THE NUMBERS OF PUBLICATIONS WRITTEN BY BIOLOGISTS

By C. B. WILLIAMS, *Rothamsted Experimental Station*

J. Dufrenoy (1938) has discussed the number of papers published by different authors which were reviewed in single years in the *Review of Applied Mycology*, particularly for the years 1932, 1934 and 1935. His data for the year 1935 are shown in the first column of Table 1 and graphically as a histogram in Fig. 1.

In this year 2229 papers by 1527 authors were reviewed, and it will be seen that the distribution of frequency of authors publishing one, two, three or more papers forms a hollow curve somewhat resembling a hyperbola.

Dufrenoy suggested that if the probability of an author to produce 1 paper was p_1, then the probability to produce 2, 3 or 4 papers would be $(p_1)^2$, $(p_1)^3$, $(p_1)^4$, etc., and so the frequency of publication of papers should be on a geometric series. He suggests for this set of data the series $(3{\cdot}2)^6$, $(3{\cdot}2)^5$, $(3{\cdot}2)^4$, ..., etc., for authors publishing 1, 2, 3, etc., papers. His figures calculated on this basis are shown in the 2nd column of Table 1.

If this series is diagrammatically transformed so that log number of authors is plotted against the number of papers per author it gives a straight line which is shown on the right-hand side of Fig. 1, together with the observed data as crosses. Dufrenoy made this transformation and suggested as a result of inspection that the writers of 1, 2, 3 and 4 papers in a year fell on the theoretical line, but the writers of 5 or more papers must be in a different category, apparently psychologically different. One might be tempted to say that they were more prolific, or more verbose, but we must remember that the data take no notice of the length of the paper, only of the numbers, and it is quite likely that one big paper would be longer than several smaller ones.

Dufrenoy's series was obtained apparently quite empirically, and it will be seen that if the number of authors producing up to 8 papers according to his geometric series are added together they total about 1560 whereas the observed number was only 1527.

A geometric series fitting the data that 2229 papers are produced by 1527 can be readily calculated. In the geometric series n_1, n_1x, n_1x^2, etc., if N is the total number of units (in this case papers) and S the total number of groups (in this case authors) it can readily be shown that $n_1 = S^2/N$ and $x = (N-S)/N$. From this it would seem that the best fit for a geometric series to Dufrenoy's data would be given by a series given by $N_1 = 1046{\cdot}1$ and $x = 0{\cdot}3149$. This series is shown in the 2nd column of Table 1, and, although it corresponds to the total number of authors and papers, it is a poorer fit to the frequency of authors with different numbers of papers. Both series underestimate the number of authors with one paper, overestimate the number with two, and underestimate the number with 5–8 papers. The straight-line transformation of the new geometric series is almost identical with Dufrenoy's series and has not been shown separately on the diagram.

I have recently had occasion to work on several biological problems with a logarithmic series first suggested by Fisher, Corbet & Williams (1943) to be applicable to problems of sampling such as the frequency of species with different numbers of individuals, or genera with different

numbers of species, and it occurred to me that it would be interesting to test this series against the data for publication of papers.

The series is $$n_1, \quad \frac{n_1 x}{2}, \quad \frac{n_1 x^2}{3}, \quad \frac{n_1 x^3}{4}, \quad \text{etc.},$$

where n_1 is the number of groups with one unit and x a constant less than unity.

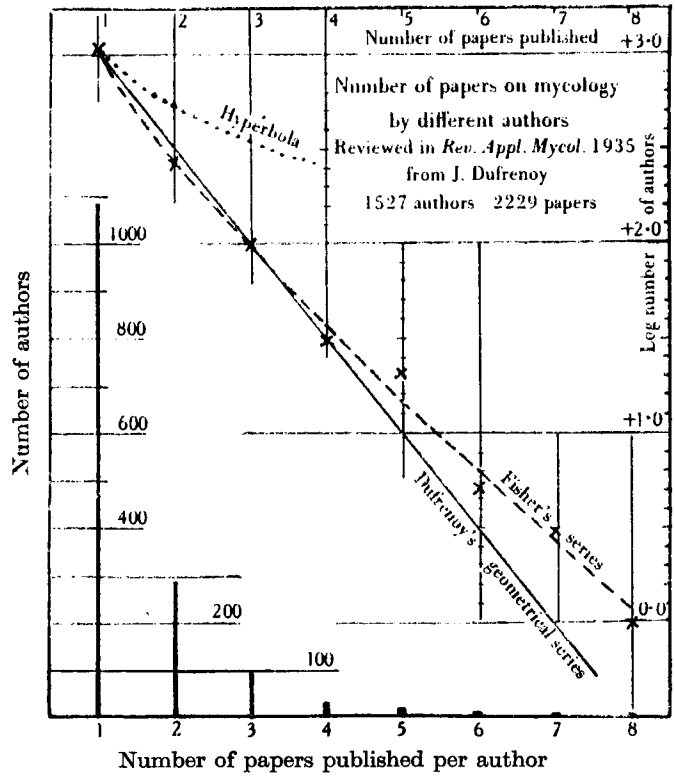

Fig. 1.

Table 1. *Publications in* Review of Applied Mycology, 1935 (2229 *papers by* 1527 *authors*)

No. of papers per author	No. of authors			
	Observed	By Dufrenoy	Calculated geometric series	Logarithmic series
1	1085	1073·7	1046·1	1091·31
2	285	333·5	329·40	278·50
3	96	104·86	103·70	94·77
4	31	32·77	32·66	36·28
5	21	10·24	10·28	14·81
6	5	3·20	3·24	6·30
7	3	1·00	1·02	2·76
8 or more	1	0·31 (for 8 only)	0·32	2·27
Total	1527	1559·58	1526·72	1527·00

The total number of groups S is $\frac{n_1}{x}(-\log \overline{1-x})$, and the total number of units N is $\frac{n_1}{1-x}$, so that if S and N are known, n_1 and x, and hence the whole series, can be calculated. In the present case n_1 is found to equal 1091 and $x = 0{\cdot}5105$. The series is shown in the last column of Table 1,

and diagrammatically, transformed to log number of author's basis, by the broken line in Fig. 1. It will be seen that in every case except that of 4 papers per author the logarithmic series gives a closer approximation to the observed figures than either geometric series; and an inspection of the figures gives no reason to suppose that all points are not related to the same series, i.e. that there is no need to suggest any fundamental break between the writers who publish more and those who publish less, so far as this is determined solely by the number of papers. This seems to be a much more likely conclusion than Dufrenoy's supposed discontinuity.

To get further data on the subject I have tabulated the number of papers published by different authors in the *Review of Applied Entomology*, vol. 1, 1913 and vol. 24, 1936. In the former there were reviewed 656 papers by 411 authors, and in the latter 2379 papers by 1534 authors. The

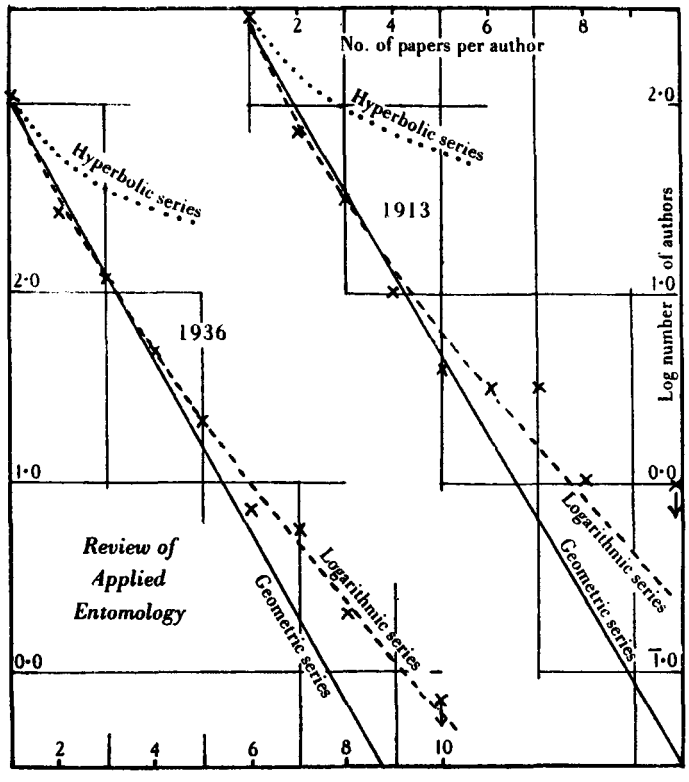

Fig. 2.

observed results, together with numbers calculated from the geometric series and the logarithmic series, are shown in Table 2 and (on a log author basis) in Fig. 2. It will be seen that in both cases the logarithmic series gives a closer fit and there is no evidence of discontinuity.

In each case the calculated number of authors of one paper is slightly below the observed, but the difference is not large, about 4 % in the smaller number of papers in 1913 and only 1·5 % in 1936.

The interpretation of the results does not seem to be easy; as it has been pointed out we are not concerned with the amount written by authors but only with the number of papers into which this is subdivided. Some journals like longer papers, others short papers. Some scientists are encouraged to write many short reports—often repetitions of the same subject—others consider that it is best to produce larger and fuller papers at longer intervals. Dufrenoy's geometric series

was based on the assumption that the probability of a man producing a second paper when he had already produced one was the same as the probability of him producing a third paper when he had already produced two. The observed results show that the numbers of authors producing several papers are higher than the numbers suggested by the geometric series. In other words

Table 2. *Publications in the* Review of Applied Entomology, vol. 1, 1913 (656 *papers by* 411 *authors*) *and* vol. 24, 1936 (2379 *papers by* 1534 *authors*)

No. of papers per author	No. of authors					
	Vol. 1, 1913			Vol. 24, 1936		
	Observed	Calculated		Observed	Calculated	
		Geometric	Logarithmic		Geometric	Logarithmic
1	285	257·50	274	1062	989·1	1046
2	70	96·18	79·79	263	351·3	293·1
3	32	35·92	30·96	120	124·8	109·5
4	10	13·42	13·51	50	44·33	45·99
5	4	5·01	6·29	22	15·75	20·61
6	3	1·87	3·05	7	5·59	9·62
7	3	0·70	1·52	6	1·99	4·62
8	1	0·26	0·78	2	0·71	2·27
9	2	0·10	0·40	—	0·25	1·13
10	1	0·04	0·21	1	0·09	0·57
11	—	—	—	1	0·03	0·29
Total	411	411·00	410·51	1534	1533·94	1533·70

there is a steady—but not apparently discontinuous—tendency for writers of several papers to be more likely to write another. The proportion of authors writing at least 6 papers who write a seventh is higher than the proportion of writers of at least 2 papers who write a third.

It would appear that while individually scientists may still cling to the idea that they are free to publish or not as they wish, yet as a group they are 'but a being that moves in predestinate grooves', and so are subject to mathematical laws.

REFERENCES

J. Dufrenoy (1938). The publishing behavior of biologists. *Quart. Rev. Biol.* **13**, 207.

R. A. Fisher, A. S. Corbet & C. B. Williams (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.* **12**, 42.