SUBJECT: Fish Notes
TO : CO, SSA, War Dept.

Report #F 14
3 April 1944

1. The formulae used for setting X patterns can be developed in a number of different ways. The following exposition is in some respects a simplification but in others and amplification. I have tried to translate from British terminology to our own and to fill in gaps which appeared in the development. These gaps could, no doubt, be easily bridged by a professional mathematician but I have attempted to set the matter forth in a way that can be understood by anyone with a knowledge of algebra and the elements of probability theory. I have, however, omitted some steps in the argument which might be necessary for a complete and formal derivation from the theorems of inverse probability. The use of ΔD_{ij} is purely illustrative. The formulae are applicable to any relation in which λ has an appreciable value. The symbol λ, incidentally, is frequently used here to indicate the "bulge", or excess over random, of any variable. It has a different meaning in this report, for example than in my report #F 5.

2.

n = message length
p = probability of dot in ΔD_{ij}
$\frac{1}{2}$ = random probability of dot

If, at a given ΔX_{ij} setting, s dots are yielded in ΔD_{ij} and if, a priori, this setting is just as likely to be right as wrong, the odds that it is right, or the factor in favor of its being right,

$$F = \left(\frac{p}{\frac{1}{2}}\right)^s \left(\frac{1-p}{\frac{1}{2}}\right)^{n-s}$$

$$= 2^n p^s (1-p)^{n-s}$$

$$\log F = n \log 2 + s \log p + n \log (1-p) - s \log (1-p)$$

$$= s\left\{\log p - \log (1-p)\right\} + n \log 2(1-p)$$

The theoretical score, s, = np. Therefore

$$\frac{\log F}{n} = p\left\{\log p - \log (1-p)\right\} + \log 2(1-p)$$

$$= p\left\{\log 2p - \log 2(1-p)\right\} + \log 2(1-p)$$

$$= \frac{2p \log 2p + 2(1-p) \log 2(1-p)}{2}$$

Let λ = the "bulge" in p; that is, $p = \frac{1}{2} + \lambda$. Then

$$\frac{\log F}{n} = \frac{(1+2\lambda) \log (1+2\lambda) + (1-2\lambda) \log (1-2\lambda)}{2}$$

Report #F 14 (continued)

If we use logs to base e the foregoing can be approximated and simplified through the logarithmic series

$$\log_e (1+x) = x - x^2/2 + x^3/3 - x^4/4 \cdots$$

$$\frac{\log_e F}{n} = \frac{(1+2\lambda)\ (2 - 4\lambda^2/2 + 8\lambda^3/3 \cdots) + (1-2\lambda)(-2\lambda - 4\lambda^2/2 - 8\lambda^3/3 \cdots)}{2}$$

$$= \frac{4\lambda^2 + 8\lambda^4/3 \cdots}{2}$$

Since λ is small

$$\log_e F \doteq 2\lambda^2 n \doteq 2(s/n - \tfrac{1}{2})^2 n$$

3. The British use the term "centibans" to measure the scores of settings. A "centiban" is 1/100th of a theoretical "ban" which is simply the log (to base 10) of the odds.

$$\log_e 2 = \log_{10} .8686$$

Therefore, $\log_{10} F = .8686\,\lambda^2 n$

Number of centibans $= 86.86\lambda^2 n$

How big a score in centibans should be aimed at in setting patterns? In the first place the a priori odds that a given setting is wrong vary from 597:1 (for X_{45}) to 1270:1 (for X_{12}). The British average this at 1000:1 for a 2 wheel run in order to avoid a separate set of computations for each pair of wheels. Furthermore, they adopt 10:1 as the standard of odds required of a favorable result. Therefore F must be 10,000 in order to make it 10:1 that the setting is correct notwithstanding a 1000:1 a priori probability of incorrectness. Since the log of F to base 10 must be 4, a score of 400 centibans is necessary. The length of message necessary to achieve this score is easily determined.

$$400 = 86.86\lambda^2 n$$

$$n = 4.6/\lambda^2$$

400 centibans is the average for a message of this length. Therefore, we should get a score of 400 centibans or greater just half the time. Similarly,

$$\lambda = \sqrt{4.6/n}$$

From this we can, if we know the characteristics of the motor or Ψ patterns, compute the bulge in ΔP_{11} necessary to achieve the required degree of success with a message of length n.

Report #F 14 (continued)

4. The foregoing represents a means of comparing the correct setting with random settings. However, many of the wrong settings are not truly random because of the self match of the patterns which is inevitable owing to the limitations on their composition. The British call this self match "slide". A "slide" of 6 means that there is a strong self match at a displacement of 6. This is a factor that cannot be lost sight of in appraising the reliability of results and the true significance of scores. Some theory has been developed on this phase of the problem which I will try to go into at a subsequent date.

5. Bream was the only circuit solved in March. Strenuous efforts have been and are being made to determine the Jellyfish X patterns statistically. The circuit is Paris - Berlin and the traffic is believed to be very important. Messages with the same QEP numbers will not read in depth so presumably Jellyfish has some auto-key element. Vesy likely it is the P_5 limitation but this is not certain. The circuit started early in 1944 and the auto-key feature has always existed. Since the traffic has never been read nothing is known about its plain text characteristics. Virtually the entire attention of Mr. Newman's section has been devoted to operational work on Bream and research on Jellyfish. A slight amount of work has been done by Maj. Tester's section on Sticklebach (Koenigsberg - Poland) which uses the X_2 limitation but not the P_5. No other circuit is being worked on at all.

6. I have asked Mr. Newman to collect some suitable messages which his group is unable to work on, to be sent to A. H. with the hope that our people can find the X wheels statistically. He is very anxious to have us try. I don't know whether the messages will be old Jellyfish or some other circuit but will send them on as quickly as possible.

Walter J. Fried
Capt. Signal Corps