June 27, 1949.

Dear Kim:

Your MS and letter arrived this morning, and provided a very welcome distraction from the heat. I'm writing you my first thoughts now, mainly because I have the time at hand.

I understood the "Delbruck" equation to refer to a Poisson distribution of infecting particles. Of course it is inapplicable owing to reactivation, but they didn't know about this at the time.

The n-hit theory is, of course, a monstrosity in its simplest terms. However, I suspect that most of its proponents have been thinking in terms of a repair process, or else of the accumulation of a cell poison, small amounts of which can be neutralized. However, suppose that there are relatively stable intermediate states and that a second hit might contribute enough energy to go all the way. I would certainly expect that some single hits would also be effective. I agree that summation without an intensity effect would be very surprising.

I've asked Luria in some detail about his correction for single-infected bacteria, and am convinced that he has the situation well in hand. In practice, he says that the correction is quite negligible. But I think you put the cart before the horse. A correction is made in y so as to take account of that part of w which is due to a residuum of active phage. I agree with you about the possible effect of unequal sensitivity, and raised this at Shelter Island. But Luria convinced me that it would not adequately account for the discrepancies. Don't ask me to repeat his argument! He thinks that most of the discrepancies may be due to some hindrance to unlimited exchanges between many particles, such as Dulbecco analysed.

In the MS I found some phraseologies and developments that were unnecessarily hard. I hope you won't mind some suggestions about rephrasing: take them for what they seem to be worth to you. In general, rather than write $P_{m,i}$, etc., I would use the term P_m. If you write a instead of x you will save the typesetter a bit of grief.

(3) should be $-\frac{dN}{dD} = KN$ (not KD!) N.B. Rate constants are usually written k, not K. Arbitrary!

In (8) the rth term is [illegible]. You have e^{-nkd}. I can't see where that comes from, but I might be wrong. Oh, I see— that's the last term. I think it would be better to write down the general term, which gives a clearer impression of the way the series goes. nCr might be rather large.

(10) might be cleaner as the sum(s): $S = 1 - \sum_{r=1}^{n} C_n^r (1-S)^r$

$\sum C_n^r = 1$

p.6 L-1 "ordinary arithmetic average" might sound less elegant than "mean".

LL. 9-10. ditto for "semilog plot of....". Might be better as "plotting log (-log 1-S) or loglog(1/1-S).

(13)-(14). After a while, I recognized this as the derivation of the Poisson series from Bernouilli's theorem. For the reader to whom the rest of this paper is addressed, I wonder if the derivation should not be a little less abrupt. A statement like that underlined would suffice. Also, I would suggest that many more biologists are acquainted with the expression ${}_nC_r$ than with $\binom{n}{r}$ It took me and several other people here quite a few minutes to recollect this.

Is this derivation of (16) given by Sommermeyer?

Between (15) and (17) the meaning of "g" is a little ambiguous. In order to plot (17) you use as implicitly defined in (16). I think that it would be much cleaner and more direct to state (17) explicitly as:

and then to show that . In this event, g should only be defined as an experimental statistic, and theoretically equal to np. In (14) np should be used for g. After all, "g" can be plotted against n even if the present theory did not hold.

I can't help feeling that there is a certain weakness in your argument for the application of (17) for the case of constant initial n and high dose. These seem to be precisely the same conditions for which (9) is specified. But again, these are also the conditions under which g closely approximates S: That is, you would have , or , which certainly holds even for moderately low S. I think that some clarification is needed that the g plot gives a linear relationship into regions of lower doses (i.e. killing) and I am not convinced that this is so. The log g plot is much more laborious to compute than (9) is, and it should have some concrete justification. I take it that you would base this on the likelihood that if a population initially showed a Poisson distribution, then there would be a linear fit closer to the origin. This is not rigorously shown. That is, while a constant n might be converted to a Poisson distribution of n at low survival, an initial Poisson might conceivably be altered in some different way,

The approach that I had in mind may be a little more rigorous in its requirements, and I suspect is what Delbruck used.

In my letter I was mistaken as to the proper computation of corrected S (S') to fit the truncation. The final expression would be

(16a) $\log -\log (1 - S(1-e^{-r}))$ .....

so I think we both have been wrong. To see this, I have had to rephrase S and S' as fractions to see what they meant.
Let N_0 be the initial observed living cells. N_0' is the fictitious class with no nuclei to start with, and N_s the observed survivors.

Then $S = N_s/N_0$ and $S' = N_s / N_0 + N_0'$.

Then $\frac{S}{S'} = \frac{N_s}{N_0} \cdot \frac{N_0 + N_0'}{N_s} = \frac{N_0 + N_0'}{N_0}$ *

$S' = S\left[\frac{N_s}{N_0 + N_0'}\right] = S\left[1 - e^{-r}\right]$.

If this is correct it should fit the boundary conditions:

$S_0 = 1 - e^{-r}$ $\quad S_\infty = 0$

which it clearly does.

Except for very small r, the correction of the actual mean by the truncation is negligible, but I think that it might appreciably affect the apparent intercept. I am notvsure but that it might not be almost as efficient to use the log S plot.

The main point of your paper seems to be the superiority of the " g plot" as compared with the S plot. You will admit that g is much more laborious: in fact, for low values of S, (i.e. high 1-S) you can't find the logs in tables, and you have to use approximations. I don't think that the paper as it stands is entirely convinci about the superiority of the g plot. It would be most useful if you could give a graphical comparison of g plots using S and S' to show the magnitude of the error involved here for r≤ 4, for example, and also compare with an S plot. The best data to use might be your own distribution of nuclei in Neurospora, corrected by whatever factor you want to use for chromatid multiplicity.

As to sending it to Delbruck, I've had my own fingers burned when I asked for a discussion from him. I really think that you might get a more constructive response from Dulbecco and Luria, both of whom are at Caltech this summer, and who will both certainly give it a sympathetic response. In fact, Luria could probably help to get it published quickly in J. Bact. I have the feeling that it's too long for PNAS (limit 6 type pages). In any event, it should be seen at Caltech (e.g. in view of Novitski's blurb) but I don't see any point in asking anyone outside of Columbia to transmit it to PNAS if you want it there. Curt Stern has told me that he is anxious for more manuscripts on Gen. of Micr. and that you get get reasonably prompt publication now in Genetics.

Have a good summer: we wish we were there withbyou—

Joshua