Cumulative Advantage and Inequality in Science

Paul D. Allison; J. Scott Long; Tad K. Krauze

CUMULATIVE ADVANTAGE AND INEQUALITY IN SCIENCE*

PAUL D. ALLISON
University of Pennsylvania

J. SCOTT LONG
Washington State University

TAD K. KRAUZE
Hofstra University

The hypothesis of cumulative advantage is widely accepted in the sociology of science, but empirical tests have been few and equivocal. One approach, originated by Allison and Stewart (1974), is to see whether inequality of productivity and recognition increases as a cohort of scientists ages. This paper extends their work by examining true cohorts of biochemists and chemists rather than synthetic cohorts. Increasing inequality is observed for counts of publications but not for counts of citations to all previous publications. It is also shown that a mathematical model of cumulative advantage does not imply increasing inequality. When the model is modified to allow for heterogeneity in the rate of cumulative advantage, however, increasing inequality is implied.

THE CUMULATIVE ADVANTAGE HYPOTHESIS

In the last decade the principle of cumulative advantage (Merton, [1942] 1973, 1968) has become a dominant theme in the study of stratification in science. At least three books have invoked the idea of cumulative advantage as a central explanatory principle (Cole and Cole, 1973; Zuckerman, 1977; Gaston, 1978). Numerous articles have sought to provide empirical support (Cole, 1970; Allison and Stewart, 1974; Faia, 1975; Reskin, 1977; Long, 1978; Mittermeier and Knorr, 1979; Hargens et al., 1980) or to develop richer or more rigorous theoretical accounts of the process (Price, 1976; Goldstone, 1979; Turner and Chubin, 1979; Yablonsky, 1980; Rao, 1980).

The essence of cumulative advantage is well captured by the commonsense notion that "the rich get richer at a rate that makes the poor become relatively poorer" (Merton, 1968). Yet the simplicity of this formulation belies the diversity and complexity of the underlying processes. In science the key form of riches is recognition from peers (prestige) for published research (Merton, 1957; Hagstrom, 1965; Storer, 1966). Scientists who are rich in recognition find it easier to get the resources that facilitate research: grants, free time, laboratories, stimulating colleagues, capable students, etc. They are also encouraged by their colleagues to continue to invest time and energy in research (Zuckerman and Merton, 1972). As a consequence their research productivity is likely to increase—or at least be maintained at high levels—and this brings additional recognition. In contrast, scientists who receive little recognition for their research efforts get little in the way of resources and encouragement, thus reducing their chances for future productivity and recognition.

Even without changes in research productivity, it appears that prestigious scientists get further recognition more easily than unknown scientists. In his seminal paper on the Matthew effect, Merton (1968) argued that for a variety of reasons scientists tend to choose their reading matter on the basis of an author's preceding reputation. As a result, two publications of equal intrinsic merit will receive differential recognition if the authors are unequal in prestige. This process can continue to operate even after a scientist has retired from research and publication (Allison and Stewart, 1975).

Why has the hypothesis of cumulative advantage won such widespread acceptance? One reason is that it goes a long way toward explaining the enormous inequality of productivity and recognition in science. It is also so intuitively plausible that for many it simply *must* be true. Finally, the hypothesis appears to be equally appealing to both critics and admirers of the stratification system in science, in part because it cuts across the established principles of universalism and particularism. On the one hand, Cole and Cole (1973) and

Zuckerman (1977) have maintained that strong adherence to universalistic evaluation actually accelerates cumulative advantage by concentrating resources among those best equipped to use them. Turner and Chubin (1979), on the other hand, have emphasized the importance of luck in the operation of cumulative advantage and, hence, have questioned whether the resulting distribution of resources corresponds to the distribution of talent. Thus, the debate has centered on the interpretation rather than the existence of cumulative advantage.

This wholesale endorsement of the cumulative advantage hypothesis is all the more remarkable in that the quantitative evidence is tenuous and often equivocal. One approach has been to test separately the many postulated causal links in the cumulative advantage process. There has been considerable interest, for example, in determining whether scientists whose early work is highly cited (and hence "reinforced") tend to publish more than others in later years, controlling for levels of initial productivity. Cole and Cole (1967, 1973) and Lightfield (1971) found evidence for such reinforcement. Long (1977), on the other hand, found virtually no effect of early citations on later productivity. Reskin (1977) found strong reinforcement among nonacademic scientists but relatively weak effects among academic scientists.

Even if all the causal links in the cumulative advantage process were confirmed, it would remain an open question whether the end result was an appreciable change in the distribution of productivity and rewards. Allison and Stewart (1974) argued that the hypothesis of cumulative advantage implies that the distribution of productivity and recognition should become more unequal as a cohort of scientists grows older. To test this implication, they took a cross-sectional sample of academic scientists in the United States, divided them into groups by professional age, and within each age group computed the Gini index of inequality for number of publications and number of citations. For both measures they found a strong tendency for inequality to increase linearly with professional age. Faia (1975) used essentially the same method and reached similar conclusions. On the other hand, Mittermeier and Knorr (1979) applied the method to a sample of European scientists and found little evidence for increasing inequality.

Allison and Stewart recognized that their method had one glaring deficiency. Instead of following a single cohort through time, they merely compared scientists of different professional ages. If all cohorts were alike, this would give a true picture of a single cohort followed through time. Yet, the same picture of increasing inequality would appear if each cohort had constant inequality over time but older cohorts were more unequal than younger cohorts. Gaston (1978) attempted to avoid this difficulty by using true cohort data. He calculated the inequality of publications in five-year intervals for several cohorts of British and American scientists and found strong increases in inequality for periods of up to twenty years after the doctorate. Unfortunately, Gaston's results are called into question by his failure to use a measure of inequality that is scale invariant—a fundamental criterion for choosing a measure of inequality (Allison, 1978, [1976] 1980a).

There is also a theoretical weakness in the analysis presented by Allison and Stewart (1974). Increasing inequality is evidence for cumulative advantage only if cumulative advantage implies increasing inequality. Intuitively this would seem to be the case. Allison and Stewart also suggested that mathematical models of cumulative advantage would imply increasing inequality, but they failed to present a rigorous demonstration. Faia (1975) correctly pointed out that the one model they discussed in detail implied increasing variance only if rather arbitrary constraints were imposed on the coefficients.

The aim of this paper is to put the work of Allison and Stewart on a firmer foundation, both theoretically and empirically. In the next section we examine changes in the inequality of publications and citations for true cohorts of chemists and biochemists. Increasing inequality is observed for publications but not, surprisingly, for citations. In a later section we examine the implications of mathematical models of cumulative advantage and again there are surprises. A plausible model does not imply increasing inequality although it does imply increasing variance. Nevertheless, a modification of the model in the direction of greater realism does lead to increasing inequality.

EMPIRICAL EVIDENCE

Using true cohort data, we now examine the hypothesis that inequality of publications and citations increases with professional age. Before describing the data, let us first consider the choice of an inequality measure. The Gini index has been the most popular measure of inequality of scientific productivity (Allison and Stewart, 1975; Cole et al., 1978; Mittermeier and Knorr, 1979), but there are many other candidates. Faia (1975) used the variance and Gaston (1978) used the variance divided by the mean. Hustopecky and Vlachy (1978) listed

thirteen additional measures for possible use in science studies.

Allison (1978, 1980b) and Schwartz and Winship (1979) have argued for scale invariance as a fundamental criterion for choosing an inequality measure. Scale invariant measures are those which remain constant when every individual's score is multiplied by a constant. The Gini index is scale invariant, as are the coefficient of variation and the standard deviation of the logarithms. The variance is not scale invariant, however, and neither is Gaston's measure, the variance divided by the mean. Why scale invariance? If every scientist's citation count were doubled, for example, the ratio of any two scientists' citation counts would remain unchanged. Moreover, each scientist's proportional share of the total pool of citations would remain unchanged. Scale invariant measures of inequality would reflect this constancy but scale dependent measures would not. For example, doubling everyone's score would double Gaston's measure.

The choice among scale invariant measures is less clear-cut. However, Allison (1980b) has suggested a modified version of the coefficient of variation which has several desirable features for studies of scientific productivity. The measure is defined as

$$C = \frac{\sigma^2 - \mu}{\mu^2} \qquad (1)$$

where μ is the mean and σ^2 is the variance. The advantages of the measure C include equal sensitivity to differences at all levels of the distribution, a built-in correction for purely random variation, and a close relationship to the negative binomial distribution, which accurately describes the observed distribution of publications and citations.[1] Accordingly, we have used this measure in both the empirical and mathematical investigations reported here.

DATA

Two different data sets were analyzed. The first is a sample of 239 chemists constructed by Reskin ([1973] 1980, 1977). Using the American Chemical Society's *Directory of Graduate Research,* Reskin drew a systematic sample of those who received doctorates in chemistry from American universities between 1955 and 1961. Counts of publications in each year were taken from *Chemical Abstracts* and *Science Citation Index.* Counts of citations to each

Table 1. Inequality, Means and Standard Deviations of Article Counts

Career Years	Biochemists Cohort I			Biochemists Cohort II			Chemists		
	C	$\overline{X}$	$\hat{\sigma}$	C	$\overline{X}$	$\hat{\sigma}$	C	$\overline{X}$	$\hat{\sigma}$
12–14	1.39	4.2	5.4						
9–11	1.22	4.2	5.1						
6–8	1.09	3.6	4.2	.99	4.1	4.6	1.82	2.5	3.7
3–5	.93	3.0	3.4	.76	3.5	3.6	1.46	1.8	2.6
0–2	.55	2.3	2.3	.41	2.5	2.2	.74	1.7	1.9
N	286			271			239		

chemist's first-authored publications were also taken from *Science Citation Index.*

The second data set has been described in detail by Long (1978) and Long et al. (1979). The sample consisted of 557 male biochemists who received doctorates from American universities in the fiscal years of 1957 or 1958 (Cohort I) and 1962 or 1963 (Cohort II). Publication counts were taken from *Chemical Abstracts* and citation counts were taken from *Science Citation Index.* The citation counts differed from those of Reskin in that citations to all articles were included, whether or not the biochemist was first author.[2] Career history data drawn from *American Men and Women of Science* made it possible to classify each individual's career as predominantly academic or predominantly nonacademic.

RESULTS

Table 1 reports inequality measures for publication counts grouped into three-year intervals.[3] For the chemists and for both cohorts of biochemists there is a strong tendency for inequality to increase with professional age.[4]

---

[1] The reciprocal of C is one of the two parameters of the negative binomial distribution.

[2] For a detailed description of the procedures used to construct the citation counts, see the Appendix in Long (1978). Long et al. (1980) present results on the use of complete counts vs. counts to first-authored publications only.

[3] The inequality measures in this and subsequent tables were estimated by the method of maximum likelihood under the assumption that the counts were drawn from a negative binomial distribution. For a discussion of this procedure, see Allison (1980b, 1980c).

[4] For each cohort, a test of the null hypothesis that there is no increase in inequality was obtained by simply regressing the inequality measures on the number of years since the doctorate. The test statistic was the usual t-statistic for the slope. The results were: biochemists (cohort I), p=.002; biochemists (cohort II), p=.04; chemists, p=.06. We therefore reject the null hypothesis in favor of the alternative that there is a linear increase with time. This test is justified under the null hypothesis that the counts are generated by a multivariate negative binomial distri-

Note that the mean and standard deviation also increase substantially. Table 2 presents similar results obtained after dividing the biochemists into academics and nonacademics. This was done to test Reskin's (1977) suggestion that processes of reinforcement may be stronger in nonacademic settings. Although the overall level of inequality is greater among nonacademics, there is an increase in inequality within each group and no clear evidence that the rate of increase varies across groups.[5]

We turn now to citations. There are several reasons to expect a more rapid increase in inequality for citations than for publications. First, if one assumes that the number of citations per paper is relatively constant for each author, then any increase in inequality of publications should be directly translated into an increase in inequality of citations. Furthermore, the Matthew effect implies that there should also be increasing inequality over time in the average number of citations per paper since the work of prestigious authors is more widely read. The compounding of these two processes should generate rapid increases in inequality of citations.

A second argument is based on the fact that the production of research articles requires inputs that tend to be limited or rigid, even for very prestigious scientists. Thus, there is only so much time that one can spend on research, and there are many facilities over which scientists have little immediate control (e.g., libraries, computers, departmental colleagues) so long as they keep the same job. One would not expect, then, that the distribution of publications would change as easily as the distribution of recognition. With citations, on the other hand, there are no obvious limitations; a scientist's citation count should respond readily to any changes in prestige or recognition.

These expectations are disconfirmed by Table 3, which reports inequality of citations to all previous articles, counted at three-year intervals for biochemists and one-year intervals for chemists.[6] As in Allison and Stewart's

Table 2. Inequality of Biochemists' Article Counts, by Employment Sector

Career Years	Cohort I		Cohort II	
	Academic	Non-academic	Academic	Non-academic
12–14	1.06	1.83		
9–11	.96	1.54		
6–8	.72	1.57	.80	1.24
3–5	.70	1.20	.65	.90
0–2	.38	.70	.30	.53
N	139	147	145	126

(1974) study, there is much greater inequality of citations than of publications. But despite large increases in the mean and the variance, there is no evidence of an increase in inequity—if anything there are slight decreases. Clearly these results do not support Mulkay's (1980:25) claim that

> the marked separation between the scientific elite and the great mass of ordinary scientists is not immediately apparent but occurs gradually as a result of cumulative processes of differentiation.

Table 3 suggests that the differentiation is immediate and relatively constant over at least the first fourteen years of the career.

While these results are unambiguous, it is difficult to explain how there can be an increase in inequality of publications without a corresponding increase in the inequality of citations to those publications. The question is complicated somewhat by the fact that publications were counted in three-year intervals while citations were counted to all previous publications. For the biochemists, we were able to get some additional insight by restricting our counts to citations to articles published in the preceding three years. Results are shown in Table 4. In this table there are consistent increases in inequality for both cohorts.[7] Furthermore, for each interval, the inequality for three-year counts is substantially higher than for the total counts. These results suggest that scientists' older publications are cited with less inequality than their more recent work, and that increasing inequality of citations to recent work is counterbalanced by *decreasing* inequality of citations to older work.[8] Further

---

bution with a constant value of C in all intervals (Allison, 1980c). This implies that the inequality measures have equal variance, have equal correlations with one another, and are approximately normally distributed. Under these conditions, ordinary least squares has its usual optimal properties (Theil, 1971:243).

[5] The greater inequality among nonacademics reflects the fact that they are a very heterogeneous group. This result also explains why in Table 1 there is more inequality among chemists, who are predominantly nonacademic, than among biochemists.

[6] In the case of chemists, it was necessary to divide the sample into two cohorts: those with degree

---

dates of 1956–1958 and those with degrees in 1959–1961. 1955 doctorates were excluded.

[7] For both cohorts, the increases are significant at the .05 level. See note 3 for a discussion of the statistical test.

[8] This explanation is consistent with the strong correlation between publications and citations. Scientists who are highly cited are also likely to publish frequently. It is reasonable to expect that citations

Table 3. Inequality, Means and Standard Deviations of Total Citation Counts

Career Year	Biochemists Cohort I			Biochemists Cohort II			Chemists Cohort I			Chemists Cohort II		
	C	$\overline{X}$	$\hat{\sigma}$	C	$\overline{X}$	$\hat{\sigma}$	C	$\overline{X}$	$\hat{\sigma}$	C	$\overline{X}$	$\hat{\sigma}$
14	2.2	46	70									
13												
12							4.1	7.5	15			
11	2.1	38	55				3.4	7.4	14			
10							3.5	6.6	13			
9							3.5	6.0	11	3.7	7.4	15
8	2.3	36	54	1.8	34	47	4.6	4.1	9	4.1	6.3	13
7										4.6	5.7	12
6										4.3	5.4	12
5	2.0	19	27	1.7	27	35				4.8	4.5	10
4												
3												
2	2.4	9	15	2.1	14	20						
N		286			270			140			84	

Table 4. Inequality of Three-Year Citation Counts for Biochemists

Career Year	Cohort I	Cohort II
14	3.80	
11	3.63	
8	3.58	2.88
5	3.46	2.59
2	3.01	2.39

investigation is needed to confirm these hypotheses.

MODELS OF CUMULATIVE ADVANTAGE

It appears that inequality of publication counts increases with professional age while inequality of citation counts remains relatively constant. What this means for the cumulative advantage hypothesis is not entirely clear. If cumulative advantage implies increasing inequality of productivity and recognition, then the constant inequality of total citation counts is a disconfirmation of the hypothesis. Moreover, while the increasing inequality of publication counts is consistent with the hypothesis, that does not rule out the possibility that other quite different processes may have produced this result.

But does cumulative advantage really imply increasing inequality? To make any headway in answering this question, we believe it is essential to express the hypothesis in mathematical form. As with any attempt to mathematize a

that might have been received by older papers are replaced by citations to recent publications for these productive scientists. Thus inequality in citations would be found primarily among citations to recent work.

verbal theory, however, there are innumerable ways to express the same basic idea, and different formulations may lead to different conclusions. Another perennial dilemma is that models which are simple enough to be mathematically tractable are usually unrealistic in one or more respects. Our strategy is to start with a relatively simple, well-known stochastic model that appears to embody the essential features of cumulative advantage and to examine its implications in detail. We then introduce modifications in order to make the model more realistic.

What are the essential features of cumulative advantage as it pertains to publications and citations? Two elements of the hypothesis stand out as being crucially important: (1) each publication increases a scientist's propensity for future publications; similarly each citation increases a scientist's propensity for future citations; (2) the occurrence of publications or citations is at least partially governed by random processes. Thus, fortuitous publications and citations are converted into lasting advantages.

Price (1976) proposed a stochastic model which possesses these characteristics but, unfortunately, it describes the equilibrium behavior of an entire population rather than the changing distribution for a single cohort. For our purposes, a more useful model is the contagious Poisson process introduced to sociologists by Coleman (1964). In the literature of stochastic processes it is commonly referred to as "linear birth with immigration" (Boswell and Patil, 1970). In the social sciences, the model has been used to represent accident occurrence (Arbous and Kerrich, 1951), purchasing behavior (Coleman, 1964), episodes of racial violence (Spilerman, 1970), and the hospitalization of mental patients

(Eaton, 1974). As a model for scientific productivity, the contagious Poisson process was first proposed by Allison and Stewart (1974) and later examined in greater detail by Allison [1976] 1980a, 1980c) and Rao (1980). Yablonsky (1980) considered a special case.

Although we shall describe the model in terms of publications, its use for citations should be obvious as well. Let X(t) be the cumulative number of articles a scientist has published by time t. We assume that t=0 corresponds to the year in which the doctorate was awarded. Let P(t) denote the instantaneous "probability" that a scientist publishes an article at time t.[9] P(t) is not really a probability since it can be greater than one. Its numerical value may be interpreted as the expected number of papers in an interval one time unit long. Accordingly P(t) can be interpreted as the propensity to publish.

We assume that all scientists start off with the same propensity to publish, denoted by α. We then assume that each new publication increases the propensity to publish by a fixed amount β, which may be interpreted as the rate of cumulative advantage. Thus, after the first publication a scientist's propensity is $\alpha + \beta$, after two publications it is $\alpha + 2\beta$. In general, recalling that X(t) is the cumulative number of papers at time t, we have $P(t) = \alpha + \beta X(t)$.

This completes the statement of the model. Clearly it embodies the notion that each publication increases the likelihood of future publications. And since the number of publications in any interval is a random variable, it also allows for the influence of chance fluctuations. Note, however, that we have said nothing about the reasons why publishing a paper increases the likelihood of future publications. Thus, the model is consistent with a variety of explanations. Our intent is simply to express rigorously the cumulation of advantage so that implications can be drawn about changes in the distribution of publications.

A well-known implication of the model is that X(t), the cumulative number of publications, has a negative binomial distribution.[10]

Less well-known is that the number of publications in any given interval of time also has a negative binomial distribution (Allison, 1980c). It is of some interest that the negative binomial distribution has, in fact, been shown to provide good fits to observed publication and citation distributions (Allison, [1976] 1980a, 1980b, 1980c; Rao, 1980). We shall not pursue this implication further, however, especially since our modified versions of the model do not specify the exact probability distribution of X(t).

What we really want to know is how inequality of publications in some fixed length of time, say a year, varies with the amount of time since the doctorate. Let $X_s(t)$ be the number of publications in an interval of length s which begins at time t. Formally, $X_s(t) = X(t+s) - X(t)$. Allison ([1976] 1980a, 1980c) has shown that the mean and variance of $X_s(t)$ are given by

$$E[X_s(t)] = (e^{\beta s} - 1)e^{\beta t}\alpha/\beta \qquad (2)$$

$$var[X_s(t)] = E[X_s(t)] [(e^{\beta s} - 1)e^{\beta t} + 1] \qquad (3)$$

Hence, both the mean and the variance increase exponentially with time. Using these results, one can easily get an expression for C, the modified coefficient of variation used as the measure of inequality in the empirical study reported above. From (1), (2) and (3), we have

$$C = \frac{var[X_s(t)] - E[X_s(t)]}{E^2[X_s(t)]} = \frac{\beta}{\alpha} \qquad (4)$$

i.e., the degree of inequality is simply the ratio of the rate at which advantage accumulates to the initial propensity to publish, both of which are constant over time.[11] Thus, the model implies that the level of inequality in publications is high when intensity of cumulative advantage is high. A discipline which has strong cumulative advantage will, ceteris paribus, have more inequality in publications than a discipline which has weaker cumulative advantage. The model further implies that the stronger the initial propensity to publish in a discipline, the smaller the inequality of publications. Still, regardless of the values of α and β, inequality will not change over time.

We therefore reach the surprising conclusion that a plausible model of cumulative advantage does not imply increasing inequality. Of

---

[9] P(t) is formally defined as follows. Consider the probability that a scientist publishes one or more articles in the interval (t,t+s) given that he has already published X(t) articles. Divide that probability by s and take the limit as s goes to zero. In symbols, we have

$$P(t) = \lim_{s \to 0} \frac{Prob[X(t+s) - X(t) \geq 1 | X(t)]}{s}$$

[10] There are also other models which imply the negative binomial distribution but which do not embody the ideas of cumulative advantage (Spilerman, 1970; Boswell and Patil, 1970). Price (1976) claimed that the negative binomial distribution is only implied

by models in which a "success" is rewarded and a "failure" is punished. That is incorrect. In the contagious Poisson process discussed here, there is nothing that corresponds to failure of punishment.

[11] It can be shown that the squared unmodified coefficient of variation actually decreases with time, but only to C as an asymptote.

course, the model is unrealistic in many respects, and only one measure of inequality has been considered. Nevertheless, this single counter-example is enough to disprove the claim that increasing inequality is a necessary consequence of cumulative advantage. Hence, any model under consideration must be carefully studied to determine whether or not it implies increasing inequality.

Since we do in fact observe increasing inequality of publications, it is natural to inquire whether the model can be modified to yield that result. We shall examine two possible modifications. An obvious limitation of the model, in its present form, is the assumption that all scientists start out with the same propensity to publish, denoted by α. It is much more realistic to assume that α varies across scientists with a positive mean and variance. For this generalization, Allison ([1976] 1980a) showed that the mean and variance of $X_s(t)$ are given by

$$E[X_s(t)] = (e^{\beta s}-1)e^{\beta t}E(\alpha)/\beta \qquad (5)$$

$$\begin{aligned} \text{var}[X_s(t)] &= (e^{\beta s}-1)^2 e^{2\beta t}\text{var}(\alpha)/\beta^2 \\ &\quad + E[X_s(t)][(e^{\beta s}-1)e^{\beta t}+1] \end{aligned} \qquad (6)$$

both of which increase exponentially with time. These results imply that the modified coefficient of variation is

$$C = \frac{\text{var}(\alpha)}{E^2(\alpha)} + \frac{\beta}{E(\alpha)} \qquad (7)$$

which, again, is a constant over time. Thus, even with initial heterogeneity we get constant inequality. Notice that this constant inequality is the sum of two components: (1) the squared coefficient of variation (i.e., inequality) of the initial propensity to publish and (2) the ratio of β, the rate of cumulative advantage, to the average initial propensity. Hence, the cumulation of advantage does contribute to the level of inequality but not to increases in inequality over time.

Having relaxed the assumption of initial homogeneity, we now turn to the parameter β which has also been assumed constant across individuals. Again, this assumption would seem to be a likely candidate for relaxation. In the eyes of their colleagues, some scientists consistently write papers that are judged outstanding while others write papers which are, at best, mediocre. Surely there is more cumulative advantage in writing good papers than bad ones. A greater degree of cumulative advantage for a scientist would mean a larger value of β for that scientist. We therefore tried to determine the implications of a model in which β varies across individuals but is constant over time for each individual.

Unfortunately, models with heterogeneous β

Table 5. Computer Simulation of Contagious Poisson Process with β Constant vs. β Varying

Career Year	β Varying			β Constant		
	C	$\overline{\text{X}}$	$\hat{\sigma}$	C	$\overline{\text{X}}$	$\hat{\sigma}$
10	7.3	6.3	17.0	1.5	3.3	4.5
9	6.9	4.2	11.0	1.5	2.5	3.4
8	6.0	2.9	7.3	1.4	1.9	2.6
7	5.3	2.0	4.8	1.5	1.4	2.0
6	4.1	1.4	3.0	1.5	1.0	1.6
5	3.7	.9	2.0	1.1	.8	1.2
4	3.2	.6	1.3	1.3	.5	.9
3	2.5	.4	.9	1.4	.4	.8
2	1.7	.3	.7	1.4	.3	.7
1	1.2	.2	.5	2.2	.2	.6

are much less tractable than those with heterogeneous α. If the probability distribution of β is left unspecified, it is impossible to obtain a closed-form expression for our measure of inequality C. And even under specific distributional families for β, we have not been able to obtain closed-form results. Nevertheless, in the Appendix we demonstrate that there is an increase in inequality for all nondegenerate distributions of β that imply a finite mean and variance for $X_s(t)$.[12]

To illustrate this conclusion, we present the results of a computer simulation of the contagious Poisson process with β allowed to vary across individuals. We fixed α at .20 and let β have a uniform distribution between .10 and .50. For a "sample" of 1000 scientists, we generated publication counts for each of the first ten years after the origin of the process. Means, standard deviations, and inequality measures are shown in Table 5. All three measures increased substantially with time. We then repeated the process, this time giving β a fixed value of .30. The results, shown in Table 5, follow the earlier predictions: increases in the mean and standard deviation with no increase in inequality.

Within the framework of the contagious Poisson process, we see that heterogeneity in the rate of cumulative advantage is a sufficient condition for increasing inequality. It may not be a necessary condition, however. It is entirely possible that alternative modifications of the model would yield the same result.

We are also well aware that the model remains unrealistic in many respects. Most obvious is the fact that the model predicts a sustained exponential increase in the number of publications in a given interval of time. While

---

[12] Some distributions of β (e.g., exponential with small values of the parameter) lead to infinite mean and variance of $X_s(t)$.

this may not be too far off the mark in the early stages of a scientist's career, the growth curve must ultimately reach an inflection point (cf., Price, 1963, on science in the aggregate). This suggests a logistic relationship between X(t) and P(t), but such a model is likely to be intractable.

How do these mathematical results relate to the empirical findings reported earlier? The publication data correspond rather well with the predictions based on the contagious Poisson process with heterogeneous β: the mean, variance, and inequality all increase. In the absence of competing hypotheses, this model appears to provide a plausible explanation of those results. More importantly, it is now clear that the failure to observe increasing inequality for citations is not necessarily a disconfirmation of the cumulative advantage hypothesis. There can be cumulative advantage without increasing inequality.

These results are still somewhat unsettling. In arguing for heterogeneous β, we noted that "good" papers are more likely to bring further advantage than "bad" papers. Is it possible that this is not true for citations, that one citation is as good as another in getting more citations? It is useful to consider the process in somewhat greater substantive detail. There is a sense in which citations may directly generate other citations by a process of diffusion. If a paper is cited, those who notice the citation will be more likely to read that paper and cite it in the future. Given that journals and papers vary widely in readership, it would be surprising if there were not substantial heterogeneity in the citation-generating potential of each citation. Nevertheless, it could be that the major portion of that heterogeneity is within rather than between individuals. In other words, there may be little systematic tendency for some scientists to get "powerful" citations and others to get "weak" citations.

It could also be argued that citations, in themselves, have little power in generating future citations. Rather, they may merely measure or reflect a scientist's prestige, which changes according to its own underlying processes of cumulative advantage (or the Matthew effect). One would then have to argue that prestige accumulates at the same rate for all scientists. While this may seem implausible, it is certainly more plausible than concluding that there is no cumulative advantage at all.

CONCLUSION

Although the basic idea of cumulative advantage is quite simple, it is actually a rather complex hypothesis involving numerous causal links organized into several feedback loops. As a consequence, it is difficult to specify just what empirical findings would disconfirm the hypothesis. Moreover, a negative result for a single causal link may seem inconsequential next to so imposing a hypothesis. Allison and Stewart (1974) took the global approach of arguing that cumulative advantage should produce increasing inequality as a cohort of scientists grows older. Their results confirmed that expectation. In this paper we have attempted to replicate their findings using true cohort data. Increasing inequality was confirmed for counts of publications. On the other hand, no increases were observed for counts of citations to all publications. Although there were increases in inequality for citations to recent publications, it seems likely that those increases are largely attributable to the increasing inequality of the number of publications, and are not the result of additional cumulative advantage processes affecting citations.

The apparent disconfirmation for citations is surprising, especially since there are reasons to expect a more rapid increase in inequality for citations than for publications. The pattern of results also suggests that some peculiar changes are occurring in the distributions of citations to older as opposed to more recent publications. Obviously a much more detailed investigation is warranted.

In light of the mathematical results reported here, the constant inequality for citations is not a strong disconfirmation. It is simply not true that models of cumulative advantage must imply increasing inequality. Indeed, a reasonable first approximation to the idea of cumulative advantage was shown to imply constant inequality, even when the model was generalized to allow for initial heterogeneity in the propensity to publish or be cited. A further generalization of the model to allow for heterogeneity in the rate of cumulative advantage does lead to increasing inequality, however.

It should be noted that the empirical results are only for two disciplines, chemistry and biochemistry, and that only the first fourteen years of the career were examined. On the other hand, even though chemistry and biochemistry are closely related in subject matter, there are some striking differences between the two fields. Most notably, while chemistry has a very low percentage of its doctorates entering academic jobs, biochemistry has a high percentage of doctorates not merely in academia but in the most prestigious universities. Furthermore, the fact that the two data sets were collected by independent investigators using different procedures gives us added confidence in the results.

Finally, we note that while interest in cumulative advantage has been largely con-

fined to sociologists of science, there is good reason to expect that similar processes occur wherever rewards are socially distributed (Goode, 1978; Broughton and Mills, 1980). Consider the accumulation of wealth, for example. Folk wisdom tells us that "the rich get richer, while the poor get poorer" and that "it takes money to get money." But there has been relatively little study of whether and how this occurs and how much importance it has for changes in the distribution of wealth (Thurow, 1975 is one exception). We suggest that some of our results for scientific productivity may also apply to the distribution of wealth. For a stochastic model of productivity, we showed that a homogeneous rate of accumulation would not lead to increasing inequality but a heterogeneous rate would produce increasing inequality. This is much easier to show for the effect of interest rates on wealth accumulation. Specifically, it can be shown that a uniform rate of compound interest will not produce increasing inequality even with great heterogeneity in initial investments. On the other hand, heterogeneity in interest rates will lead to increasing inequality of wealth over time. Moreover, the rate of increase in inequality will be faster if initial investments and interest rates are correlated.[13] Of course interest rates are only a small part of the picture of wealth accumulation, but this example does point out the possibility of a general approach to accumulation processes that could be applied to many areas of social life.

APPENDIX

PROOF THAT HETEROGENEITY IN THE RATE OF CUMULATIVE ADVANTAGE IMPLIES INCREASING INEQUALITY

We assume a population of individuals each producing publications (or citations) according to a contagious Poisson process. More specifically, the propensity to publish defined in (2) is given by $P(t) = \alpha + \beta X(t)$ where $X(t)$ is the number of publications between time 0 and time t. We also assume that α is a constant but that β is a random variable (across individuals) with a nonzero variance and a cumulative distribution function $F(.)$. The modified coeffi-

---

cient of variation applied to publications in the interval (t, t+s) is defined as

$$C[X_s(t)] = \frac{\text{var } [X_s(t)] - E[X_s(t)]}{E^2[X_s(t)].} \quad (A1)$$

Increasing inequality means that the derivative of C with respect to t is positive. That is what we shall prove.

Using general formulas for the mean and variance in terms of conditional means and variances (Maddala, 1977), the mean and variance of $X_s(t)$ are

$$E[X_s(t)] = E\{E[X_s(t)|\beta]\}$$

$$\text{var}[X_s(t)] = \text{var}\{E[X_s(t)|\beta]\} + E\{\text{var}[X_s(t)|\beta]\}$$

where $E[X_s(t)|\beta]$ is given by (3) and $\text{var}[X_s(t)|\beta]$ is given by (4). After some simplification, (A1) takes the form

$$C[X_s(t)] = \frac{A + B}{\alpha D^2} - 1 \quad (A2)$$

where

$$A = \int_0^\infty \alpha\beta^{-2}(e^{\beta s}-1)^2 e^{2\beta t} dF(\beta) \quad (A3)$$

$$B = \int_0^\infty \beta^{-1}(e^{\beta s}-1)^2 e^{2\beta t} dF(\beta) \quad (A4)$$

$$D = \int_0^\infty \beta^{-1}(e^{\beta s}-1) e^{\beta t} dF(\beta) \quad (A5)$$

Applying the Mean Value Theorem to expressions for A, B, A+B and D we obtain

$$A = \alpha a^{-2}(e^{as}-1)^2 e^{2at} \quad (A6)$$

$$B = b^{-1}(e^{bs}-1)^2 e^{2bt} \quad (A7)$$

$$A+B = (\alpha+k)k^{-2}(e^{ks}-1)^2 e^{2kt} \quad (A8)$$

$$D = d^{-1}(e^{ds}-1) e^{dt} \quad (A9)$$

where constants a,b,d,k,s, and α are positive. Substitution of (A8) and (A9) into (A2) gives

$$C[X_s(t)] = re^{2t(k-d)} - 1$$

where r is a positive constant. The derivative of $C[X_s(t)]$

$$\frac{dC[X_s(t)]}{dt} = 2r(k-d)e^{2t(k-d)} \quad (A10)$$

is positive when k−d is positive. We will show now that k−d>0, thus completing the proof.

As a preliminary step, we show that $A > \alpha D^2$. Let $g(\beta) = \beta^{-1}(e^{\beta s}-1)e^{\beta t}$. Clearly $A = \alpha E[g^2(\beta)]$ and $D = E[g(\beta)]$. Calling on the well-known fact that the variance is always greater than or equal to zero, we have

$$0 \leqslant \text{var}[g(\beta)] = E[g^2(\beta)] - E^2[g(\beta)] = \alpha^{-1}A - D^2$$

and, hence

$$A \geqslant \alpha D^2$$

This is an equality if $g(\beta)$ is a constant, i.e., if β is a constant. But since we assume that β has a positive variance, equality is disallowed and, hence $A > \alpha D^2$. A straightforward corollary of this result is that a>d.

The final step is to show that k=a and hence (k−d)>0. From (A6), (A7) and (A8) we have

$$\alpha a^{-2}(e^{as}-1)^2 e^{2at} + b^{-1}(e^{bs}-1)^2 e^{2bt} =$$
$$(\alpha+k)k^{-2}(e^{ks}-1)^2 e^{2kt}$$

The equality above holds for arbitrary $t>0$ if and only if $k=a=b$; this can be shown, for example, by expanding both sides of the equality into Taylor series and comparing the coefficients of the same exponents of t. From (A10) we now see that the derivative is positive which completes the proof.

The formula (A10) allows one to compute higher derivatives of C which are also positive. Thus, the inequality increases at an increasing rate. With only minor modifications, this proof holds for the more general case in which α varies across individuals. The same method can also be used to show that the derivative of $C[X(t)]$ with respect to time is positive. Thus, inequality in total career productivity increases with time since Ph.D.

REFERENCES

Allison, Paul D.
{1976} Processes of Stratification in Science.
1980a New York: Arno Press.
1978 "Measures of inequality." American Sociological Review 43:865–80.
1980b "Inequality and scientific productivity." Social Studies of Science 10:163–79.
1980c "Estimation and testing for a Markov model of reinforcement." Sociological Methods and Research 8:434–53.
Allison, Paul D. and John A. Stewart
1974 "Productivity differences among scientists: evidence for accumulative advantage." American Sociological Review 39:596–606.
1975 "Reply to Faia." American Sociological Review 40:829–31.
Arbous, A. G. and J. B. Kerrich
1951 "Accident statistics and the concept of accident proneness." Biometrics 7:340–432.
Boswell, M. T. and G. P. Patil
1970 "Chance mechanisms generating the negative binomial distribution." Pp. 3–22 in G. P. Patil (ed.), Random Counts in Biomedical and Social Sciences. University Park, PA: Pennsylvania State University Press.
Broughton, Walter and Edgar W. Mills, Jr.
1980 "Resource inequality and accumulative advantage: stratification in the ministry." Social Forces 58:1289–1307.
Cole, Jonathan R. and Stephen Cole
1973 Social Stratification in Science. Chicago: University of Chicago Press.
Cole, Stephen
1970 "Professional standing and the reception of scientific discoveries." American Journal of Sociology 76:286–306.
Cole, Stephen and Jonathan R. Cole
1967 "Scientific output and recognition: a study in the operation of the reward system in science." American Sociological Review 32:377–90.
Cole, Stephen, Jonathan R. Cole and Lorraine Dietrich
1978 "Measuring the cognitive state of scientific disciplines." Pp. 209–51 in Yehuda Elkana, Joshua Lederberg, Robert K. Merton, Ar-

nold Thackray, and Harriet Zuckerman (eds.), Toward a Metric of Science. New York: Wiley.
Coleman, James S.
1964 Introduction to Mathematical Sociology. New York: Free Press.
Eaton, William W., Jr.
1974 "Mental hospitalization as a reinforcement process." American Sociological Review 39:252–60.
Faia, Michael
1975 "Productivity among scientists: a replication and elaboration." American Sociological Review 40:825–29.
Gaston, Jerry
1978 The Reward System in British and American Science. New York: Wiley.
Goldstone, Jack A.
1979 "A deductive explanation of the Matthew effect in science." Social Studies of Science 9:385–91.
Goode, William J.
1978 The Celebration of Heroes. Berkeley: University of California Press.
Hagstrom, Warren O.
1965 The Scientific Community. New York: Basic Books.
Hargens, Lowell L., Nicholas C. Mullins and Pamela K. Hecht
1980 "Research areas and stratification processes in science." Social Studies of Science 10:55–74.
Hustopecky, J. and J. Vlachy
1978 "Identifying a set of inequality measures for science studies." Scientometrics 1:85–98.
Lightfield, E. Timothy
1971 "Output and recognition of sociologists." The American Sociologist 6:128–33.
Long, J. Scott
1977 "Productivity and position in the early academic career." Ph.D. dissertation, Cornell University, Ithaca, NY.
1978 "Productivity and academic position in the scientific career." American Sociological Review 43:889–908.
Long, J. Scott, Paul D. Allison and Robert McGinnis
1979 "Entrance into the academic career." American Sociological Review 44:816–31.
Long, J. Scott, Robert McGinnis and Paul D. Allison
1980 "The problem of junior-authored papers in constructing citation counts." Social Studies of Science 10:127–43.
Maddala, G. S.
1977 Econometrics. New York: McGraw-Hill.
Merton, Robert K.
{1942} "The normative structure of science." Re-
1973 printed in Robert K. Merton, The Sociology of Science. Chicago: University of Chicago Press.
1957 "Priorities in scientific discovery: A chapter in the sociology of science." American Sociological Review 22:635–59.
1968 "The Matthew effect in science." Science 159:56–63.
Mittermeier, Roland and Karin D. Knorr
1979 "Scientific productivity and accumulative advantage: a thesis reassessed in the light of

international data." R & D Management 9:235–39.

Mulkay, Michael
1980 "Sociology of science in the West." Current Sociology 28:1–184.

Price, Derek de Solla
1963 Little Science, Big Science. New York: Columbia University Press.
1976 "A general theory of bibliometric and other cumulative advantage processes." Journal of the American Society for Information Science 27:292–306.

Rao, I. K. Ravichandra
1980 "The distribution of scientific productivity and social change." Journal of the American Society for Information Science 31:111–22.

Reskin, Barbara F.
{1973} Sex Differences in the Professional Life 1980 Chances of Chemists. New York: Arno Press.
1977 "Scientific productivity and the reward structure of science." American Sociological Review 42:491–504.

Schwartz, Joseph and Christopher Winship
1979 "The welfare approach to measuring inequality." Pp. 1–36 in Karl F. Schuessler (ed.), Sociological Methodology 1980. San Francisco: Jossey-Bass.

Spilerman, Seymour
1970 "The causes of racial disturbances: a comparison of alternative explanations." American Sociological Review 35:627–49.

Storer, Norman W.
1966 The Social System of Science. New York: Holt, Rinehart & Winston.

Theil, Henri
1971 Principles of Econometrics. New York: Wiley.

Thurow, Lester C.
1975 Generating Inequality. New York: Basic Books.

Turner, Stephen P. and Daryl E. Chubin
1979 "Chance and eminence in science: Ecclesiastes II." Social Science Information 18:437–49.

Yablonsky, A. I.
1980 "On fundamental regularities of the distribution of scientific productivity." Scientometrics 2:3–34.

Zuckerman, Harriet
1977 Scientific Elite. New York: Free Press.

Zuckerman, Harriet and Robert K. Merton
1972 "Age stratification in science." Pp. 292–356 in Matilda W. Riley, Marilyn Johnson and Anne Foner (eds.), A Sociology of Age Stratification, Vol. 3 of Aging and Society. New York: Russell Sage.