GRANT EOAR 63-93

DD 8

AD668207

CONSTANT ERRORS IN WEIGHT JUDGEMENTS AS A FUNCTION OF THE SIZE OF THE DIFFERENTIAL THRESHOLD.

By Helen E. Ross.

From the Psychological Laboratory, University of Cambridge.

Constant errors in psychophysical procedures are known to increase in magnitude with the intensity of the standard, and with the temporal or spatial separation of the standard and comparison stimuli. It is often assumed that this is due to increasing "adaptation" of the effective standard. It is argued here that the increase is due to a statistical artefact, since the measures of constant error and of differential threshold normally used are not independent, so that any factor which increases the difficulty of discrimination will increase the magnitude of the constant error.

It is shown that for differential thresholds ranging from 1-25 gm the absolute size of the constant error increases with the size of the threshold, but the direction of the error is variable and remains unexplained.

27

1. INTRODUCTION.

In psychophysical experiments it is often found that a difference arises between the "point of subjective equality" (that stimulus which is judged equal to the standard) and the physical value of the standard. This difference is called the "constant error" (CE). CEs most frequently arise in comparison methods, when the standard is distinguishable from the comparison stimulus because it is always presented first (or second), or on the left (or right). With temporal separation the CE is usually called the "time error" and with spatial separation the "space error".

The present investigation was undertaken as a result of a previous experiment by Ross and Gregory (1963) in which it was shown that the size-weight illusion influenced the differential threshold (DL): when the physical weight was kept constant apparently heavy weights gave rise to higher DLs than apparently light weights. In this experiment the Method of Constant Stimuli was used, with a standard of 200 gm., and the comparison weights were lifted either simultaneously or before or after the standard. It was also observed that larger CEs were associated with the apparently heavier weights, and that there was a correlation between the DL and the absolute size of CE. (The modulus sign, /CE/, will be used to indicate that only the size and not the direction of CE is being considered). This correlation is to be expected since the measures of CE and DL normally used are not statistically independent. (A proof of this is given in the appendix). However, some authors have considered the CE to be an independent variable due to some kind of adaptation, and if this were so the adaptation might affect the DL and thus give rise to a correlation between /CE/ and DL apart from the statistical artefact.

Kohler (1923) suggested that the "time error" was due to a "fading trace" left by the first stimulus with which the second stimulus was compared and overestimated. However, all explanations based on "fading" assume that the trace does not merely become disturbed with time, but represents a signal of lower intensity: they therefore fail to explain the fact that the direction of the error is sometimes to underestimate the second weight (Woodrow, 1935). Moreover, time errors also occur in judgements of quality (as opposed to quantity or intensity) where "fading" could not apply. For example, Koester and Schoenfeld (1946) found time errors in pitch discrimination.

These authors unfortunately give no data on the effect of the time interval upon the DL, so we cannot be certain that they were increasing together. However, Piaget and Lambercier (1943) include a measure of the DL in their work on the "space error" in the comparison of lines: in this case the DL increases with the spatial separation of the lines, and adequately accounts for the increase in the CE.

According to Weber's Law the DL increases with the intensity of the standard, and there is considerable evidence showing that the s... of the CE also increases. It is usually found that comparison stimuli are overestimated in relation to intense standards and underestimated in relation to weak ones, but that the absolute size of the error increases with the intensity of the standard. Hollingworth(1910) proposed the theory of the "central tendency of the standard" to explain both the size and the direction of the CE: he supposed that the value of the standard in some way gravitated towards the mean of the series under consideration, with the result that there was no CE with standards at the mean but large CEs in opposite directions with standards at the extremes of the series. It is not well established that there are large errors of underestimation at low intensities, though over-estimation at high intensities has been shown in several modalities: for example, by Holling worth (1910) for judgements of length, Woodrow (1933) for weight, and Needham (1935) for loudness. The errors are normally asymmetrical, with considerably more overestimation than underestimation, so Woodrow suggested that in addition to "central tendency" there was a continual tendency for the value of the standard to sink with time. Helson(1947) proposed a similar but more complex "adaptation level" theory, in which the value of the effective standard gravitated towards a weighted geometric mean of the series, with a negative constant added for the "time error". These theories are unsatisfactory because they assume the occurrence of a negative "time error" of constant size without proposing any method of measuring it independently of "central tendency", or allowing for the correlation between the DL and the magnitude of any measure of constant error.

It seems advisable to treat the direction and the magnitude of the CE as two separate questions. Most of the "central tendency" data are presented as group means taking the direction of error into account; since individual subjects vary in the direction of their error, the underlying Weber relationship is often disguised, especially with weak standards when subjects may change from overestimation to underestimation. The only data inconsistent with Weber's Law were those of Needham (1935) on loudness discrimination, where individuals gave increasingly large CEs at low intensities. However, we cannot be certain that this result conflicts with Weber's Law, since no data on the DL were included, and an unstable measure of CE was used. This measure was D%, which is found by taking the difference between the number of judgements "heavier" and "Lighter", divided by the total number of judgements and expressed as a percentage. Since "equal" judgements are allowed, the size of D% will vary inversely with the number of such judgements, in addition to increasing with the DL.

Since it was not clear from the existing data whether the CE was proportional to the standard at low intensities, it was decided to perform an experiment on weight discrimination similar to the previous experiment of Ross and Gregory but using much lighter weights. Only one time order was used, as this was not intended as a study of all the different variables which might affect the direction and size of the error.

2. METHOD.

Eight subjects were used. Seven were undergraduates and graduates of both sexes, aged between 19 and 29 years, and one a boy of 13.

Three sets of weights were used, each set consisting of one standard and nine comparison weights. The weights were made from tins containing lead shot and candle wax, and fitted with wire loops on the lids with which to lift them. All the tins were painted black. One set was made from Nescafé tins 3" high by 3" diameter, one from Kodak film tins 1¾" by 1¼", and one from Ferrania film tins 2¼" by ¾". The smallest tins (Ferrania) ranged from 6 to 14 gm in 1gm. intervals, with a standard of 10 gm. equal to the middle weight. The medium sized(Kodak) and large sized (Nescafé) tins both ranged from 65 to 85 gm in 2.5 gm intervals, with a standard of 75 gm. The only difference between these two sets of weights was size, with the effect that the smaller set felt heavier (size-weight illusion). The purpose of this was to attempt to obtain results at a lower intensity comparable to the previous results of Ross and Gregory. In that experiment the size-weight illusion was shown to influence the DL and CE, and in this experiment it was hoped to discover whether the same effect occurred at low intensities.

The subject and the experimentor were seated at a table with a screen between them. Each subject was tested on all three sets of weights, the order of sets being varied for different subjects, but the same procedure used for each set. The subject was instructed to use his preferred hand throughout the experiment, and to lift first the standard and then the comparison weight. The comparison weights were hidden behind the screen, and were presented to the subject in a predetermined random order, each weight appearing ten times. The subject lifted the weights up and down smartly by the loop once, and then said whether he thought the comparison weight was heavier or lighter than the standard. "Equal" judgements were not allowed, and if uncertain he was asked to guess.

3. RESULTS.

The standard deviation (SD) and constant error (CE) for each subject and each set of weights were calculated according to the method of averaged z-scores, as described by Woodworth and Schlosberg (1955, p.205) In this method the probability is calculated of judging each comparison weight heavier than the standard, and each probability score is transformed into a z-score (normal deviate score). It is assumed that the z-scores are linearly related to the weight of the stimulus, and a straight line is fitted through two points which are found by averaging the z scores for the lighter half and the heavier half of the comparison weights separately. The SD is a measure of the slope of this line and is found by calculating the number of gms on the x axis corresponding to one standard deviation on the y-axis. The CE is given by the difference between the weight of the standard and the weight of the comparison stimulus judged heavier on 50% of the trials. The response bias (RB), or the degree of overestimation or underestimation of the standard, was also calculated. This measure is given by the probability of judging the comparison weight as heavier when it is equal to to the standard, subtracting 0.5. Thus CE and RB are functionally

related variables, since CE is the intercept on the x-axis and RB the intercept on the y-axis of the same response curve at p = 50 and x = S (standard) respectively.
Fig.1. shows typical response curves for subjects with little error and with positive and negative errors. The y-axis here shows probability scores in place of the corresponding z-scores, but so spaced as to give the same straight-line transformation for the response curves.

As a measure of correlation Kendall's τ was calculated for each condition separately between SD and the absolute size of CE and RB (/CE/,/RB/) The correlations are shown in Table 1. In every condition there was, as expected, a strong correlation between /CE/ and /RB/ (p 0.001); but the correlation between /CE/ and SD was never significant, and that between SD and /RB/ was even smaller and in some cases negative. However, when there is a sufficiently large variation in SD the correlation between SD and /CE/ becomes significant. This is shown graphically in Fig.2, where the results for individual subjects are shown from all the experiments in which the standard was lifted first. The variation in the range of weights used produces a large variation in SD, and the correlation is significant at the 1% level on a graphical test of monotonic association (Quenouille, 1959, No: 28) The same results are shown in Fig.3, except that /RB/ is used instead of /CE/. A similar trend is present as in Fig.2, but the correlation just fails to be significant at the 5% level on the same test.

Since /RB/ is less affected by SD than is /CE/ it is used in Fig.4 to show the distributions of errors obtained under different conditions. The first three pairs of histograms are derived from the data of the previous experiment, and the last three histograms from this experiment. The sign of RB is the opposite of CE: thus overestimators have a positive RB but negative CE, and vice versa. The median SD for each group is shown, and it can be seen that in the previous experiment there are both larger SDs and larger RBs than in the present experiment. There are also changes in the direction of RB under different conditions. In general there is overestimation with the standard lifted first, underestimation with the comparison stimulus lifted first, and slight overestimation with simultaneous lifting. There is also a tendency for the size-weight illusion to affect the direction of the errors: when the small (Kodak) tins are compared with the large (Nescafé) tins it can be seen that more subjects show errors in the predominant direction, producing a shift in the means of the distributions. This difference only reached a significant level in the previous experiment in the case where the standard was lifted first ($p=0.01$ Wilcoxon test). In the present experiment no significant difference in either SD or CE was found between these two sizes of tins, perhaps because too few subjects were used.

4. DISCUSSION.

The method of averaged z-scores may result in slightly inaccurate estimates of SD and CE when the CE is large, since the method is based on the assumption that the responses form a normal cumulative probability curve, with the point of subjective equality (PSE) in the centre. For large CEs the PSE is not central, and the distributions do not appear to be normal. The inaccuracy of this method is small compared with the observed correlation between /CE/ and SD, but it suggests that the assumption of normality should not be made in calculating the results, Thurstone (1928) suggested that in view of Weber's Law the response distributions should be plotted on a logarithmic scale. This has little effect over the small range of weights used in these experiments. Moreover, this would only account for positive CEs, whereas CEs can occur in either direction producing response curves which are symmetrically skewed: overestimators produce curves which rise more steeply over the heavier than the lighter weights, and the converse is true for underestimators. (See Fig.1)

Regardless of the question of the most suitable method of calculating the constant error and differential threshold, the pattern of results is fairly clear: /CE/ is a function of both /RB/ and SD, and in any set of data it correlates most highly with the factor showing the larger fluctuations. In these experiments SD was varied by using different subjects, weights, sizes of tin and methods of presentation. In any one condition the variation in SD among different subjects was not great enough to give rise to a significant correlation, but when the intensity of the weights was also varied the correlation became significant. /RB/ is not significantly correlated with SD, though with large variations in SD it shows a small positive correlation due to the non-independence of the measures. In future investigations of the factors affecting the size of the constant error it would be advisable to use /RB/ as a measure, and to reduce variation in SD as far as possible.

Changes in the direction of CE are much harder to explain than changes in size. The direction is known to be influenced by the temporal or spatial order of stimuli, the size of the temporal or spatial interval, the intensity of the standard, and by practice (Needham, 1934). The CE is notoriously unstable, and can be affected by the form of question and response (Brown 1910). In the present experiment there was a change to more positive CEs with the larger (and apparently lighter) Nescafé tins as opposed to the Kodak tins, but this trend was not continued with the small and very light Ferrania tins. This does not necessarily contradict Woodrow's results, as he obtained positive errors for light standards only when the weight of the standard was varied throughout the experiment instead of being kept constant for each series. Moreover, he used the same size of tin for all the weights, which would enhance the lightness of the light weights (size-weight illusion). Perhaps some kind of "contrast" is necessary (whether due to the size-weight illusion or to other weights in the series) before positive errors are obtained with the standard presented first.

Explanations in terms of peripheral adaptation (e.g. Helson, 1948) hardly seem appropriate, since CEs occur in such a variety of modalities and conditions of presentation that it is difficult to postulate suitable adaptation mechanisms for all of these (e.g. the simultaneous presentation of lines of different lengths)) Whether adaptation is thought to be peripheral or central it is implied that the CE is an independent variable due to some perceptual change. If this were the case there should be a negative correlation between SD and /RB/, since the PSE is regarded as fixed, and a flatter slope of the response curve must make a smaller intercept on the y-axis, and thus a smaller /RB/ (see Appdndix). The results do not support this conclusion, but suggest instead that /CE/ is a function of /RB/, with SD acting as a scaling factor: we should regard /RB/ as fixed, so that a flatter slope of the response curve will make a larger intercept on the x-axis, and thus a larger /CE/.

It has been pointed out by Irwin (1958) and Luce (1959, pp 30-4) that subjects may show a preference for one response which is unrelated to the ability to discriminate or solve problems. A response bias of this sort might be considered similar to the gambling strategy described by Bruner, Goodnow and Austin (1956), when subjects were asked to sort different pictures of aeroplanes into two categories. With no knowledge of results they tended to pick un an irrelevant cue and use it persistently with little regard for the outcome; while with some knowledge of results they made an effort to categorize correctly, and tended to guess in the proportion in which they thought the two categories were occuring. Psychophysical tasks are similar in that they demand categorization with incomplete knowledge of results. This suggests that CEs might represent a form of gambling in favour of one response when uncertain: the gamble could be completely haphazard or it could be in favour of the category believed to be more frequent or more desirable. CEs do not seem to be haphazard since there is some regularity in their direction; nor do most subjects believe that the categories are unequally distributed - indeed, they are frequently disturbed by the apparent inequality introduced by their CEs; nor is there any obvious reason why one category should be regarded as more desirable. Possibly one response is favoured by verbal conditioning: we say "heavier or lighter" "louder or softer" rather than the other way round, and this may produce a preference for the first response. (R.J.Audley, personal communication, 1962). Unfortunately this provides no explanation for errors in the reverse direction.

Explanations in terms of response bias assume that the subject is guessing rather than discriminating. In the statistical decision theory of Tanner and Swets (1954) it is assumed that there is almost no limit to the subject's ability to discriminate between stimuli of different intensities, but that the perceived intensity of a given stimulus is variable. The subject's task is to decide whether a given perceived intensity is more likely to have arisen from one or another physical stimulus, In making this decision he is influenced not only by the relative probabilities involved but by the different rewards and penalties for success or failure in either decision. In view of these he adopts a "criterion" or"cut-off' point" for sorting the perceived stimuli into two categories. In the case of differential thresholds the standard is given as the criterion, and there seems to be no reason why CEs should arise unless the rewards and penalties in the two categories are sufficiently unequal to make the subject adopt some other criterion. As already stated, it is difficult to see why this should be so. However, experiments in which the probabilities and values were deliberately altered might reveal how far these factors influence CEs.

Decision theory is probably too precise in supposing that subjects are always discriminating rather than guessing. Senders and Sowards (1952) found that subjects tended to respond in short repetitive runs rather than random sequences. This suggests that they are not always maintaining a constant decision rule, since each response is partly determined by the previous responses. Luce (1963) quotes evidence in favour of a discrimination threshold and suggests a threshold model incorporating response biasses. In practice it is very difficult to distinguish between guessing and discriminating at threshold level, though there is a theoretical difference. Analysis of response sequences may reveal more about the nature of response biasses than can be found from cumulative response distributions.

This work was carried out while supported by the U.S.A.F. European Office of Aerospace Research, Grant A.F. EOAR 63-93. I am grateful to Mr. R. L. Gregory, Miss Violet Cane and Mr. D.R.J. Laming for much valuable discussion and advice, and to Professor O.L. Zangwill for providing facilities.

TABLE 1.

The Kendall rank correlation coefficient between SD, /CE/ and /RB/ under different conditions. The correlation between /CE/ and /RB/ is always significant ($p < 0.001$ 2 tails), but the correlation between SD and /CE/ or /RB/ is never significant.

Condition	Weights.	Weight of standard.	No. of subjects.	SD & /CE/	SD & /RB/	/CE/ & /RB/
Standard first	Small	200 gm	16	0.167	0.017	0.834
Standard first	Large	200 gm	16	0.050	0.017	0.860
Comparison first	Small	200 gm	16	0.227	-0.113	0.700
Comparison first	Large	200 gm	16	0.133	-0.017	0.683
Simultaneous	Small	200 gm	20	0.100	-0.116	0.759
Simultaneous	Large	200 gm	20	0.290	0.147	0.854
Standard first	Small	75 gm	8	0.097	0.032	0.387
Standard first	Large	75 gm	8	0.097	0.0	0.353
Standard first	Verysmall	10 gm	8	0.161	0.129	0.404

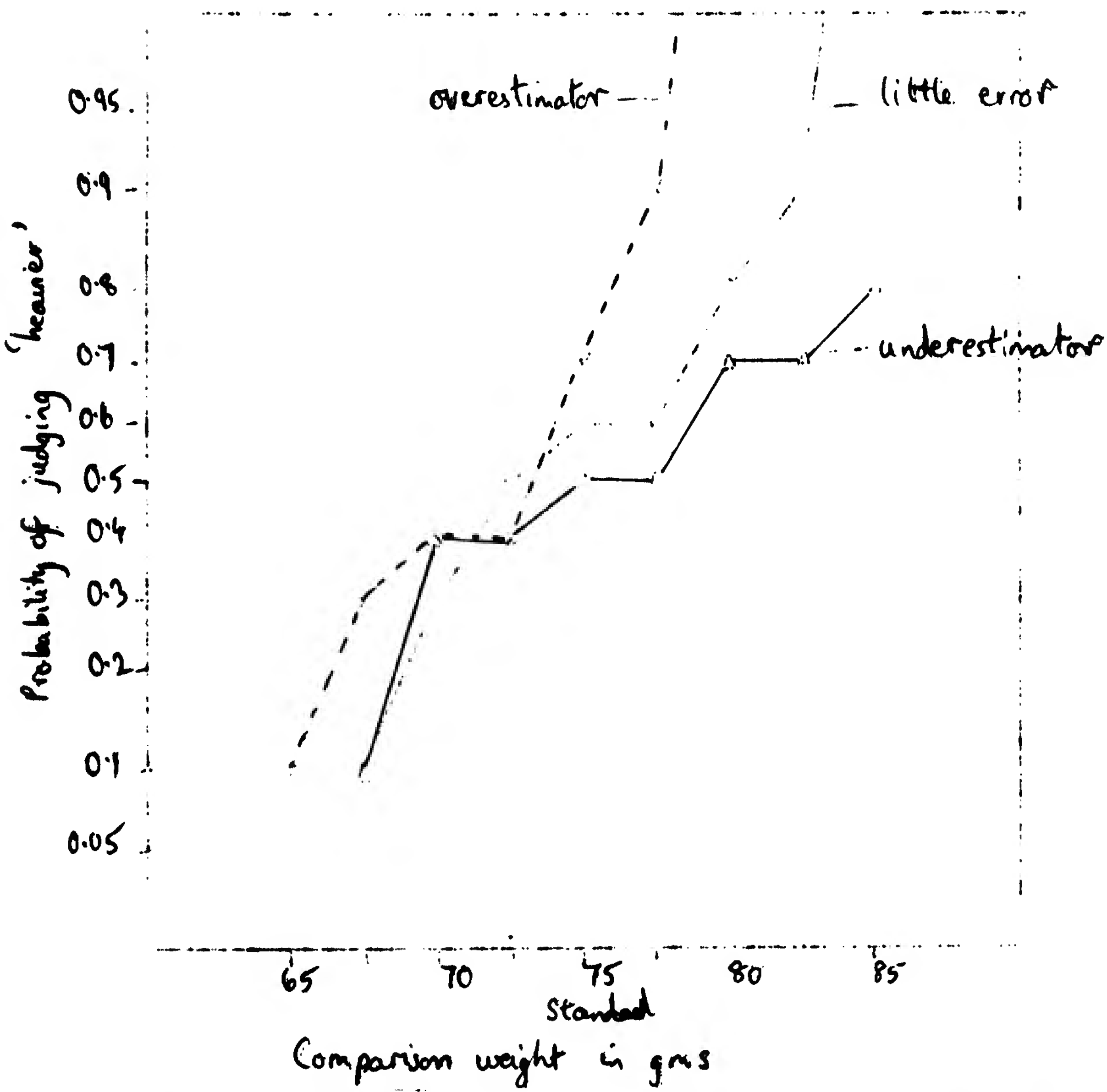


Fig.1.Typical response curves from subjects with positive and negative errors, and with little error. The curves are plotted on probability paper, so that a normal ogive should appear as a straight line.

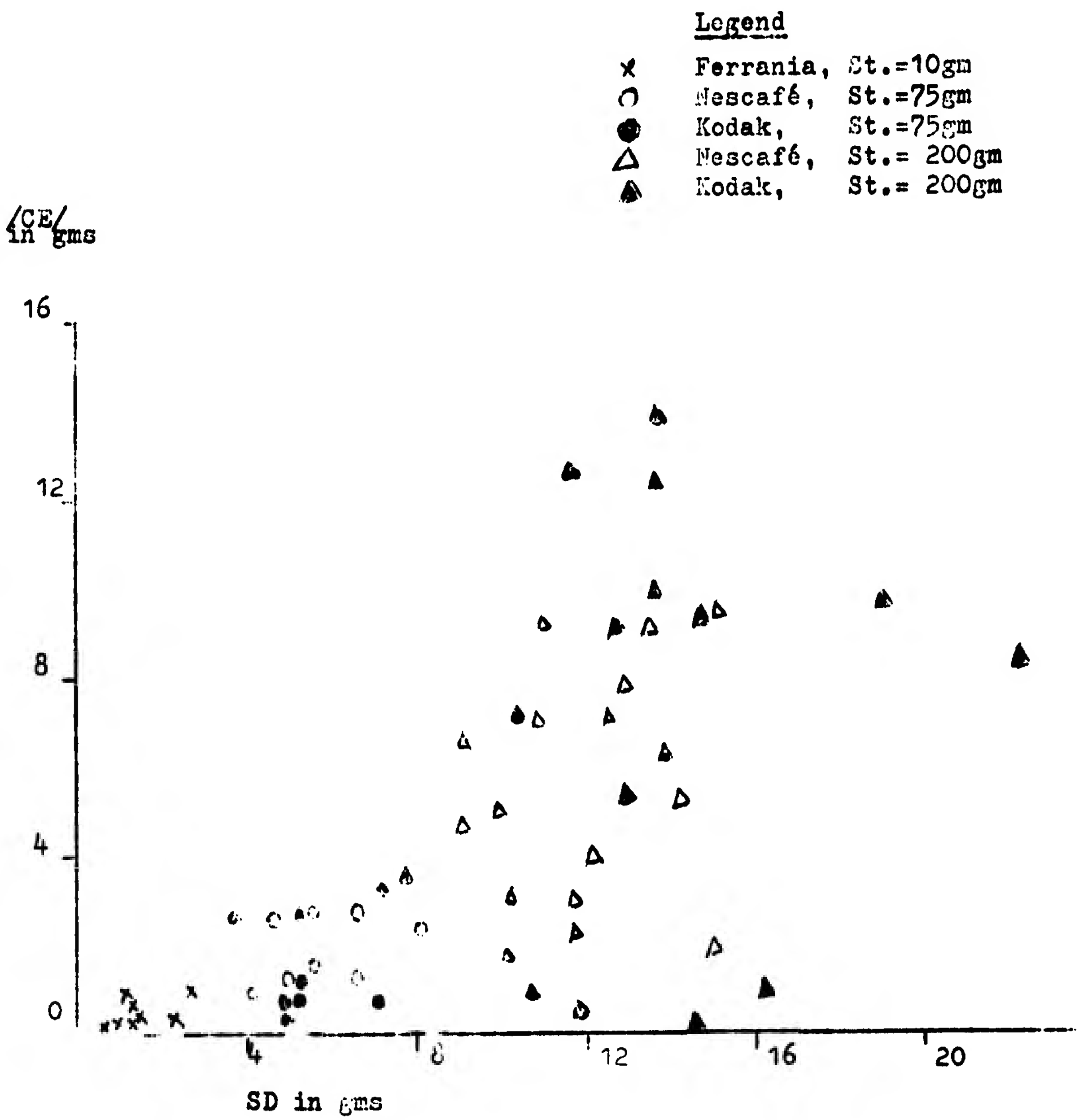


Fig.2. Scatter diagram showing a correlation between Standard Deviation and the absolute magnitude of the Constant Error. The data are taken from all the conditions in this and the previous experiment where the standard was presented first.

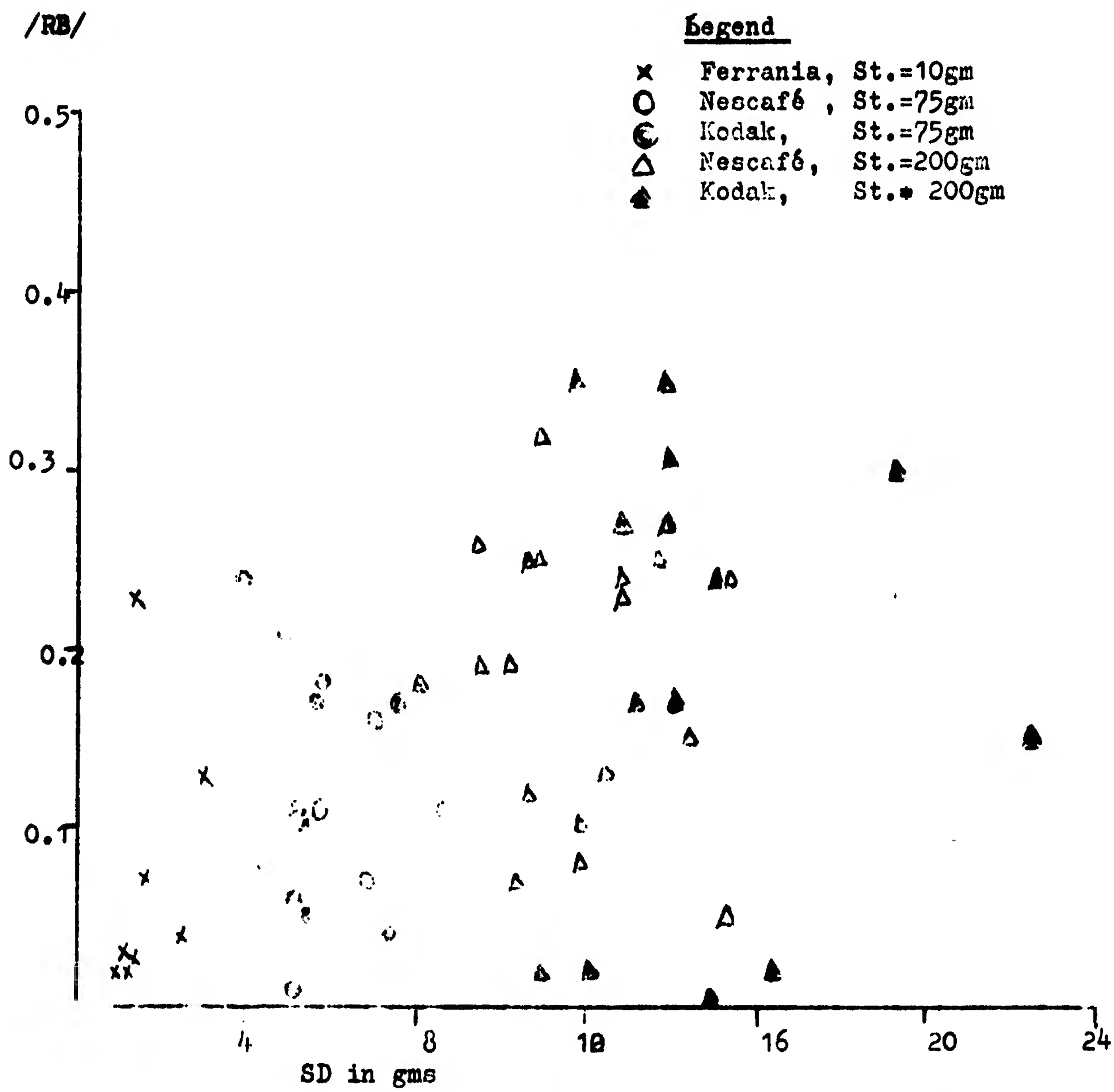


Fig.3 Scatter diagram of the same data as in Fig.2, except that the absolute magnitude of the Response Bias is shown instead of the Constant Error. The correlation between SD and /RB/ is not significant.

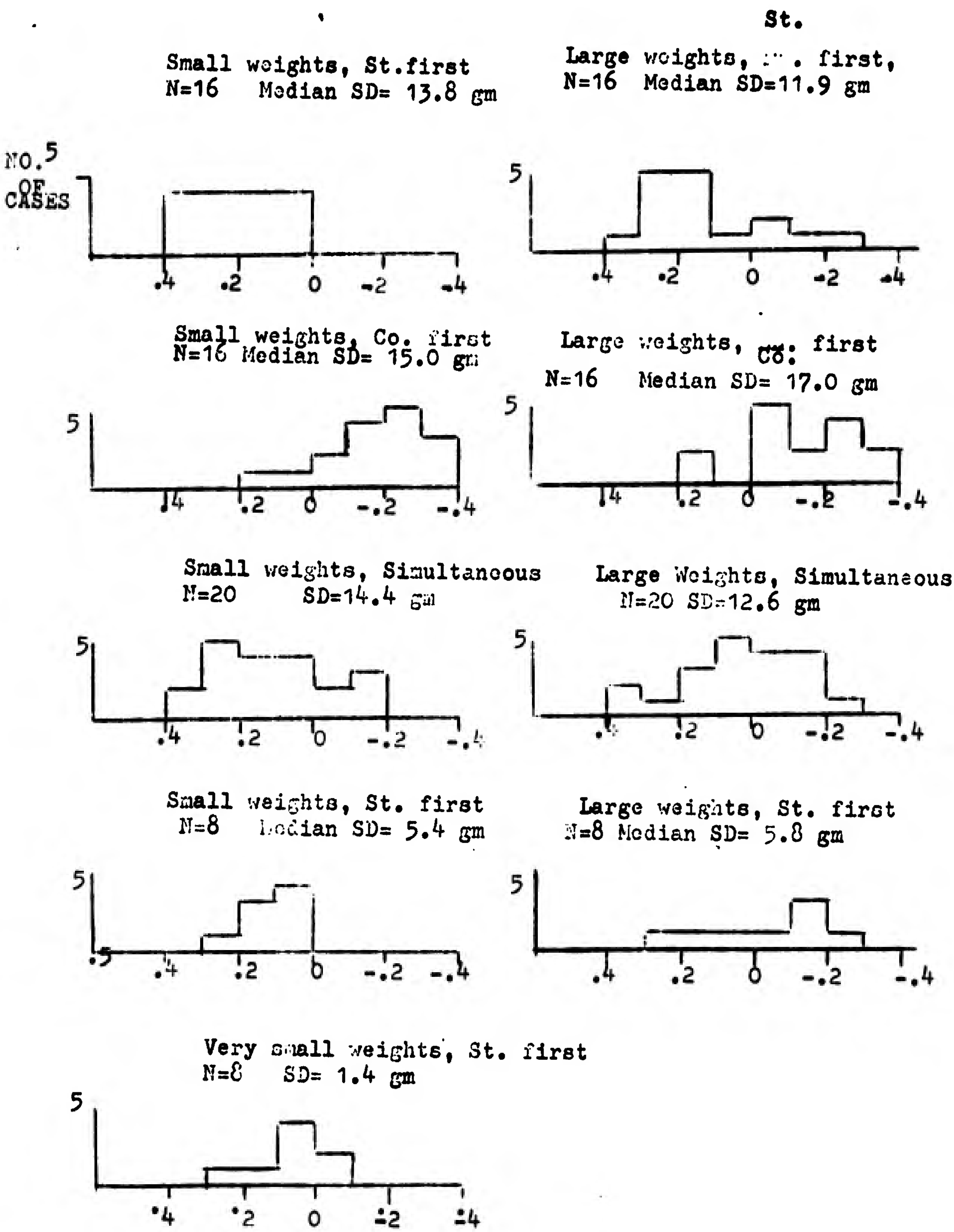


Fig.4. Histograms showing the distributions of positive and negative Response Biases obtained under different conditions. The first three pairs of histograms are taken from the previous experiment, and the last three histograms from this experiment.

REFERENCES.

BROWN W. (1910) The judgement of difference. Univ.of Calif.pub. in Psychol. 1. No.1

BRUNER, J.S. GOODNOW, J.J. & AUSTIN G.A. (1956) A Study of thinking. Ch.7. New York, Wiley.

HELSON, H (1947) Adaptation-level as a frame of reference for prediction of psycho physical data. Amer.J.Psychol. 60,1-29

HELSON H. (1948) Adaptation-level as a basis for a quantitative theory of frames of reference. Psychol. Rev. 55, 297-313.

HOLLINGWORTH, H.L. (1910) The central tendency of judgement. J. Philos. Psychol. Sci. Meth. 7, 461-469.

IRWIN, F.W. (1958) An analysis of the concepts of discrimination and preference. Amer. J. Psychol. 71, 152-165

KOESTER, T & SCHOENFELD, W.N. (1946) The effect of context upon judgement of pitch differences. J. exp. Psychol. 36, 417-430.

KOHLER W. (1923) Zur Theorie des Sukzessivvergleichs und des Zeitfehler. Psychol Forsch. 4,115-175.

LUCE, R.D. (1959) Individual choice behaviour. New York, Wiley.

LUCE, R.D. (1963) A threshold theory for simple detection experiment Psychol. Rev. 70, 61-79

NEEDHAM, J.G. (1934) The time-error as a function of continued experimentation. Amer. J. Psychol. 46, 558-567

NEEDHAM, J.G. (1935) The effect of the time interval upon the time error at different intensive levels. J. exp. Psychol. 18,530-543.

PIAGET, J & LAMBERCIER (1943) Recherches sur le developpement des perceptions: 2 Archives de Psychol. Geneve, 29,123-253

QUENOUILLE, M.H. (1959) Rapid Statistical calculations. London:

ROSS H.E. & GREGORY R.L. (1963?) Is the Weber fraction a function of Physical or perceived input? Q.J.exp.Psychol.

SENDERS & SOWARDS A. (1952) Analysis of response sequences in the setting of a psychophysical experiment. Amer.J.Psychol. 65,358-374

TANNER W.P. & SWETS J.A. (1954) A decision-making theory of visual detection. Psychol.Rev. 61, 401-409.

THURSTONE, L.L. (1928) The phi-gamma hypothesis. J. exp. Psychol. 11, 293-305.

REFERENCES (cont).

WOODROW, H (1933) Weight discrimination with a varying standard. Amer. J. Psychol. (45, 391-416.

WOODWORTH, R.S. & Schlosberg, H. (1955) *Experimental Psychology.* London, Methuen.

GRANT EOAR 63-93

APPENDIX. A proof of the statistical dependence between the Constant Error and the Standard Deviation.

by D. J. Laming.

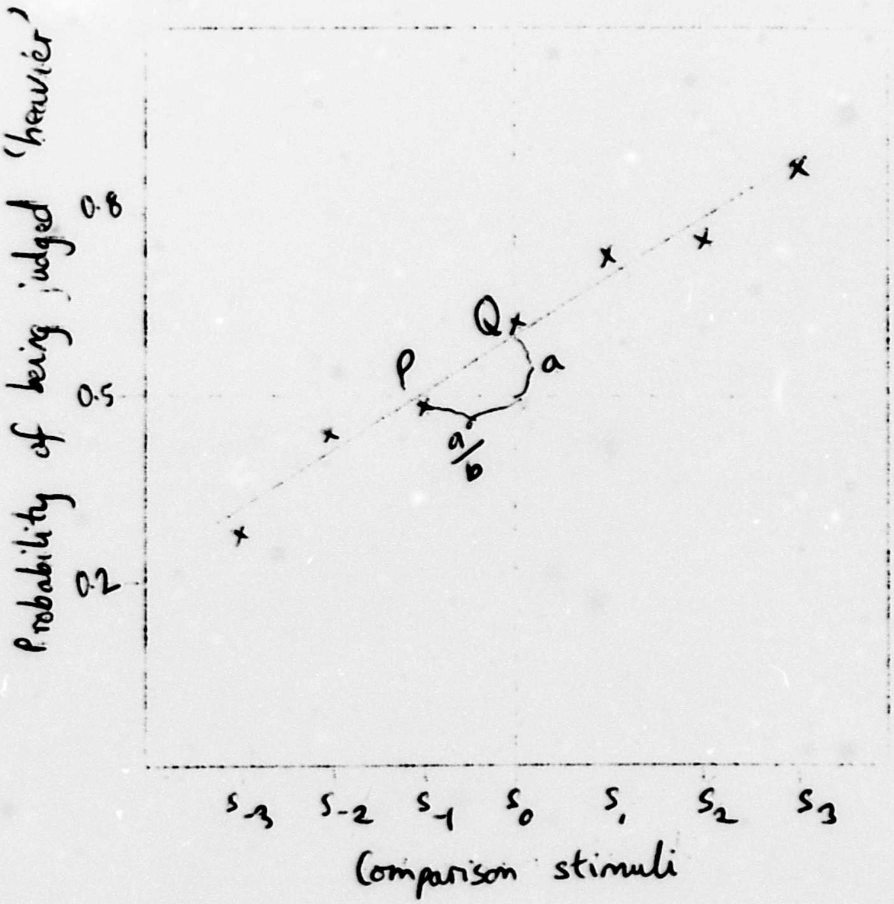
from the Psychological Laboratory, University of Cambridge.

The data from a psychophysical experiment using the method of frequency (Woodworth and Schlosberg 1954, p.205) are in the form represented in Fig. 1. s_0 is the standard stimulus and s_i ($i = \ldots -1, 0, 1, \ldots$) represent the values of the different stimuli to be compared with it. Each stimulus s_i is presented n_i times for comparison with the standard, s_0, and is judged greater than the standard r_i times.

It has been customary to analyse data such as these on the assumption that the relationship between the value of the stimulus, s_i, and the probability of judging it greater than the standard can be represented by a normal probability integral: in other words, that the threshold value is normally distributed. The mean and standard deviation of this distribution are estimated by probit analysis. This method is based principally on an outward similarity of the data to a normal probability integral, and is unnecessarily restrictive.

We shall suppose here that the frequency function of the threshold value is reasonably smooth so that its distribution function may be represented to a first approximation over the centre of its range by a straight line, $p = \frac{1}{2} + \alpha + \beta(s_i - s_0)$ where p_i is the probability of being judged greater than the standard, and α and β are constants of the functional relationship. We shall suppose that the range of this approximation covers both P ($p_i = \frac{1}{2}$), and Q ($p_i = \alpha + \beta s_0$)

Fig. 1. (appendix). Form of data from a psychological experiment using the Method of Constant Stimuli.



We shall also suppose that the form of the threshold distribution is constant among the population of experimental subjects but will differ in dispersion, and perhaps also in location. (Thus the inverse slope of the line, $\frac{1}{\beta}$, is proportional to the standard deviation). All these suppositions are implicit in the much stronger assumption of normality, but they are true for a much wider class of distributions, skew ones in particular. By this means we can estimate the median and percentage points of the threshold distribution.

For any pair of estimates (a,b) of (α, β) the likelihood function,

$$L = \prod_i \binom{n_i}{r_i} \left(\tfrac{1}{2} + a + bs_i\right)^{r_i} \left(\tfrac{1}{2} - a - bs_i\right)^{n_i - r_i}$$

where s_i is measured relative to s_0 and i runs over the range for which the straight line is a suitable approximation.

Maximising this:

$$\frac{\delta \log L}{\delta a} = \sum_i \left(\frac{\frac{r_i}{n_i} - \left(\frac{1}{2} + a + bs_i\right)}{n_i\left(\frac{1}{2} + a + bs_i\right)\left(\frac{1}{2} - a - bs_i\right)} \right) \qquad \frac{\delta \log L}{\delta b} = \sum_i s_i \left(\frac{\frac{r_i}{n_i} - \left(\frac{1}{2} + a + bs_i\right)}{n_i\left(\frac{1}{2} + a + bs_i\right)\left(\frac{1}{2} - a - bs_i\right)} \right)$$

The factor $\left[n_i\left(\frac{1}{2} + a + bs_i\right)\left(\frac{1}{2} - a - bs_i\right)\right]^{-1}$ is a weighting factor due to the variance of $\frac{r_i}{n_i}$, determined by the underlying functional relationship and the number of presentations of s_i. To a first approximation we may replace it by a constant w_i, so that both a and b are weighted sums of the r_i. Assuming that all judgement are independent, the r_i are binomial random variables, so that the distribution of a is approximately a generalized binomial times a scaling factor. In particular the variance of a increases monotonically with $\frac{1}{\beta}$, the inverse slope of the functional relationship.

It is customary to abstract two statistics from these data:

The Constant Error (CE) = OP = $-\frac{a}{b}$

The Standard Deviation (SD) = tan OQP = $\frac{1}{b}$

If there is no systematic CE with a given subject $E(a) = 0$. Provided $s_0 = \frac{\Sigma n_i s_i}{\Sigma n_i}$ (i.e. the standard is at the centre of gravity of the stimulus series) a and b are independent variables, so that the distribution of CE contains SD as a scaling factor. /CE/ CE and SD are uncorrelated, but not independent. /CE/ and SD are correlated. In face E (/CE/) is proportional to SD. If we now consider the entire population of experimental subjects, so that β varies, we increase the variation in b and hence the size of the correlation. Under variation of β, a and b are no longer independent variables. But the effect on a of increasing $\frac{1}{\beta}$ (see above) is such as to strengthen the correlation between /CE/ and SD.

If now there is a systematic CE, it may happen that this is essentially a shift either in a horizontal direction in Fig.1 or in a vertical direction. If we have a shift in a horizontal direction $E(CE) = c$, $E(a) = -\beta c$. In this case the correlation between /CE/ and SD is weakened, and in the limit, when c is so large that all the observed values of CE fall on the same side of zero, the correlation is zero. But we now have a negative correlation between /a/ and SD. If, on the other hand, we have a shift in a vertical direction, giving rise to a systematic CE, $E(a) = c'$, $E(CE) = \frac{c'}{\beta}$ In the limit we have no correlation between /a/ and SD, but a positive correlation remaining between /CE/ and SD.

The evidence reported in this paper is not conclusive between these two possibilities, but favours the suggestion that systematic CEs are due to shifts in the vertical direction in Fig. 1 rather than the horizontal direction. We should therefore focus our attention on the response bias when the variable is equal to the standard rather than the CE.

This appendix does not prove that there is no functional relationship between the Standard Deviation and the magnitude of the Constant Error, but that the methods hitherto used to investigate such a relationship are seriously confounded with a statistical artifact. However, the pattern of results hitherto reported, particularly in this paper, is consistent with the conclusion that the observed effect is due to the operation of this artifact alone.