Prior expectations in cross-modality matching

Donald Laming*

Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge, CB2 3EB UK

Abstract

A mathematical model is proposed to accommodate both the ‘regression’ effect in cross-modality matching and the central tendency of judgment. The basic idea is that the subject has two sources of data, (a) the stimulus to be judged and (b) previous stimuli in the experiment which afford some idea what magnitude of stimulus to expect. The minimum variance estimate of stimulus magnitude is a weighted average of these two. That weighted average biases judgments towards the prior expected value; at the same time, the variability of the judgments is reduced below the value which would attach to judgments based on datum (a) alone.

This basic idea is packaged in a process model which purports to describe the minutiae involved in making a cross-modality match. The process model adds this one substantive assertion to the mathematical idea: that prior expectations cannot be disregarded (notwithstanding that they are irrelevant to the judgment) because the prior experiences from which they are derived are an integral component of the subject making the judgment. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Cross-modality matching; Judgment; Process model; Prior expectations

1. Introduction

The model which follows addresses both the ‘regression’ effect (Stevens and Greenbaum, 1966) and the central tendency of judgment (Hollingworth, 1910).

1.1. The ‘regression’ effect

Stevens and his colleagues have shown many times over that subjects can assign
numbers to stimulus magnitudes in such a way that the geometric average increases approximately according to the power law relation

\[ N = aX^\beta, \]  

where \( N \) is the number uttered by the subject and \( X \) is the physical stimulus magnitude. Subjects can also adjust the magnitude of a stimulus to match a given number. Reynolds and Stevens' (1960) subjects first judged the loudness of seven different levels of broadband white noise. Then they adjusted the level of the noise to match six different numbers, numbers chosen to lie within the range used in the first (estimation) phase of the experiment. Both estimations and productions conformed reasonably to a power law (see Stevens and Greenbaum, 1966, Fig. 5); but the production data required a higher exponent (\( \beta \)) in Eq. (1) than did the estimates. A similar discrepancy is found between the two values of the exponent when subjects adjust one stimulus to match another. Stevens and Greenbaum (1966) asked 10 observers to match the duration of a light to the duration of noise; then the same 10 observers matched the duration of the noise to that of the light. Their geometric mean matches are shown in Fig. 1.

The dotted line of gradient 1 in Fig. 1 shows how the data would lie if matching was veridical. It is clear that both sets of matches are biased. The power law exponent,
expressing the regression of log duration of the light on the log duration of the noise is 0.87 when the light is matched to the noise, but 1.30 when the noise is matched to the light. These and other similar findings have been succinctly summarised by Stevens (1975).

The effect is systematic. The difference in gradient is often small in data aggregated from a small cohort of observers, but always lies in the same direction. ‘The observer tends to shorten the range of whichever variable he controls’ (Stevens, 1971, p. 426). Magnitude productions always have a steeper gradient than estimates. The difference in gradient can, however, be large for individual subjects. Stevens and Greenbaum (1966, Fig. 10) exhibit two examples from Stevens and Guirao (1962) for the magnitude estimation and production of levels of noise.

This systematic difference between the estimated gradients for magnitude estimation and production was named the ‘regression’ effect by Stevens and Greenbaum (1966). But that is misleading. It invokes an analogy with linear regression when there are errors in both variables (Kendall and Stuart, 1967, Ch. 29). In that case (errors in both variables; call them \( x \) and \( y \)) maximum likelihood analysis must correctly partition the variability between the two. Assigning all the variability to one variable, as in linear regression, gives a biased result; the regression of \( x \) on \( y \) has a different gradient to the regression of \( y \) on \( x \). The gradients of the two regression lines always differ according to the pattern of Fig. 1, whence Stevens and Greenbaum’s (1966) label. But in the present data one of the variables, either the stimulus magnitude to be estimated or the number to be matched, is a mathematical variable known (to the experimenter) without error. This means that regression of the response variable on the stimulus value assigns the variability correctly and is unbiased as a method of analysis. There is no statistical artefact (Cross, 1973) and we are looking at a real phenomenon of judgment.

Stevens’ (1971, p. 426) assertion that observers tend to compress the range of whichever variable they control does no more than describe what is observed to happen. This is equally true of Poulton (1989) who categorises the ‘regression’ effect as a stimulus contraction bias. Baird (1997) has proposed a sensory interpretation. Envisage that each stimulus engages a population of neurons and that the magnitude perceived depends on the aggregate discharge. During the time between presentation of the stimulus and production of the response, ‘the memory representation (trace) of the stimulus drifts in that the traces of the slow-firing neurons drop out of the aggregate’ (Baird, 1997, pp. 112–113). If, now, stimulus and response continua are interchanged, as in cross-modality matching, this adaptation of the neural response acts in the contrary direction. But no reason is given why the drift of the neural response should compress the range of the stimuli as perceived by the subject, and not simply reduce all perceived magnitudes by the same proportion.

This article offers an explanation of the ‘regression’ effect in terms of prior expectations.

1.2. The central tendency of judgment

An experiment by Hollingworth (1909) on the reproduction of guided movements points the way to an understanding of the ‘regression’ effect. There was a narrow slit cut
in a strip of cardboard which was then pasted onto a similar strip to form a guide. This guide could be readily traced with a pencil held by a blindfolded subject. Two seconds after tracing the guide, the subject attempted to produce a movement of the same length unguided, on a sheet of plain paper, and 2 s later a second reproduction. Different target lengths were presented in random order, each followed by two reproductions, until a total of 50 reproductions of each target had been recorded. Fig. 2 shows the mean reproductions of each target length.

In different experimental sessions different sets of target lengths were presented for reproduction: Series A ranged from 10 to 70 mm, Series B from 30 to 150 mm, and Series C from 70 to 250 mm. Within each series, the shorter movements were overproduced while the longer movements were underproduced, and log mean reproduction is approximately a linear function of log target length. Mean reproductions are biased towards a fixed point internal to each stimulus series; but different series are biased towards different fixed points, showing that the bias depends on the particular lengths the subjects have recently traced. But if one target length is presented repeatedly, by itself, in a separate session (filled circles in Fig. 2), the bias is negligible. Hollingworth (1910) dubbed this phenomenon the ‘central tendency of judgment’.

In respect of its formal design, this experiment has recently been replicated by Marks

![Fig. 2. Mean reproductions of target lengths of guided movement from Hollingworth (1909). Series A, B, and C are from different sessions in which different ranges of lengths were presented for reproduction. The filled points display the results of three further sessions in which one length only was presented repeatedly. (Figure from The Measurement of Sensation by Laming (1997, p. 22).)
(1993, Exp. 1), using magnitude estimation of 500 Hz tones instead of blindfold reproduction of lines drawn under guidance. Like Hollingworth (1909), Marks (1993) used sets of seven stimuli drawn from the ranges 25–55, 40–70, and 55–85 dB SPL; but, unlike Hollingworth, the stimulus sequences drawn from these ranges were all presented within the one session, in blocks, without any demarcation or other indication to the subject. Nevertheless the configuration of the geometric mean estimates in Fig. 3 is very similar to that of the open symbols in Fig. 2.

The faint dotted line is an aggregate relation for all three sets of stimuli taken together. It is a power law between judged magnitude and stimulus intensity fitted to the geometric mean estimates from each whole set. Mean estimates for individual stimuli within each set are biased towards a fixed point internal to the set; with different sets biased towards different fixed points, just as in Fig. 2.

Baird (1997, p. 192) presents a simulation of Marks’ (1993) data. The essential idea (I deliberately do not go into detail) is that the judgment on trial \( n \) assimilates to the previous judgment on trial \( n-1 \), the strength of assimilation increasing with the similarity between the successive stimuli. This means that there is greater assimilation between stimuli within each range than there is between ranges. This is equivalent to an assimilation of the perceived stimulus magnitudes which Poulton (1989) would categorise as a stimulus contraction bias. The idea I suggest below is related; it is,

![Fig. 3. Geometric mean magnitude estimates of the loudness of 500 Hz tones from Marks (1993, Exp. 1). The low, medium, and high sets of stimuli were presented in blocks within one continuous series. Data from Marks (1993).](image-url)
approximately, the net effect of such an assimilation, taking expectations over a sequence of stimuli drawn at random from each range.

1.3. Other experimental paradigms . . .

An article on sequential effects in psychophysical judgment inevitably invokes questions of the form ‘What about this experiment?’ In the interests of simplicity I exclude any consideration of category judgment. In such an experiment the fixed range of the available responses introduces an additional complication that I can well do without. For example, Johnson (1955, p. 354) writes:

‘. . . this phenomenon of central tendency is one manifestation of the regression effect that always occurs when the series of data are not perfectly correlated.’ (italics original)

1. Johnson (1955, p. 354) is commenting on the category ratings of colour differences from Philip (1947) in which errors of assignment of the extreme stimuli must necessarily put them in a category closer to the centre of the response range. There is no such constraint on the cross-modal matching of one stimulus by another, or on its reproduction, or on magnitude estimation.

2. Johnson’s implicit explanation is another version of Stevens and Greenbaum’s (1966) ‘regression’ idea. It does not apply to the data shown in Figs. 1 and 2 because there was no constraint in those experiments from a small fixed response range.

3. The data set out in Figs. 1–3 are (geometric) means of the original observations. There is no reason why that kind of calculation should be appropriate for category judgments which, to my mind, require a quite different kind of analysis (cf. Braida and Durlach, 1972).

1.4. . . . and other sequential effects

An article on sequential effects in psychophysical judgment also invokes questions of the form ‘What about this effect?’ Again, for reasons of simplicity, I confine my attention to the phenomena displayed in Figs. 1–3. There is, for example, a remarkable autocorrelation of log numerical estimates (e.g. Baird et al., 1980) which is related to other effects in both magnitude estimation and category judgment (see Laming, 1984, 1997). That also is excluded from consideration. Envisage that there are several intrinsic constraints operating trial-to-trial in psychophysical judgment tasks. I intuit — I may be wrong, but I intuit — that the data in Figs. 1–3 (and other similar data) exhibit one of those elementary trial-to-trial processes relatively uncontaminated by the others. Let’s go after that one.

I am concerned in what follows with an approximately linear relation of log response (matching stimulus magnitude or reproduction or numerical estimate) to log stimulus magnitude that systematically shows the response range to be compressed in a manner specific to the set of stimuli presented for judgment. This particular ‘regression towards the mean’ is not a statistical artefact. It requires a psychological explanation.
2. Two ideas

Hollingworth’s (1909) experiment shows that the length of a reproduced movement is a compromise between the actual length of the guided movement on that particular trial and a length which the subject has come to expect on the basis of previous trials, recently experienced. So the psychological idea is that the observer uses his experience of previous stimuli in the experiment to provide some prior idea of the likely length or duration of the next stimulus. To that psychological intuition I add the mathematical idea that the influence of previous stimuli on perception of the present one can be modelled as the optimal combination of data from two independent sources.

Envisage two experiments providing independent data to test the same hypothesis. In the present context one experiment is the presentation of a stimulus $S_i$ of magnitude $X_i$. That experiment provides a datum $j_i$, where

$$f(j_i|S_i) = (2\pi\sigma^2)^{-1/2} \exp\{-\frac{1}{2} (j_i - \log X_i)^2 / \sigma^2\}. \quad (2)$$

The second experiment subsumes the presentation of all previous stimuli and is assumed, for present purposes, to be equivalent to the presentation of a notional prior stimulus $S_p$ of magnitude $X_p$, giving a datum $j_p$, where

$$f(j_p|S_p) = (2\pi\sigma^2_p)^{-1/2} \exp\{-\frac{1}{2} (j_p - \log X_p)^2 / \sigma^2_p\}. \quad (3)$$

Notwithstanding that $j_p$ results from the observation of some other stimulus, it is interpreted (for the purposes of the model) as though it provided information about $S_i$.

Statistical theory now tells us that the minimum variance estimate of (log) stimulus magnitude is

$$\frac{I_f(j_i|S_i) + I_f(j_p|S_p)}{I_f(S_i) + I_f(S_p)}, \quad (4)$$

and thereby generates a gradient of the cross-modal matching relation equal to $I/I_f(S_p)$ which is always less than 1.

---

1. Assume, for the purposes of choosing an estimator of log stimulus magnitude, that the expectation of $\xi_i$ is log $X$. Consider weighted averages of $\xi$ and $\xi_p$ such that the weights sum to 1 (i.e. of the form $a\xi + (1-a)\xi_p$); all such estimators have expectation log $X$ and are unbiased. The variance of such an estimator is the

$$[a\log\xi + (1-a)(\log X)]^2 = a^2\sigma^2 + (1-a)^2\sigma^2_p,$$

Differentiation with respect to $a$ and solution of the resulting equation gives

$$a = \sigma^2_p/(\sigma^2 + \sigma^2_p), \quad (5)$$

so that

$$a^2\sigma^2 + (1-a)^2\sigma^2_p.$$
The complement of this quantity, $I_p/(I + I_p)$, is the relative weight of the prior expectation. Table 1 shows the values of prior expectation and their associated weights estimated from the experiments in Figs. 1–3.

The gradient of the regression line ‘light adjusted to noise’ in Fig. 1 is 0.87. This is the weight attaching to the stimulus to be matched and the weight attaching to prior expectation in Table 1 is its complement. When the level of noise is adjusted to match the light, the gradient is 1.30. This has reciprocal 0.772, which is the weight attaching to the stimulus to be matched, and its complement is 0.229. In this way the optimal combination of data from two independent sources generates the observed direction of Stevens’ ‘regression’ effect.

A similar calculation gives the weights attaching to prior expectation in Fig. 2 because, again, unbiased responses would show a gradient of 1. But the calculation for the data in Fig. 3 is more complex. The faint dotted line in that figure is an aggregate relation (assumed unbiased) for all three sets of stimuli taken together. It has equation

$$\log N = -1.424 + 0.0344 \log X,$$

where the stimulus intensity (log X) is expressed in dB. If, for stimulus values selected from the low set, log X is substituted by $\{a \log X + (1-a) \log X_p\}$, the gradient with respect to stimulus intensity is reduced to 0.0344a which (for the low set) is estimated to be 0.0293. The weight attaching to the prior expectation $[\log X_p]$ is $(1-a)$ which is then equal to 0.148 ($= 1-0.0293/0.0344$).

When the statistical significance of prior expectation is tested against the variation about each fitted regression line, it is highly significant; this is apparent even from inspection of the figures. If, more rigorously, significance (in the experiments by Hollingworth and Marks) is tested against the variation in weight between experimental conditions, it is typically significant at the 5% level, sometimes better.

This model has practical application. On the night of 15th October 1987 southern England was buffeted by its most severe storm since 1703. The British meteorological office signally failed to forecast the strength of that storm and one weatherman was on

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Experimental condition</th>
<th>$X_p$</th>
<th>$I_p/(I + I_p)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matching of duration</td>
<td>Light adjusted to match noise</td>
<td>1.282 (s)</td>
<td>0.129</td>
</tr>
<tr>
<td>Stevens and Greenbaum (1966)</td>
<td>Noise adjusted to match light</td>
<td>0.511 (s)</td>
<td>0.229</td>
</tr>
<tr>
<td>Reproduction of guided movement</td>
<td>Series A: 10–70 mm</td>
<td>41.48 (mm)</td>
<td>0.238</td>
</tr>
<tr>
<td>Hollingworth (1909)</td>
<td>Series B: 30–150 mm</td>
<td>73.38 (mm)</td>
<td>0.228</td>
</tr>
<tr>
<td></td>
<td>Series C: 70–250 mm</td>
<td>128.43 (mm)</td>
<td>0.349</td>
</tr>
<tr>
<td>Magnitude estimation of loudness</td>
<td>Low set: 25–55 dB SPL</td>
<td>34.63 dB</td>
<td>0.148</td>
</tr>
<tr>
<td>Marks (1993)</td>
<td>Medium set: 40–70 dB SPL</td>
<td>59.89 dB</td>
<td>0.308</td>
</tr>
<tr>
<td></td>
<td>High set: 55–85 dB SPL</td>
<td>67.89 dB</td>
<td>0.366</td>
</tr>
</tbody>
</table>
television earlier that very day saying categorically that there would be no hurricane. But there was! Meteorological observations are fragmentary; they have to be extrapolated forwards sometimes up to 24 h and then interpreted to produce the forecast. In that process the forecast is modified by purely human expectations based on past experience of weather in southern England in mid-October. Hurricanes do not happen in southern England and no meteorologist is going to forecast one.

To this point the interpretation set out above, in terms of the optimal combination of data from independent statistical sources, has delivered a weighted mean response of the form of Eq. (4). I comment that, subject only to a reinterpretation of the coefficients, the same equation could have been derived from other ideas — simple response assimilation, for example. That is to say, the ideas put forward here constitute one possible interpretation only of the data in Figs. 1–3. If, however, the present statistical interpretation be accepted, it generates a specific problem of its own and (via Eq. (5)) an answer to it.

2.1. Variance of judgments

If prior expectation is actually irrelevant to present judgment, why does it produce the bias that it does? The answer to that question comes in two parts. The first part, which I give here, links bias in judgment to an increase in repeatability — specifically to a decrease in the variance of a sample of judgments, the shift in the mean being undetectable by the person making the judgments. The second part of the answer, much later, asserts that prior expectation is simply the prior state of the subject so far as that prior state interacts with the making of the judgment and that any idea of disregarding prior expectation would be a contradiction in terms.

An experiment by Sinha (1952) points the way to the first part of my answer. Sinha asked his subjects to fixate a rotating spiral. When the spiral was stopped suddenly, the subjects experienced a strong sensation of contrary movement (expansion because the direction of rotation made the spiral appear to be continually contracting). Subjects were asked to measure the duration of the after-effect they experienced with a stopwatch. There were two sessions for each subject 5 or 6 days apart with 10 measurements in each session. In the first session subjects started the stopwatch as soon as they observed the contrary movement begin and stopped the watch when the movement ceased, but were not themselves allowed to look at the watch or to know how long their after-effect had lasted. Before the second session, however, subjects were told ‘...Incidentally it may be of interest to you that I have tested about 50 subjects and in the majority of cases, the average reaction time was found to be ... [different averages given to different subjects] ... with 30 seconds exposure’ (Sinha, 1952, p. 9). Subjects were then allowed to read out their measured duration themselves ‘as it was the last day with them’. Table 2 shows the summary data from two subjects.

Comparing day 1 with day 2, each subject’s mean shifts towards the ‘group’ average announced by the experimenter. In addition, the variability of successive judgments is reduced. These two phenomena go hand-in-hand and suggest that prior expectations have the effect that they do because they reduce the uncertainty of judgment.
Table 2
Duration of spiral after-effect (Sinha, 1952)

<table>
<thead>
<tr>
<th>Subject</th>
<th>1st day</th>
<th>2nd day</th>
<th>1st day</th>
<th>2nd day</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ('Standard' duration 15 s)</td>
<td>22.5</td>
<td>17.6</td>
<td>20.7</td>
<td></td>
</tr>
<tr>
<td>2 ('Standard' duration 20 s)</td>
<td>14.6</td>
<td>14.6</td>
<td>0.94</td>
<td></td>
</tr>
</tbody>
</table>

*Excluding the first measurement from the series of 10.*

Fig. 4 shows the variable errors\(^2\) of the reproductions of movements in Fig. 2 plotted against the means, now with a linear abscissa. For the three sessions in which one target length only was presented (filled circles) the variable error increases almost in proportion to the mean. That control data may be adequately described by the equation

$$\text{Variable error} = \sqrt{1.87 + 0.005(\text{mean reproduction})^2}$$

(6)

where the constant element of variability 1.87 mm\(^2\) is arguably the limiting accuracy to which a line can be drawn blindfold with the left hand. But the variable errors from the experimental (mixed) sessions are systematically greater and this requires some explanation.

Holland and Lockhead (1968) and Ward and Lockhead (1970, 1971) have shown that in absolute identification tasks the judgment of each stimulus assimilates to its predecessor in the same way that the reproductions of movements in Fig. 2 assimilate to the geometric mean of the stimulus set; and Cross (1973) and Ward (1973) have demonstrated a similar assimilation in magnitude estimation. Such assimilation appears to be a general phenomenon, and it appears likely that the global bias in Figs. 1–3 is simply the trial-by-trial assimilation of matches, reproductions, and estimations to preceding stimuli averaged over a sequence presented in random order. Certainly the effect of stimulus frequency on two-alternative choice-reaction times, on both reaction times and errors, can be related to sequential effects operating trial-by-trial in this way (Laming, 1968). On this basis the prior stimulus in Hollingworth’s (1909) experiment will chiefly represent the target length on the preceding trial, and the expected length of reproduction on trial \(n\) will be (rewriting Eq. (6) Fig. 4)

$$E\{y\} = (I \log X_n + I_y \log X_{n-1})/(I + I_y).$$

(7)

The preceding target length \(X_{n-1}\) will, of course, be different for different presentations

\(^2\text{If } y \text{ is the length of reproduction of some target length } X \text{ in Fig. 2, and if } \bar{y} \text{ is the mean reproduction of length } X, \text{ then } (\bar{y} - X) \text{ is the constant error and } (y - \bar{y}) \text{ is known as the variable error. The variable error therefore measures the precision of repeated reproductions about their mean, while the constant error measures the bias. The data in Fig. 4 are averages of the absolute variable error } |y - \bar{y}| \text{ (rather than, as one would calculate today, root mean squares). If the distribution of variable error is normal, the absolute error is equal to 0.80 times the standard deviation.}\)
Fig. 4. The variable errors of reproduction from Hollingworth (1909, cf. Fig. 2). As before, Series A, B, and C are from different sessions in which different ranges of amplitudes were presented for reproduction. The filled points display the results of three further sessions in which one amplitude only was presented repeatedly; the continuous curve is Eq. (6). (Figure from The Measurement of Sensation by Laming (1997, p. 37).)

of a given length $X_i$ (i.e. $X_n$ in Eq. (7)) and will contribute an additional element to the variable error.

It should be noted that prior expectation applies equally to the single stimulus values presented in the control series in Fig. 4 so that the coefficient 0.005 in Eq. (6) is an estimate [of $0.8^2$ times] the reduced variance $(I + I_n)^{-1}$; but when only one target length is presented, $X_{n-1} = X_n$ on every trial and no bias shows. When, however, successive target lengths vary at random, the variance of the aggregated reproductions contains the additional contribution $I_n \text{Var}[(\log X_{n-1})/(I + I_n)]$ which raises the open data points in Fig. 4 above the control values. This provides a more detailed perspective on the mechanism of prior expectations. The idea of a prior stimulus introduced above can now be seen to be a convenient mathematical fiction which subsumes the pattern of sequential interactions operating on a random sequence of stimuli at a trial-to-trial level.

3. Process model

I next set out a process model which purports to describe the minutiae of cross-modal matching. It provides packaging for the mathematical idea introduced above in the sense that it shows how Eqs. (2) and (3) might fall out of the operation of selecting a
cross-modal match. I emphasise that this particular model is far from unique; many other equivalent models could be devised to serve the same purpose. The model does not itself add anything of quantitative substance, except to suggest (later) that prior expectations cannot be ignored. As a matter of convenience I envisage a matching task in which the subject can adjust the matching stimulus ad lib (unlike the example experiment in Fig. 1), but in which both stimulus and match are values on the same continuum. Different matching tasks will, in fact, require slightly different process models, and the model which follows can easily be modified to suit different experimental tasks.

The hypothesised sequence of operations is set out in Fig. 5.

3.1. Prior state

Each stimulus to be matched is presented to a subject who has been instructed what to do and has accumulated prior experience in the experiment. That prior experience has

![Diagram](image-url)
already been shown to be material to the choice of match. The initial stage of the process model characterises the prior state of the subject by a random variable $\xi$. This variable takes different values on different trials of the experiment in recognition of the fact that the stimuli are presented in random order. In principle $\xi$ is a multi-dimensional random variable of arbitrary dimensionality. But, in practice, it only has to describe those matches the subject would make if asked to guess in the absence of any physical stimulus and can therefore be taken to be a unidimensional variable such as might be contributed by the presentation of a notional prior stimulus $S$. The prior stimulus $S$ characterises the subject’s distribution of prior expectations on different trials. I set the density function of $\xi$ to be (as in Eq. (3) above)

$$f(\xi\mid S) = (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{1}{2}(\xi - \log X)^2 / \sigma^2\right).$$

### 3.2. The stimulus to be matched

The presentation of a stimulus $S$ generates another internal state variable $\xi$, again taken to be unidimensional, according to the density function of Eq. (2)

$$f(\xi\mid S) = (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{1}{2}(\xi - \log X)^2 / \sigma^2\right).$$

The variable $\xi$ encapsulates the subject’s entire knowledge about the magnitude of the present stimulus and is the internal value needing to be matched.

### 3.3. Selection of candidate match

A subject is assumed to proceed by looking at a series of possible matches until one is found which will do. These candidate matches are not selected entirely at random, but according to a prior distribution $\pi(\mu)$ which depends, in turn, on the prior state $\xi$. A candidate match $S_\mu$ (of mean $\mu$) is selected in proportion to the probability that, as prior stimulus, it would generate the internal state $\xi$. Using Bayes’ Theorem,

$$f(S_\mu\mid \xi) = f(\xi\mid S_\mu)\int f(\xi\mid S) f(S_\mu) d\xi = (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{1}{2}(\mu - \xi)^2 / \sigma^2\right).$$

### 3.4. Sampling of candidate match

The candidate match, now realised as a physical stimulus, is sampled, generating an internal state variable $\xi$ with mean $\mu$ and density $f(\xi\mid S_\mu)$ according to Eq. (2). The candidate match will be accepted if $\xi$ is sufficiently close to $\xi$. The joint probability of selecting $S_\mu$ and generating the internal state variable $\xi$ is

$$f(\xi\mid S_\mu) f(S_\mu) = (4\pi^2 \sigma^2)^{-1/2} \exp\left(-\frac{1}{2}[(\mu - \xi)^2 / \sigma^2 + (\xi - \xi)^2 / \sigma^2]\right).$$
3.5. Criterion of acceptance

If \( \xi_\mu \) is acceptably close to \( \xi_r \), the matching process terminates; otherwise a fresh candidate match is selected for trial. This termination criterion can be approximated by a test of equality, so I set \( \xi_\mu = \xi_r \) at the termination of the trial.

3.6. Posterior distribution of match

This is calculated from Bayes’ Theorem as

\[
\frac{f(S | \xi_\mu, \xi_r) = f(\xi_r | S_\mu) f(S_\mu | \xi_\mu)}{\int_{-\infty}^{\infty} (4\pi \sigma^2 \sigma_r^2)^{-1/2} \exp \left\{ -\frac{1}{2} \left[ \frac{(y - \xi_r)^2}{\sigma^2} + \frac{(\xi_\mu - y)^2}{\sigma_r^2} \right] \right\} dy}
\]  

(10)

When the integral in Eq. (10) is evaluated and divided into the numerator, the match is found to have

\[
\text{mean} = (I \xi_i + I_\sigma \xi_\mu)/(I + I_\sigma)
\]

(11)

and

\[
\text{variance} = (I + I_\sigma)^{-1},
\]

(12)

writing \( I \) for \( \sigma^{-2} \) and \( I_\sigma \) for \( \sigma_r^{-2} \) as previously.

3.7. Unconditional distribution of matches

Eqs. (11) and (12) give the mean and variance of the match in terms of the internal state variables \( \xi_\mu \) and \( \xi_r \) which are unobservable. An unconditional distribution of matches is required. Since \( \xi_\mu \) and \( \xi_r \) are normal, respectively \((\log X_\mu, \sigma^2_\mu)\) and \((\log X_r, \sigma^2_r)\), the unconditional distribution will also be normal with

\[
\text{mean} = (I \log X_i + I_\sigma \log X_\mu)/(I + I_\sigma).
\]

(4)

This, of course, is Eq. (4).

Since \( \xi_\mu \) and \( \xi_r \) are random variables, there is an extra element of variability in the unconditional distribution equal to \((I + I_\sigma)^{-1}\). The unconditional distribution of matches therefore has

\[
\text{variance} = 2(I + I_\sigma)^{-1}
\]

(13)

which is exactly twice the value of Eq. (5).

These features of the process model are worthy of note.

1. The internal state variables do not appear in the ultimate formulae. So the same formulae would result from any other assumptions about the internal state of the subject, provided only that that assumed internal state was sufficient to support the
normal model of Eq. (2). It follows that while the applicability of the normal distribution is an empirical assumption and admits experimental test, the use of a unidimensional internal state variable does not.

2. If there were no prior expectation, \( I_o(=\sigma_o^{-2}) \) would be zero (prior density of infinite variance), and the variance of the observed distribution of matches \( 2/I \). Because \( I_o \) is positive, \( 2/I \) is an upper limit to the variance specified by Eq. (13). That is, prior expectations reduce the variability of the judgment. This reduction in variability comes at the expense of the matching bias (the intrusion of the term \( \log X \) in Eq. (4)).

3. Eq. (13) has the multiplier 2 because the variances \( \sigma^2 \) and \( \sigma^2_o \) enter into the matching process twice over, first in the perception of the stimulus and the prior stimuli and again in the selection of the candidate match. If the target experiment had been a magnitude estimation/production pair, then the second involvement of the variances \( \sigma^2 \) and \( \sigma^2_o \) would not have been needed (numbers are mathematical quantities known exactly). I take this opportunity to re-emphasise that an appropriate process model depends on the particular experiment, but that in all cases it can be made to deliver formulae of the form of Eqs. (4) and (5).

4. The mean match is \( (I \log X + I_o \log X_o)/(I + I_o) \) which assigns a relative weight \( I_o/(I + I_o) \) to the prior stimulus. This weight is greatest when \( I \) is small (\( \sigma^2 \) large). That is to say, the effect of prior expectations shows most strongly with the most difficult judgments.

3.8. Can prior expectations be disregarded?

I now return to the question: if prior expectations are actually irrelevant to present judgments, why do they have the effect that they do? and propose the second part of the answer. The process model uses \( \xi_t \) to describe the state of the subject at the point in time when the stimulus is presented. Although, mathematically, prior expectation is modelled as just another source of data which the subject might use or disregard as he pleases, at this point the process model introduces a substantive assumption. Within the terms of that model, prior expectations are (a component part of) the subject. For this reason they cannot be disregarded.

The assertion that prior expectations cannot be disregarded is an empirical matter and requires collateral support. I cite one example from the problems of flying ‘blind’.

When a pilot (I am talking of light aircraft, flown manually) can see the horizon, orientation by sight is satisfactory. In cloud or at night, on the other hand, the pilot must rely on instruments to indicate the attitude of the aircraft. But natural behaviour is to ignore the instruments and rely on postural and vestibular sensations; this is what people do when standing or walking with their eyes closed. Problems arise because the interpretation of postural sensations is fallible. For example, the vestibular organ is sensitive to gravity and to linear acceleration (especially critical when an aircraft is banking), but not to velocity per se. So illusions occur when bodily orientation can no longer be monitored visually. A pilot will ‘feel’ the aircraft to be at a different attitude to that which it really is. One study of judgments of aircraft attitude by experienced, but blindfold, USAF pilots showed 60% errors (Jones et al., 1947).
Learning to fly blind means learning to disregard postural and vestibular sensations and to rely exclusively on instruments. But, when an inadequately trained pilot flies into cloud, it has repeatedly been found that he is unable to disregard bodily sensations such as the feel of the seat beneath his bottom, even though he knows (intellectually) that those sensations are not to be trusted. This is so, notwithstanding that the pilot’s attention to his postural and vestibular sensations usually leads to loss of control and to loss of life. An inadequately trained pilot in a light aircraft is unable to ignore information that is not only known to be irrelevant, but also fatal.

4. Summary

The global effects of prior stimuli on present judgment can be modelled mathematically as the optimal combination of statistical data from independent sources. That model accommodates the bias apparent in both Stevens’ regression effect (Fig. 1) and Hollingworth’s central tendency of judgment (Fig. 2) and also the reduction in the variability of judgment seen in Table 2 (Sinha, 1952). But the psychological idea that prior expectation is simply another source of data which the subject can take notice of or disregard as he pleases needs to be resisted. Prior expectation is the subject making the judgment.

Acknowledgements

I thank Professor L.E. Marks for kindly providing the numerical data in Fig. 3.

References

Jones, R.E., Milton, J.L., Fitts, P.M., 1947. An investigation of errors made by pilots in judging the attitude of an aircraft without the aid of vision. AAF memorandum report TSEAA-694-13, Engineering Division, Air Materiel Command, Wright Field, Dayton, OH.


