ANNEX to #F 71

Elementary Screed on Delta D Counts and Colossus Runs

1. Delta D Counts.

There can be no universal criteria for determining the correctness of a de-Chi along the lines of insisting on a certain specified letters having a high frequency in Delta D.

Delta D = Delta P + Delta Psi' Whereas Delta Psi' has a count that is reasonably settled (given "a"), Delta P will vary strongly from message
to message.

The factors determining the shape of the Delta D count are, of course, many: there are a few that we normally take into consideration and that we believe to dominate the count.

(i) Doodling Habits. Some operators nearly always double a 5, others a 5 and an 8; some double the 9 between words. These vices, of course, bump up / in Delta P, Some tend to put 89 between words, or to indulge in strings of 898989 : this sends up the count of 5 in Delta P.

(ii) The proportion of punctuation. (A lot of punctuation is likely to be due to many abbreviations). A high frequency of full stops sends up the counts of U and 5 and (to a lesser extent) of A or O or both.

(iii) The order of 8 and 9. Most operators use 89 to return to letter-shift : some use 98. Since the commonest letter before 89 (or 98) is M this influences the frequencies of A and O.

(iv) The proportion of plain German. The Delta P counts of plain German differ notably from those arising from punctuation, figures and abbreviations etc. The typical Delta P count of message strong in plain German is rich in J, F, 3, fairly rich in S and U.

A given Delta P count will be largely interpretable in terms of these factors. The characteristics carry over also into the Delta D count with the bulges, of course, much feebler and with smaller antipodal bulges on the opposites and near-opposite, (e.g. if the Delta P is strong in 5' s, the Delta D wll in consequence be strengthened in 9's and to a lesser extent in /, S, I, N and H). One should therefore
? line missing ?

We are liable to get fantastic counts on / or 5 (Doodlers): or to messages with a lot both of 5 and U (Punctuation): or to get messages with 3, J, F as the highest letters (plain German).

Linear combinations of the factors, of course, give linear combinations of the effects; it is however important to keep in mind the constituent parts.

2. Details of Techniques.

We normally get the first four X (Chi) - settings in pairs, - from a run for X1 and X2 followed by a run for X3 and X4 or for X4 and X5. Any run in which two X's are to be set is called a long run : any run for only one X is called a short run.

Long runs. A long run is liable to produce a purely random score in the neighbourhood of 3.2 sigma ; scores of 3.5 to 4 sigma are far from being rarities. It is important to improve judgment by scoring the best scores of runs; a chart exists, making use of the second best competitor, giving the decibanage in favour of the best score. As a rough rule of thumb, it has not in the past been our practice to accept as a basis for further runs a X1 X2 result below 3.5 sigma.

If a X1 X2 result of probability p is accepted. it can be much improved if a good result is obtained on a run using these settings as a basis. Suppose that, on the (3+4)x/1x2x run, a result is obtained whose probability (lusing the chart) works out to be q. We can assume that the probability of the X34 settings being right and the X12 wrong is negligible. (This can be checked by doing the (3+4)x count at the place in question).

we now compare 4 theories (a) X1234 all right.
(b) X12 right 34 wrong.
(c) X12 wrong 34 right.
(d) X1234 all wrong.

The odds of these are pq : p(1-q) : 0 : (1-p) (1-q) [see Appendix] In fact the odds in favour of (a) are pq : 1-q - or, the odds on the

the first to give the odds of the first theory.

Example. On the X12 run the settings 16, 09, give a score of 3.8 sigma with a rival of 3.4 sigma. Thus the settings 16, 09 are 2 decibans up : or p = 8/13, 10log10 p = -2.

(a) A (3+4)x/1x2x run gives X34 settings of 13, 20 with a score of 4.5 sigma and the nearest rival is 3.5 sigma : the decibanage in favour (from the chart) is 15. Therefore 10log10(p*q/(1-q)) = 13, in fact the whole story is 13 decibans up or 20 to 1 on.

(b) On the other hand, if the (3+4)x/1x2x run gives only a 4 sigma reading with a 3.5 sigma rival, its decibanage is 6 up and the whole story is then only 4 decibans up, or 5 to 2 on.

These methods enable one to estimate the odds on the combined settings of X1234.

On the other hand, the X34 run may be a disappointment, and if it is quite flat, this is evidence against the original X12 settings. The factor lost by the X12 settings is a function of the probability of the correct answer, being below the observed highest answer. This will vary with the link and the run used : experience shows that on a Berlin end or a Paris Jelly or Salonica Cod, a reputable score occurs about half the time - say a score of 3.5 sigma or over. So the original hypothesis loses a factor of 2 if the (3+4)x/1x2x run has no score above 3.5 sigma. On Rome Bream or Zagreb Gurnard on the run 4=5=/1=2, the factor is probably in the region of 5.

Two modifying considerations.

(i) The length of the message. The chart is based on a standard length message and on the probability of the right score being in the various ranges. For a very short message the right score cannot be expected to be so high and the chart is over severe; conversly for long messages it is lenient. A standard length might be taken to be 3600.

(ii) Slides. The highest score may well be correct on one wheel and wrong on the other. Wheels often have good slides - (i.e. two different

?possible line missing?
higher than the correct position itself. This means that these are serious rival hypotheses - such as that X134 are correct and X2 is a good slide of the correct position. This is a type of hypothesis worth bearing in mind if, after four X's have been set with apparent satisfaction, the last X is obstinate.

The hypothesis can be checked by doing short runs for each separately on letters determined for the purpose from their strength in the 16 - letter count.

Short Runs. On a short run the random scores are liable to be as high as 2 sigma and we are not accustomed to take much interest in an unsupported score of under 2.5 sigma. In average circumstances such a score is liable to make the corresponding setting about evens. Anything above 3 sigma is worth taking pretty seriously : 4 sigma must on no account be ignored : (it may be only a good slide or even a good anti-slide, but the setting eventually selected must account for the occurrance of such a score).

The deciban chart also applies to short runs but is less reliable : the expected bulges are extreemly variable. However, no harm is likely to arise if it is used faithfully provided that the flnal 32-letter count is always looked at carefully. All exceptional features or the count should be explained in terms of the factors mentioned in the first part of the screed (allowing of course for reasonable random deviation).

The final 32-letter count. There are two main reasons for a wrong de-X being sent over with a confident comment.

(i) A feeble story being over-estimated at each stage. For instance a 3.6 sigma followed by a 3.5 sigma, that sets its last X with a 2.8 sigma. Such a count is unlikely to be more than 3 decibans up in all, and should not be described as "all certain" although there may be no one impulse that strikes the eye as being doubtful. The comment "all certain" is defined in this Section to mean "Each separate setting is better than 10:1 on".

(ii) Having one setting wrong. This can arise from a deceiving slide at

?one line missing?
at each impulse separately when the 32-letter count has been finished.

In examining for a single impulse being wrong, we look at pairs of letters which differ only in that impulse.

First examine the last impulse only; use the additional evidence ot the pairs not used in that X, comparing the split on those pairs with the corresponding splits on the average counts. (these counts are available at Colossus). Particular note should be taken of splits in the wrong direction or failures to achieve an expected large split.

These can often be explained. Suppose X3 is being exmIned If there are a high number of 5's and many fewer 8's, we can expoect a high number of 5's in Delta P to against Total Motor x's and these will give rise to 9's in Delta D about twice as often as they will to /'s. Consequently it is not a matter for surprise to flnd in such a count 9's above /'s. In the same count there may be a similar, rather smaller effect causing N to be as high as - or higher than - 3.

Similarly it the Delta D is rich in /, we will often find 8's above 5's and, as a similar secondary effect, V may come above C and K above J.

These examples apply to a scrutiny of the X3 setting. similarly, if the X5 setting Is being investigated, in a message strong in /'s it is not to be wondered at it if X comes up to F, or even Q to U.

A rule of thumb may be suggested but it is extremely crude :-

(a) If X3 is in question. Assume the following 16 lettes are expected to be better than their X3 pairs - /,H,O,3,R,G,P,I,U,Q,5,J,F,X,Y,S. Count in the in the supposed Delta D how many of these letters are better than their pairs (not equal to). If the number is lower than 11, that is a fact that requires explanation.

(In a sample run Bream, Gurnard, Stickleback, Jellyfish and Cod of 50 messages completely set - (an unsatisfactory sample for statistics, theoretically) - the numbers having from 8 to 16 of the letters going the

?possible missing line?
so the 11-rule would admit ??% of this sample).

(b) If X5 is in question. Assume the following 12 letters are expected to be above their X5- pairs :-

/,9,M,G,L,P,A,U,5,8,D,P. If in the supposed Delta D fewer than 8 go the right way, again there is reason for suspicion.

(The corresponding results were :-

5	6	7	8	9	10	11	12
1	3	6	12	12	12	6	4

The 8-rule would admit ?30%? of the sample).

Two cautions should be given
(i) Many of the counts included in the sample which were rejected by the rules were unquestionably essentially correct on examination of a songle pair. This rule is far from being sufficient grounds for final rejection.

(ii). Often the doubt about a setting arises from a good slide; in this case it can hardly be expected that the rule will discriminate. The method for this case will be described later.

When the last impulse set has been found satisfactory, look at the other impulses. Take some dominant letter (say 5) and compare its score, with the scores of letters differing from it in a single impulse. For instance, G may seem unduly high and this casts suspicion on X1, if, however, U is high and I is on the low side, it will probably be permissible to accept X1. In case of doubt short runs can be done.

Decibanning of settings when rivals are good slides.
Suppose we have 3 rival settings of X3, - 01, 03 and 05 - and that there is a good slide of 2 on Delta X3, Suppose also that we are sure one is correct, then a fairly accurate assessment can be made.

We do all three 32-letter counts; they will be very similar and they can be used as a sample to guess the approximate score in the right place of each letter. (Say, take the average of the three scores). If the estimated number of J's is 143 and of K's is 106, then a theory about ?possible missing line?

10(log10 143 - log10 106) gives the decibanage in favour of a theory for each occurences of a J. Each pair of letters can be decibanned in exactly the same way. Now we can take (say) the count for X3 = 01 as standard, and the other two counts can be decibanned up or down according to the excess or defect of the good letters over those in the 01 count.

A warning should be given about the decibanning of pairs of letters that go the wrong way; unless this feature can be satisfactorily explained, the scores should not be accepted at their face value. No satisfactory way of allowing for this is to hand : it probably be safest to leave out such pairs.

When the total decibannages have been added up, the relative probabilities of the 3 settings can be irnmediately stated.

Summary of Techniques.

(i) For long runs. - 3.5 sigma is borderline, 4 sigma is liable to be about 3:1 on, 4.5 sigma is generably reliable.

(ii) For short runs, -2.5 sigma is borderline, 3 sigma is liable to be about 3:1 on, 3.5 sigma is generally reliable.

(iii) If a long run gives settings with probability P, and these are used to get further settings, the odds in favour of the further settings (worked out from the chart) ?comment? P gives the odds in favour of the wholwestory.

If a run based on other settings fails, those settings lose a factor of 2.

(iv) Final counts must always be looked over
(a) to check individual impulses.
(b) to see that there are no abnormal and inexplicable bulges.

For X3, 11 of the 16 pairs should go the right way
For X5, 8 of the selected pairs should go the right way.
Slide rivals can be decibanned from the full counts.

Runs in common use.

(a) Break-ins	1+2/	Alll links : weaker as the pure German proportion rises.
Note: 3+4/x is crossed out and (3+4)x/ pencilled in	3+4/x	Fair except when / is common, as in RB, z Gd., Sttick. Stronger on punctuation as pure German rises.
	4+5/	Moderate an all links : best on /-type messages (RB, Z Gd, Stick). Bad on pure German.
	2+5/	Moderate to feeble on all links not bad in pure German, bad in punctuation.
	2+4/	Only used in /-type links.

On Colossus I the normaI break-In is to do simultaneously 1+2/ (also on multiple test), 4+5/, 2+5/ and either (3+4)x/ or 2+4/ : the latter only for /-type links.

On Colossus II (and later Colossi) we do break-ins singly and on multiple testing. I +2/ - is the invarlable first choice, (3+4)x/ the second except on /-type links.

The set total is normally taken as 2.5 sigma.

(b) Short runs after Break-in		
(i) Given 1 and 2	4=/1=2) 5=/1=2)	Only on /-type links.
	3x/1x 2	Moderate on all links :~ stronger with increase of pure German.
	3/12	Good on /-type links.
(ii) Given 3 and 4	5+/4,(3+4)x) 5/3x 4)	Not very fully tested as yet.
	1+/3	Good on pure German, rarely successful.
	2x/3 4x	Dubious.

?possible line missing?
Colossus II it hardly pays to do them : it is generally as quick to do a long run and certainly more powerful.

The set total is normally taken at 1.5 sigma.

(e) Long runs after Breah-in.

(i) Given I and 2 (3+4)x/1x2x) Good except on /-type links
(3+4)x/(1+2))

(4+5)/1x2x Good on most links

4=5/=1=2 Extreemly powerful on /-type links, and good on all. Hopeless on pure German.

(ii) Given 3 and 4 1x2x/(3+4)x Good on punctuation..
(1+2)/3,4x Good on language on punctuation.

(lii) Given 4 and 5 1=2/=4=5 Good on /-type links.
1x2x/(4+5) Good on punctuation.

(d) Short runs for final

(i) Given 1,2,3,4. UUU Nearly always good.
555 Good except on pure German.
888 Fair
111) Occasionally good.
999)

333) Good in language and so
FFF) worth trying if 555 fails.
GGG)

There is a good composite run 5+/4, 1x2x : if this Is used, care must be taken to see which of the other runs done are independent of this.

The set total is normally taken at 2.5 sigma.

(ii) Given 1,2,4 and 5

555) One of these is usually good,
///) rarely both; the rule is that, if xx?xx considerably exceeds ..?.., 555 is likely to be the better, otherwise ///.
UUU neary always good.

JJJ	Good on language (bad if / is frequent).
333	Good on language (bad if 5 is frequent).
999	Good if 5 is frequent.

There is a good composite run (3x/4)x 2x ; if used, care must be taken to see which of the other runs done independent of this.

On Colossus II (and later Colossi) it is posible, by use of the double negation, to run an accumulation of good letters all added together. Opinions about the advisability of this are divided.

The set total is normally taken at 1.5 sigma, but for difficult messages 1 sigma (or even 0.5 sigma) is sometimes used, thus enables one to detect settings which often score but never spectacularly.
24/7/44

Appendix p1

Note on Odds of the result of a second run, based on a previous run.

A run has been done for X12 giving a result. "This result is correct" we call A.

A second run, using the result of the X12 run, has been done for X34, say (3+4)x/1x2x, giving a certain set of scores. The fact that these scores were obtained we call z, and the statement "the highest score in this run gives the correct setting" we call B.

P(X,Y) means "the probability of X given Y",

O(X,Y) means "the odds on X given Y",

nA means "not A"

We. make use of the theorem

(1) P(X,Y).P(Y) = P(X&Y)

We use the deciban chart after the first run to flnd P(A) = p, and after the second run to find O(B, Z&A) = o.
(AES Note: I have used "d" in place of lower case Greek Delta, "e" in place of epsilon and "m" in place of)

Let P(z,A) = d(1+o), P(z & nB, nA) = d' , P(z, nA) = d'(1+)

Then, by a slight extension of (1),

P(Z & B, A) = od
P(Z & nB, A) = d

and

P(Z & B, nA) = ed'
P(Z & nB, nA) = d'

Hence

P(Z) = P(Z & nA) + P(Z & A)
= P(Z, nA).P(A) + P(Z ,A)P(A)
= d(1+e)(1-p) + d(1+o)p

and

P(A & B,Z) = P(A & B & Z)/P(Z)
= P(Z & B, A).P(A) / P(Z)
= pod/P(Z)

or

O(A & B, Z) = pod/P(Z) - pod
= pod/d'(1+e)(1-p) + dp

i.e. putting d'/d = 1-m

(a) $e = P(Z \& B, nA)/P(Z \& nB, nA) = P(B, Z \& nA)/P(nB, Z \& nA)$
$= O(B, Z \& nA)$

The odds on the highest score of a run (3+4)x/1x2x where the wrong X1 and X2 have been used cannot be very high. There are two reasons why they might be greater than 1/ 29x26.

(i) The score for B on (3+4)x may be so enormous that it shews up even on the random sample of the text given by the wrong X12 setting. This can be tested by counting (3+4)x.
(ii) A may be wrong, but a good slide of correct settings. This is a more troublesome possibility.

(b) $1-m = d'/d = P(Z \& nB, nA)/P(Z \& nB, A)$

This is the ratio of chances of getting the observed scores with the highest one wrong, starting from wrong and right X12 settings

We shall generally be considering cases where Z contains a pretty good score, so that we shall not often be dealing with Z such that P(Z, nA)>P(Z,A). Let us then also consider It unlikely that P(Z & nB, nA)> P(Z & nB, A).

On the other hand we may get P(Z & nB, nA) <P(Z & nB, A) through the possibility of B's being wrong but a good slide of the correct settings.

These considerations shew that e and m are small compared with 1, therefore we can take it that

$O(A \& B, Z)$ approx $=P(A). O(B,Z \& A) = po.$

If e is not negligible, it is because X1 or X2 is only a good slide. In that case po is an <u>overestimate</u>.

if m is not negligible it is (probaby) because of slides in X3 and X4 improving P(Z & nB, A) at the expense of P(Z & nB, nA). In fact if there are strong rivals on tte second run m is not negligible and po <u>is an understatement</u>.