gini-coefficients across the European data suggested a negative relationship between publication (and patent) dispersion and research experience, thereby suggesting that the inequality of scientists with respect to number of publications (and patents) actually *decreased* over time. These results not only conflicted with the Allison and Stewart data, but ran counter to our most pessimistic expectations for non-confirmation. Indeed, these peculiar results suggested a methodological problem. However, when the data were controlled for scientists who had produced nothing during the time period under examination, the initial counter-intuitive trends vanished from the data. Since zero-producers can account for the decrease of inequality in that their number is almost bound to decline with increasing time spent in R & D, and since the youngest age stratum—which overlaps to a large degree with our non-producers—was excluded from Allison and Stewart's analysis, the group of zero-producers were removed from further analysis. We feel that these preliminary investigations lend support to the frequent speculation that the first publication is often the most difficult to secure.

The results thus obtained are shown in Table 1 (for publications) and Table 2 (for patents).

In displaying our results we tried to make them comparable to those reported by Allison and Stewart.

In the upper part of each table the number of respondents ( $N$ ) and the Gini index ( $G_i$ ) for each respondent category classified according to the years of research experience the respondent had, are given for the total sample as well as subsamples stratified according to supervisory position, organizational affiliation<sup>7</sup> and scientific field. (The 3 rows in the centre of each column.) In the row 'producers' the number of individuals who actually produced at least one scientific 'product' (publication, patent) during the last 3 years before the survey was taken is

<sup>5</sup> A detailed discussion of the advantages and drawbacks of various indicators of scientific performance—including citation counts—can be found in Whitley & Frost (1971) or in Mittermeir (1978).

<sup>6</sup> Allison and Stewart's performance indicators relate to a five-year period before the data were collected, while ours relate to a three-year period. The use of citation rates in the present study was made impossible by the fact that some European countries are only insufficiently covered by the Science Citation Index.

<sup>7</sup> Actually a typology variable, classifying research units according to organizational affiliation and scientific discipline has been used (Cole, 1978). This typology is based on output patterns and was proven to be stable across various organizational variables.

Table 1. Inequality with respect to publications authored depending on research experience

Years with R & D experience	Total sample		Heads of units		Staff scientists		Academic natural science research unit		Academic medical and social science research unit		Academic applied science		Governmental and cooperation research units		Industrial research units		Mathematics and physics		Chemistry		Biology		Agriculture	
	N	Gi	N	Gi	N	Gi	N	Gi	N	Gi	N	Gi	N	Gi	N	Gi	N	Gi	N	Gi	N	Gi	N	Gi
	0-2	175	0.247	9	0.357	166	0.223	82	0.253	18	0.316	28	0.113	29	0.221	18	0.238	13	0.300	45	0.200	28	0.327	14
3	188	0.248	15	0.293	173	0.278	91	0.288	23	0.263	31	0.273	30	0.262	13	0.294	14	0.254	48	0.269	39	0.260	11	0.277
4-5	379	0.328	40	0.332	339	0.325	199	0.315	35	0.259	45	0.377	62	0.342	38	0.274	50	0.280	88	0.335	70	0.311	27	0.373
6-7	263	0.346	47	0.342	216	0.346	130	0.349	29	0.309	34	0.343	34	0.329	36	0.298	36	0.361	58	0.350	49	0.354	21	0.292
8-10	360	0.336	97	0.324	263	0.336	190	0.321	34	0.299	39	0.346	64	0.330	33	0.388	35	0.310	76	0.328	85	0.321	24	0.310
11-15	400	0.335	166	0.329	231	0.329	173	0.294	50	0.339	52	0.346	83	0.347	42	0.374	38	0.258	67	0.336	85	0.298	46	0.323
16-21	397	0.314	253	0.284	144	0.328	164	0.269	47	0.267	62	0.309	90	0.345	34	0.384	24	0.274	83	0.323	87	0.250	35	0.264
22-26	464	0.301	380	0.283	84	0.349	219	0.266	67	0.310	59	0.33	93	0.296	26	0.387	33	0.334	93	0.280	116	0.247	44	0.311
<i>Sample size</i>																								
Producers	2,626		1,007		1,619		1,248		303		350		485		240		243		558		559		222	
All relevant	4,000		1,196		2,804		1,684		376		566		737		637		341		817		698		329	
All (MD included)	4,057		1,222		2,835		1,702		381		573		744		657		341		825		708		331	
<i>Regression coefficients</i>																								
<i>a</i>	0.308		0.336		0.291		0.305		0.289		0.284		0.299		0.289		0.287		0.298		0.320		0.316	
<i>b</i> × 10 <sup>2</sup>	0.030		0.152		0.191		0.089		0.049		0.200		0.084		0.346		0.084		0.038		-0.201		-0.050	
<i>r</i>	0.113		0.669		0.552		0.342		0.206		0.322		0.225		0.709		0.277		0.093		-0.623		-0.182	

Table 2. Inequality with respect to patents depending on research experience (replication for categories not containing a sufficient number of cases excluded)

Years with R & D experience	Total sample		Heads of units		Staff scientists		Academic natural science research units		Academic applied science research units		Industrial research units		Academic governmental research units	
	N	Gi	N	Gi	N	Gi	N	Gi	N	Gi	N	Gi	N	Gi
0-2	19	0.199	3	0.133	16	0.199	3	0	4	0.167	6	0.167	6	0.233
3	24	0.265	3	0.167	21	0.227	4	0	4	0	5	0.280	10	0.278
4-5	72	0.267	17	0.261	55	0.260	13	0.173	12	0.242	14	0.222	30	0.242
6-7	44	0.258	13	0.268	31	0.240	6	0.167	5	0.171	8	0.257	24	0.222
8-10	66	0.279	23	0.271	43	0.282	21	0.250	5	0	12	0.290	28	0.268
11-15	88	0.279	41	0.266	47	0.256	16	0.196	15	0.287	17	0.275	37	0.285
16-21	105	0.276	61	0.275	44	0.278	21	0.304	20	0.214	20	0.289	38	0.271
22-56	102	0.287	100	0.285	22	0.272	40	0.310	19	0.266	22	0.276	29	0.215
<i>Sample size</i>														
Producers	540		261		279		124		84		202		104	
All relevant	4,006		1,203		2,803		1,678		567		644		739	
All (MD included)	4,057		1,222		2,835		1,702		573		657		744	
<i>Regression coefficient</i>														
<i>a</i>	0.248		0.209		0.236		0.088		0.118		0.259		0.238	
<i>b</i> × 10 <sup>2</sup>	0.131		0.270		0.134		0.733		0.424		-0.064		-0.159	
<i>r</i>	0.580		0.586		0.581		0.750		0.467		-0.293		-0.459	

listed. For sake of comparison, the next two lines list the number of individuals who gave a valid answer to the respective questions (i.e. 'producers' + those reporting a '0' for their respective product over the last 3 years). Finally the total sample size for each subcategory is reported.

In the last three lines the regression coefficients and the correlation coefficient between median age per category and the respective Gini-coefficient is reported. It should be noted, however, that the *b*'s are (as in Allison and Stewart) multiplied by 100 to ease readability.

For patents only the subsections of the sample with a reasonable amount of respondents are shown.

As can be seen from the tables the negative relationships obtained by looking at all respondents vanish except for a few cases in which they appear to be rather low. However the results do still not corroborate the assumption that publication- (patent) inequality increases linearly with age. Neither do the Gini-coefficients rise uniformly or substantially nor does the linear regression of the Gini measures on research experience confirm such an increase. Beta coefficients remain generally very low. Looking at publications, inequality increases are highest in industrial research and negative in biology—the only result which points in the same direction as those of the previous article, which also reports lowest Beta coefficients (0.00480) for biology. It is also in industrial research where the correlation assumes a noteworthy positive value, while it is significantly below zero in biology (both industrial and academic) and remains negative for academic natural science in general. It is also negative with regard to heads of research units, which supports the contention that a steady increase in publication output would represent diminishing returns for those who had attained a certain position.

It is worthwhile to note that some of these results appear reversed in the case of patents. There is a slightly negative regression coefficient of the Gini indices on age in industrial research units, while the highest, positive beta is obtained for academic natural science, and a low but equally positive one for heads of units in general. Thus, if there is any increased inequality with respect to the output measures studied, it will be for patents in academic natural science. However, low population values in the first age stratum render the significance of this coefficient problematic, as do low values in general for several of the categories studied in regard to patents. There appears to be no marked difference between staff scientists not yet holding a position and those who are leading a research unit, as was the case with publications.

We conclude that, with the possible exception of patents in academic natural science, our hypothesis that accumulative advantages will not necessarily lead to increased publication inequality with age is substantiated by our data. If the trends suggested by the low beta coefficients we obtained allow for any interpretation, it is that the effect will vary not only with respect to different fields and kinds of output, but also with the position a scientist holds in his field. This latter point is strongly supported by a study which found that differences in publication rate among individual scientists can be at best explained by their respective research positions (Knorr *et al.*, 1978). Since Allison and Stewart did not take position into account, we cannot know how their results would have turned out if they had proceeded from different assumptions. The question left open, of course, is why their results are still so different from ours, if only professional age alone is taken into account. This difference cannot be explained by the fact that Allison and Stewart only investigated university scientists, since we found that, if any-

thing, the correlations are stronger for industrial and government research, and are partly negative for academic science.

Another attempt to explain the differences can be made by referring to other variations within the respective populations. Of the 1,947 scientists analyzed by Allison and Stewart, only 89% were reported to be currently engaged in research (Hagstrom, 1974) whereas scientists not active in research were excluded from the present study. Since the 11% not engaged in research might be disproportionately located in higher age strata and have had either administrative positions not involving any publications, or high rank supervisory positions implying a high number of publications due to automatic co-authorship,<sup>8</sup> they might account for the increased dispersion within these strata of Allison and Stewart's data. Finally, an explanation might be sought in the cultural differences between the U.S. and many European countries, rather than from any subtle differences in approach between the two studies. One could venture the hypothesis that in the U.S. academic system position is more identified with publication (and citation) output, than in the European countries studied, although other forms of credit clearly do exist too. Thus, while the thesis that position determines the forms of credit sought and the means by which they are attained remains generally valid, it is possible that established academic science in the U.S. in the traditional fields of chemistry, physics and mathematics can be thought of as a special case in which position and publication and citation-rates are closely tied together, and where the latter dominate other forms of credit with respect to position. Of course, this suggests that further research may have to control for an even larger number of variables, and to take into account such factors as position in a field or institution as well as scientific discipline, type of organization, and the social structure of the country in which the investigation takes place.

#### ACKNOWLEDGMENT

We are grateful to F. Andrews and G. Cole for many helpful comments on an earlier draft of this paper. We also want to acknowledge the corporate effort of the international research team which collected the data we have been using.

#### REFERENCES

- Aichholzer, G., Mittermeir, R. & Waller, G. (1977) 'On the differential importance of human-relations—aspects for research activities, a comparison between academic and industrial research units. Conference of the I.S.A. Committee on the Sociology of Science, Budapest.
- Allison, P. D. & Stewart, J. A. (1974) 'Productivity differences among scientists: evidence for accumulative advantage.' *American Sociological Review*, Vol. 39, 596–606.
- Andrews, F. M. (1975) 'Social and psychological factors which influence the creative process', in: Taylor, Getzels (ed) *Perspectives in Creativity*, Chicago: Aldin Publ.
- Bayer, A. E. & Folger, J. (1966) 'Some correlates of a citation measure of productivity in science.' *Sociology of Education*, Vol. 39, Fall, 381–389.
- Bourdieu, P. (1975) 'The specificity of the scientific field and the social conditions of the progress of reason.' *Social Science Information*, Vol. 14 (6), 19–47.
- Britton, J. P. (1964) 'The productivity of scientists: a prelude to manpower studies.' Mimeographical paper, Yale University. Cit.: Allison and Stewart (1974).
- Cole, G. (1979) 'Patterns of influence and performance—a typology of research units', in: F. M. Andrews (ed.) *Scientific Productivity: The Effectiveness of Research Groups in Six Countries*. London: Cambridge University Press.
- Cole, J. & Cole, S. (1967) 'Scientific output and recognition; a study in the operation of the reward system in science.' *American Sociological Review*, Vol. 32, June, 377–390.
- Cole, J. & Cole, S. (1973) *Social Stratification in Science*. Chicago: Chicago University Press.
- Davis, H. T. (1941) *The Analysis of Economic Time Series*. Trinity University Press.
- Hagstrom, W. O. (1974) 'Competition in science.' *American Sociological Review*, Vol. 39, Feb. 1–18.
- Knorr, K. D., Mittermeir, R., Aichholzer, G. & Waller, G. (1979) 'Individual publication productivity as a social position effect in academic and industrial research units', in: F. M. Andrews (ed.) *Scientific Productivity: The Effectiveness of Research Groups in Six Countries*. London: Cambridge University Press.
- Knorr, K. D. (1977) 'Producing and reproducing knowledge: descriptive and constructive? Toward a model of research production.' *Social Science Information*, Vol. 16 (6), 669–696.
- Knorr, K. D., Mittermeir, R., Aichholzer, G. & Waller, G. (1975) 'Internationale Vergleichsstudie über die Organisation und Effektivität von Forschungseinheiten.' Vol. I. *Kapazität und Qualität naturwissenschaftlich-technischer Forschung in Österreich*. Wien: Institut für Höhere Studien.
- Leavens, D. H. (1953) 'Communication from Dickson H. Leavens.' *Econometrica*, Vol. 21, October, 630.
- Lotka, A. J. (1926) 'The frequency distribution of scientific productivity.' *Journal of the Washington Academy of Sciences*, Vol. 16, 317–323.
- Merton, R. K. (1968) 'The Matthew effect in science.' *Science*, Vol. 59, 56–63.
- Mittermeir, R. (1978) 'Leistungsdeterminanten von Forschungseinheiten.' Ph.D. Thesis, Wirtschaftsuniversität Wien.
- Price, D. J. de Solla (1963) *Little Science, Big Science*. New York: Columbia University Press.
- Shockley, W. (1957) 'On the statistics of individual variations of productivity in research laboratories.' *Proceedings of the Institute of Radio Engineers*, Vol. 45, March, 279–290.
- Taylor, C. W. & Barron, F. (1963) *Scientific Creativity: In Recognition and Development*. New York: John Wiley.
- Whitley, R. & Frost, P. A. (1971) 'The measurement of performance in research.' *Human Relations*, Vol. 24 (2), 161–178.
- Zuckerman, H. & Merton, R. K. (1971) 'Patterns of evaluation in science, institutionalization, structure and functions of the referee system.' *Minerva*, Vol. 9, 66–100.

<sup>8</sup> We are referring to the fact that in many sciences laboratory heads are automatically entitled to be mentioned as a (last placed) co-author of studies done in their lab.