Perspectives

Anecdotal, Historical and Critical Commentaries on Genetics

Edited by James F. Crow and William F. Dove

Haldane's Solution of the Luria-Delbrück Distribution

Sahotra Sarkar

Boston Center for the Philosophy and History of Science, Boston University, 745 Commonwealth Avenue, Boston, Massachusetts 02215

IN the 1940s it was still controversial whether bacterial mutants, such as those resistant to a particular phage, arose spontaneously (that is, at random) during the growth of a bacterial culture or in response to the presence of phage in the environment (SARKAR 1991). In order to answer this question, LURIA and DELBRÜCK (1943) devised their famous "fluctuation test." Several bacteria of the same genotype are allowed to grow for many generations in separate test tubes. The contents of these tubes are then separately inoculated onto plates containing phage. Sensitive bacteria die on the plates while resistant ones flourish and produce visible colonies. The number of such colonies, then, is equal to the number of resistant cells present in the generation of bacteria first plated.

Suppose that the resistant bacteria arose through spontaneous mutations during the growth of the bacteria in each of the tubes prior to plating. If there is always a small finite probability of mutation, mutations would occur very early in a few tubes and many mutant cells would be present in a generation much later because of exponential growth of the mutants through ordinary cell division (assuming back-mutations are negligible). In many other tubes there would be no mutations at all or only one in the last few generations. The distribution of mutants observed after plating would reflect this fact and would show considerable variance. Suppose, however, that mutations arose only because of some interaction between the bacteria and the phage. Then, in each of the plates, the only mutants observed would be those produced during the last generation, that is, after plating. The probability of a large number of mutants so produced in only a few tubes is low, and the distribution of observed mutants would have much lower variance than in the former case. This insight into the highly differing variances expected under the two possibilities of mutagenesis lies at the core of the fluctuation test.

The distribution of the number of mutants in the former case has since been routinely called the Luria-Delbrück distribution (SARKAR 1990). In the latter case, it becomes the familiar Poisson distribution with a mean and variance equal to the product of the probability of mutation of a bacterium and the number of bacteria. LURIA and DELBRÜCK's insight was not completely new. As LEDERBERG (1989) has pointed out, YANG and BRUCE WHITE (1934) had made a similar point while studying the resistance of *Vibrio cholerae* to cholera phage. However, LURIA and DELBRÜCK were the first to develop the insight systematically. In order to apply this test to their cultures, they had to investigate the theoretical properties of the Luria-Delbrück distribution. They managed to derive approximate expressions for the mean and the variance of the distribution but, citing "considerable mathematical difficulties," could not devise a way to compute the distribution itself.

Not surprisingly, LURIA and DELBRÜCK's results attracted the attention of J. B. S. HALDANE, then Weldon Professor of Biometry at University College, London. HALDANE found LURIA and DELBRÜCK's conclusion about mutagenesis convincing but set out to "improve" their statistical treatment. What he provided, in the process, was a combinatorial method for obtaining approximate but explicit expressions for the distribution. The results were recorded in a handwritten manuscript, "The Statistical Theory of Bacterial Mutations." HALDANE sent the manuscript to DELBRÜCK in 1946. It was read by LURIA, HERSHEY and others, and DELBRÜCK gave a talk based on it during the Phage Course at Cold Spring Harbor. It was attended, DELBRÜCK (1946) observed, "by those who took the phage course this year and by a few outsiders, mostly people to whom algebra is more

strange than Chinese." DELBRÜCK also had the manuscript typed by a student and both the original and the typed copy were returned to HALDANE. He further circulated a copy to a few individuals, including LEDERBERG, who wrote to HALDANE trying (unsuccessfully) to enlist the latter's help for some related statistical problems (LEDERBERG 1946). This was the only exposure that the manuscript enjoyed. Though HALDANE's was the first solution of the Luria-Delbrück distribution, he never published it; possible reasons for this will be discussed below. The original manuscript is currently preserved in the archives of University College Library among a large collection of HALDANE's papers "rediscovered" there in the early 1980s.

The purpose of this article is to present HALDANE's main results with just enough detail to indicate how they may be obtained. Details of the method and its relation to subsequent work on the Luria-Delbrück distribution will be published separately. Besides their obvious historical interest, HALDANE's calculations are interesting for three other reasons. (i) Ever since the publication of the results of CAIRNS, OVERBAUGH and MILLER (1988) suggesting the possibility of directed mutagenesis in bacteria, the power and limitations of fluctuation analysis have once again become a matter of controversy (see SARKAR 1991 for a review). (ii) No closed analytic solution of the Luria-Delbrück distribution yet exists and, surprisingly, HALDANE's approximate method of calculation comes closest. (iii) It turns out that HALDANE, with his usual prescience, noted some difficulties (including deviations due to various factors) with the use of the distribution (and, consequently, fluctuation analysis) which would only resurface later. In describing HALDANE's results, some trivial algebraic errors have been corrected; any remaining errors are likely to be mine, not his. Some of HALDANE's mathematical claims were left without proof. These have been checked.

In passing it should be noted that R. A. FISHER, too, might have solved the Luria-Delbrück distribution in 1946. CROW (1990) notes that he was convinced by LURIA and DELBRÜCK's experiments but found their mathematical treatment "shoddy and confusing." In 1946 he approached FISHER with the problem. FISHER "leaned back in his chair, thought for perhaps a minute, took a scrap of paper, and wrote a generating function." CROW, not immediately understanding the result, set the paper aside to work on later but lost it along with interest in the problem. Unless that scrap of paper is rediscovered it is improbable that FISHER's possible solution will ever be brought to light.

HALDANE'S RESULTS

Assume, initially, that all the bacteria have grown from a single genotype and that divisions are synchronous, that there are no deaths and no back-mutation, and that mutation occurs only during cell division with only one of the resultant cells being mutant. These are the conditions of an "ideal" experiment. HALDANE treated this case first before studying the effect of four different factors on the resulting distribution. Let n be the number of cell generations after which phage is added. Then $N = 2^n$ is the number of bacteria in a culture. Let m be the probability of mutation during a division, x the number of mutants in a culture of $N = 2^n$ bacteria, and P_x the probability of finding just x mutants. Let $g = mN/2$.

HALDANE argues that, for a successful experiment, N must be so chosen that $g \approx 1$. If it is much smaller, virtually all cultures will contain no mutants. If it is much larger, the number of mutants will be practically uncountable. Now, P_0 is the probability that no mutants are present, that is, no mutations have occurred in $N - 1$ divisions. Thus, for $g \approx 1$,

$$P_0 = (1 - m)^{N-1}$$
$$= e^{-2g}[1 + 2g(1 - g)N^{-1} + O(N^{-2})] \approx e^{-2g}$$

where $O(N^{-2})$ means neglected terms of order N^{-2} or less. When no mutations occurred during the first $n - 2$ generations and one mutation occurred in the last $(n - 1)$ generation, exactly one mutant is present. Thus

$$P_1 = (1 - m)^{N-2}(mN/2)$$
$$= ge^{-2g}[1 + 2g(2 - g)N^{-1} + O(N^{-2})] \approx ge^{-2g}.$$

When $x > 1$, all the different ways in which x mutants could have arisen have to be considered. For $x = 2$, either one mutation occurred two generations ago, or two occurred one generation ago. Thus

$$P_2 = (1-m)^{N-4}(mN/4) + (1-m)^{N-3}m^2(N/2)(N/2-1)$$
$$= (1/2)e^{-2g}\{g[1 + 2g(4-g)N^{-1}]$$
$$+ g^2[1 - 2(2 - 3g + g^2)N^{-1}] + O(N^{-2})\}$$
$$\approx [g(1 + g)/2]e^{-2g}.$$

HALDANE observes that the number of ways corresponds to all partitions of x into powers of 2 (including $1 = 2^0$). The number of such partitions is the coefficient of t^x in the expansion of $1/[(1 - t)(1 - t^2)\cdot(1 - t^4)\ldots]$ in increasing powers of t. Each partition corresponds to a different way of generating x mutants.

HALDANE considers the case $x = 5$ in detail. There are four possible partitions of 5: $1(2^2) + 1(2^0)$; $2(2^1) + 1(2^0)$; $1(2^1) + 3(2^0)$; and $5(2^0)$ (see Figure 1). For each of these mutually exclusive cases, the probability has to be computed and then summed to obtain P_5. Consider the third partition, $1(2^1) + 3(2^0)$. One mutation occurred during the $N/4$ divisions two generations

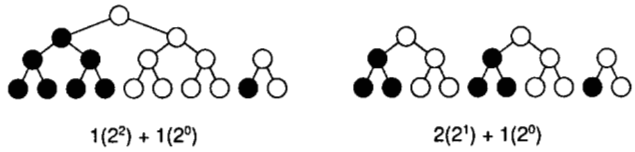

FIGURE 1.—Mutant pedigrees corresponding to partitions of 5 into powers of 2. Black circles represent mutants; white circles, nonmutants. Each of the four figures represents a possible set of pedigrees where only the part of the pedigree relevant to the mutation is shown. The numbers inside parentheses represent the type of mutation classified by when it occurred; the power of 2 is the number of generations ago that the mutation occurred. The numbers outside parentheses are the numbers of such mutations.

ago; its probability is $mN/4$. The other three took place in the nonmutant cells during the last $N/2$ divisions; their probability is $(1/3!)m^3(N/2 - 1)(N/2 - 2)(N/2 - 3)$. Thus the total probability associated with this partition is

$$(1 - m)^{N-6}(mN/24)m^3(N/2 - 1)(N/2 - 2)(N/2 - 3)$$
$$= (1/12)g^4e^{-2g}[1 - 2(6 - 6g + g^2)N^{-1} + O(N^{-2})].$$

Computing probabilities for the other partitions in this fashion,

$$P_5 = (g^5 + 10g^4 + 15g^3 + 30g^2)e^{-2g}/5! + O(N^{-1}).$$

In general, according to HALDANE, a partition of x into $\Sigma_{i=0,k}a(i)2^i$ gives rise to a term $e^{-2g}g^{\Sigma a(i)}/\prod_{i=0,k}a(i)2^{ia(i)}$. These values are accurate only when N is infinite. Consequently, HALDANE notes, the moments of the distribution of x cannot be obtained in this manner because, as N goes to infinity, the moments diverge. For finite N, however, HALDANE calculates the moments by deriving recursion relations for the change of the moments with n (the number of generations). This gives the following expressions for the mean (k_1), the variance (k_2) and the third moment (k_3):

$$k_1 = g(\log N)/(\log 2) + O(N^{-1}),$$
$$k_2 = gN + O(N^{-1}),$$
$$k_3 = gN^2/2 + O(N^{-1}).$$

HALDANE observes that these results show that the mutation rate, m, cannot be reliably obtained from the mean. If values of x are obtained from s populations of N, the variance of the mean is $1/s$ times the variance of x and, therefore, extremely large (when N/s is large, as is usual).

FIGURE 2.—The two forms of the terminal section of pedigrees when cell division is asynchronous. In form 1, both cells (after the first division) have undergone further division; in form 2, only one has.

HALDANE next gives a systematic treatment of the effects on these results of deviations from the conditions of the ideal experiment. Four cases are considered. *First*, he considers the effect of plating only a fraction of the total population, gives expressions for P_x in this case and calculates the moments of the distribution. As previously, he shows that the mean does not give a reliable measure of the mutation rate. *Second*, he considers nonsynchronous divisions. He notes that P_0 and P_1 are unaffected because, in the formation of N bacteria, exactly $N - 1$ divisions must have taken place and the calculation of these two numbers depends on nothing else. P_2, however, is diminished. This is so because the terminal sections of the pedigrees have one of the two forms indicated in Figure 2. Let there be a instances of the first form and b instances of the second. Then $N = 4a + 3b$. Now, from the first form there can be 2 related mutants present if either (but not both) of the products of the first division is a mutant. From the second form, however, the cell that undergoes further division must be mutant to produce 2 related mutants. Nonrelated mutants can occur as before. Hence, approximately,

$$P_2 = [m(a + b/2) + (m^2/2)(N/2)^2](1 - m)^N + O(N^{-1})$$
$$\approx g[(2a + b)/(4a + 3b) + g/2]e^{-2g}.$$

This number is smaller than the case when divisions are synchronous; $(2a + b)/(4a + 3b)$ would be replaced by $1/2$. However, a similar analysis shows that P_3 increases. The general effect is that the distribution becomes smoother, that is, the variance decreases.

Third, HALDANE considers the effect of the death of some bacteria. Let h be the probability that a bacterium dies before it divides and assume that this is constant. If divisions are synchronous, $N = 2^n(1 - h)^n$. The average number of last divisions is $N/2(1 - h)$. Consequently P_0, P_1, etc., all turn out to be smaller than before and the distribution once again becomes smoother. *Fourth*, HALDANE argues that mutants must have a lower growth rate than nonmutants in the original medium because their frequencies are so low. This can happen because of a lower fitness or

a back-mutation rate greater than the forward rate, though this latter possibility would have only a negligible effect on P_x when x is small. Let the average growth rate of mutants be q times that of the normal ($q \leq 1$). Suppose that 2^n bacteria would have resulted in the absence of any mutation. If $x = 0$ or 1, the actual number (N) of bacteria would be unaffected. P_0 will thus be unaltered. However,

$$P_1 = (1 + q/2)ge^{-2g} + O(N^{-1})$$

because only a fraction q of mutants which would otherwise have divided will actually have done so. Thus P_1 will be slightly increased. Similar results hold for the other P^x. In either of these last two situations, Haldane observes, it can be shown that the moments are finite.

DISCUSSION

At the time that HALDANE completed these calculations no solutions of the Luria-Delbrück distribution, approximate or otherwise, were available. Though his correspondence suggests that he ultimately wanted to publish the results, other work, especially in human genetics, as well as departmental reorganization after the War seems to have taken precedence. Meanwhile, in 1948, C. A. COULSON submitted to the *Journal of Genetics*, of which HALDANE was editor, a paper by COULSON and D. E. LEA entitled "The Distribution of the Number of Mutants in Bacterial Populations." LEA, who had collaborated earlier with HALDANE on a mathematical approach to chromosomal rearrangements (LEA and HALDANE 1947), had unexpectedly died in June, 1947. LEA and COULSON had not been able to provide a closed analytic solution of the Luria-Delbrück distribution but, treating the problem using differential equations, had developed an explicit procedure that permitted accurate numerical calculation of the distribution, though only under the conditions of the ideal experiment. HALDANE agreed to print the paper and apparently wanted only to add an appendix containing the argument about asynchronous divisions outlined above (COULSON 1948). However, COULSON was in a hurry and HALDANE appears not to have had time to write this appendix. The paper by LEA and COULSON (1949) was published without it and HALDANE never returned to the problem. Given his general tendency to treat the solution of theoretical problems only to the extent of their numerical evaluation (with a hope of eventually connecting the theory with experiment), HALDANE probably felt that his method had been superseded by that of LEA and COULSON or, at least, that there was not enough important novelty in his method to make it worth pursuing, especially given the wide variety of his interests. In any case, LEA and COULSON's paper has since routinely been considered the first solution of the Luria-Delbrück distribution (SARKAR 1991).

However, HALDANE's method permits writing closed analytic (though only approximate) expressions for P_x, which no other method does to date. Moreover, the combinatorial argument used to obtain these values is elegant and simple, which adds didactic value to this treatment. It is also the only combinatorial treatment of what is essentially a combinatorial problem. Further, no one seems to have discussed the effect of asynchronous division in such explicit detail. The effect of deaths seems also to have been largely ignored. Differential growth rates and the effect of sampling were first treated by ARMITAGE (1952) who gave a much more rigorous treatment of the problem than did LEA and COULSON (1949). ARMITAGE also treated phenotypic lag, the one major confounding factor ignored by HALDANE. That this would cause a problem had already been noticed by DELBRÜCK (1946) in the letter to HALDANE which accompanied the returned manuscript. MANDELBROT (1974) attempted to explain why the various moments of the distribution diverged in the way they did. He also provided an analytic form for the Laplace transform of the Luria-Delbrück distribution, though this is of hardly any use in computation.

STEWART, GORDON and LEVIN (1990) have provided the most extensive procedure to date for the numerical calculation of the Luria-Delbrück distribution under a wide variety of "nonideal" conditions. W. T. MA, G. vH. SANDRI and S. SARKAR (unpublished results) have provided the simplest and most efficient procedure for calculating the Luria-Delbrück distribution and also a new integral representation. Virtually all the new concerns treated comprehensively by STEWART, GORDON and LEVIN (1990) are strikingly present in HALDANE's original attempt. Most importantly, he had first seen that the effect of these other factors would decrease the variance of the distribution. Such a shift alone, therefore, does not demonstrate the occurrence of any directed mutations in bacteria and it is precisely this point, and claims that the late-appearing mutants follow a Poisson distribution and seem to be under genetic control, which lie at the core of the current controversy about directed mutations that began with the publication of the experimental results of the CAIRNS group in 1988 (SARKAR 1990). As in so many other areas, HALDANE had shown remarkable prescience.

Thanks are due to JOHN CAIRNS and FRANK STEWART for comments on an earlier draft of this paper. Moreover, this paper would not have been possible without the assistance of G. FURLONG and the staff of the archives of University College Library, London, R. D. HARVEY, Archivist, John Innes Institute, Norwich, the archival staff of the National Library of Scotland, and the encouragement of JOHN MAYNARD SMITH, GRAEME MITCHISON and NAOMI MITCHISON. The work reported here was supported by an archival research grant from the American Philosophical Society. TRACY

LUBAS prepared the diagrams. It is a pleasure to acknowledge these debts. This is Contribution No. BTBG 90–4 from the Boston Theoretical Biology Group.

LITERATURE CITED

ARMITAGE, P., 1952 The statistical theory of bacterial populations subject to mutation. J. R. Stat. Soc. B **14:** 1–40.

CAIRNS, J., J. OVERBAUGH and S. MILLER, 1988 The origin of mutants. Nature **335:** 142–145.

COULSON, C. A., 1948 Letter to J. B. S. Haldane, 12/14/48. Haldane Archives, National Library of Scotland.

CROW, J. F., 1990 R. A. Fisher, a centennial view. Genetics **124:** 207–211.

DELBRÜCK, M., 1946 Letter to J. B. S. Haldane, 8/20/46. Haldane Papers, University College Library, London.

LEA, D. E., and A. C. COULSON, 1949 The distribution of the numbers of mutants in bacterial populations. J. Genet. **49:** 264–285.

LEA, D. E., and J. B. S. HALDANE, 1947 A mathematical theory of chromosomal rearrangements. J. Genet. **48:** 1–10.

LEDERBERG, J., 1946 Letter to J. B. S. Haldane, 10/12/46. Genetical Society Papers. John Innes Institution, Norwich.

LEDERBERG, J., 1989 Replica plating and indirect selection of bacterial mutants: isolation of preadaptive mutants in bacteria by sib selection. Genetics **121:** 395–399.

LURIA, S. E., and M. DELBRÜCK, 1943 Mutations of bacteria from virus sensitivity to virus resistance. Genetics **28:** 491–511.

MANDELBROT, B., 1974 A population birth-and-mutation process. I. Explicit distributions for the number of mutants in an old culture of bacteria. J. Appl. Prob. **11:** 437–444.

SARKAR, S., 1990 On the possibility of directed mutagenesis in bacteria: statistical analyses and reductionist strategies, pp. 111–124 in *PSA–1990*, Vol. 1, edited by A. FINE, M. FORBES and L. WESSELS. Philosophy of Science Association, East Lansing, Mich.

SARKAR, S., 1991 Lamarck contre Darwin, reduction versus statistics–conceptual aspects of the directed mutagenesis in bacteria controversy, in *Organism and the Origin of Self*, edited by A. I. TAUBER. Kluwer, Dordrecht (in press).

STEWART, F., D. GORDON and B. LEVIN, 1990 Fluctuation analysis: the probability distribution of the number of mutants under different conditions. Genetics **124:** 175–185.

YANG, Y. N., and P. BRUCE WHITE, 1934 Rough variation in *V. cholerae* and its relation to cholera-phage. J. Pathol. Bacteriol. **38:** 187–200.