Psychological Review
1984, Vol. 91, No. 1, 68–111

A Theory of Criterion Setting With an Application to Sequential Dependencies

Michel Treisman and Thomas C. Williams
University of Oxford, Oxford, England

Three models of sequential effects in psychophysical tasks are defined and experimental results described. These appear most consistent with a model in which the momentary value of the criterion is modified by memory traces, an independent trace being retained for each relevant past event. On this basis, a theory of criterion setting is developed: A long-term process determines an initial value for the criterion, and two short-term processes adjust the criterion (a) to match current changes in the expected probabilities of signals and (b) to maximize the information transmitted by the subject's responses. The theory is applied to results in the literature.

Although signal detection theory (SDT; Green & Swets, 1966), which is now widely applied, includes the hypothesis of a decision criterion and puts forward normative prescriptions for its value, the problem of how a subject maintains his criterion at the appropriate value and adjusts that value to take account of relevant intercurrent events has largely been ignored. It is regrettable that criterion setting should generally be treated as a given but unexamined fact, because its proper understanding may clarify important psychological aspects of decision in all areas and may help to explain some still unresolved problems such as the causation of sequential effects. The object of the present article is to present a theory of criterion setting developed by the first author. As this theory arose from a consideration of some features of sequential dependencies, we begin with a brief discussion of the latter.

It is a familiar but still unexplained observation that the outcomes of successive psychophysical decisions are related (Verplanck, Collier, and Cotton, 1952). Early studies were concerned with determining whether response assimilation might be due, on the one hand, to factors extrinsic to the decision process, such as adaptation and other peripheral sensory consequences of stimulation (Collier, 1954b) or to spontaneous variations in the response of sense organs, or, on the other hand, to direct effects of the events on an earlier trial on the determinants of the decision on a later one. Collier (1954a) found that the association between responses decreased with the length of the intertrial interval, and an important experiment by Howarth and Bulmer (1956) provided evidence against the theory of endogenous fluctuations in sensitivity. They used a method introduced by Verplanck et al. (1952) in which only a single (faint, visual) signal intensity, chosen to give about 50% detection, was used, and this same intensity was presented on every trial. This method demonstrated *assimilation.* Howarth and Bulmer found that the probability of a *YES* response on trial i was increased if the response on trial $i - 1$ was *YES.* (We represent responses by italic upper case letters and stimuli by italic lower case.) At random intervals they omitted the signal for a sequence of three blank trials, so as to force three *NO* responses. After three such forced *NOES*, the same degree of assimilation was shown on the next trial as after three naturally occurring negative responses, which would not be expected if the latter were normally explained by a spontaneous trough in sensitivity.

This work was done while the authors were in receipt of support from the Medical Research Council of Great Britain. Thomas C. Williams was responsible for running the experiments reported and Michel Treisman for the theoretical analysis.

We would like to thank D. M. Green and B. S. Rosner for useful comments.

Requests for reprints should be sent to Michel Treisman, Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, England.

Although *assimilation* and *contrast* are well established terms, they can be confusing. The difficulty is that an implicit model appears to underlie them, although it is not usually clearly stated. Assimilation may be understood to imply that in some way the current stimulus input is made more similar to a preceding stimulus input than it would be if the former had not occurred; contrast implies the reverse. Thus these terms trail theoretical connotations rather than simply describing the phenomena. If we use these terms they will be descriptive only, to avoid confusing implications, but for the most part we avoid them and instead describe an increase in the probability of response R_i following a preceding R_i (assimilation) as a *positive sequential dependency* and a decrease in $P(R_i)$ following a previous R_i (contrast) as a *negative dependency.*

SDT (Green & Swets, 1966) made new concepts available that may be used to explain sequential phenomena. Thus, performance in a simple detection task may be considered to be limited by a noise distribution on a decision axis z. We assume the noise distribution to be normal with a mean M_n of 0 and a standard deviation of 1 and the choice between the responses YES and NO to be determined by a criterion z_c. This is illustrated in Figure 1-1 (for the moment, ignore z_d and z_c'). A number of alternative principles for defining an optimal or practical criterion have been proposed (Birdsall, 1955; Dusoir, 1975; Thomas & Legge, 1970). Theoretical treatments have usually preferred to assume that subjects define their criterion as the maximum expected value criterion β, although there are cases where it may not be possible to apply this criterion and the Neyman-Pearson criterion may be more appropriate (Treisman, 1964; Treisman & Watts, 1966). The latter places the criterion z_c at that position at which it cuts off a proportion of the noise distribution on its right equal to ϵ, the acceptable limiting false alarm rate. We usually express the criterion as a deviation from the mean of the noise or the signal distribution measured on a standardized decision axis, rather than as a likelihood ratio.

In the *single intensity procedure* (Verplanck et al., 1952) a signal of the same value is presented on every trial. The model for this case is illustrated in Figure 1-2, which shows a signal distribution $f_s(z)$ whose mean M_s coincides with the criterion z_c so as to ensure 50% of detections.

These models allow the probability of detection, $P(Y)$, to be modified in a number of ways. In a detection task, $P(Y)$ could change because some factor acts directly on the criterion to modify its value or because of a shift in the mean of the noise distribution or of the signal distribution. Thus if M_n shifts to the left and z_c is a Neyman-Pearson criterion, it should follow M_n to the left, and this consequential shift would cause an increase in $P(Y)$ for an unaltered signal distribution. An increase or decrease in the variance of the noise distribution would also cause a consequential shift in a Neyman-Pearson criterion. Endogenous fluctuations in sensitivity might work through such effects on the noise and signal distributions. However, despite some evidence for variations in sensitivity that are not modality specific (Long, 1980), we assume in the present discussion that sequential dependencies are brought about by changes in the decision process mediated by direct effects on the position of the criterion.

A number of paradigms have been used to study sequential dependencies, and a variety of explanations have been offered for them. The major paradigms discussed in this article are presented in Figure 1. Detection studies have usually shown positive dependencies (e.g., Collier & Verplanck, 1958; Howarth & Bulmer, 1956; Levitt, 1967). An opposite result was reported by Sandusky and Ahumada (1971), who presented a 500-Hz tone gated on simultaneously with a 0.1-s burst of noise. They found that $P(Y)$, the probability of YES, increased following noise on the previous trial. (Whether this was due to the noise, or to the response NO, is confounded.) Also, both $P(Y|s)$ and $P(Y|n)$ increased when the probability of the signal, $P(s)$, decreased. These are negative effects, and Sandusky and Ahumada noted that they contrast with the positive effects usually obtained. They suggested that the latter require the stable frame of reference provided by a continuous background noise.

In the *identification procedure* two stimuli that differ by a small amount on a sensory dimension are presented in random order (e.g., two tones, s_1 and s_2, with s_1 louder than s_2). One of the stimuli is presented on each trial, and the subject must make an identification response, such as *LOUD* or *SOFT*. A Thurstonian model for this task is shown in Figure 1-3. There are two distributions of sensory effects on a decision axis, and this axis is partitioned by a criterion, z_c, into regions giving the responses S_1 and S_2.

This model is formally similar to the SDT detection model, but neither explains sequential dependencies. Parducci and Sandusky (1965) applied the identification procedure to a point of light that could appear in either of two positions. The subject was required to say whether the light was *LEFT* or *RIGHT*. They found that responses without feedback were more accurate for stimulus alternations than for stimulus repetitions (that is, there was a negative effect). Furthermore, the overall probability of a *LEFT* response approximated .5 whether *left* was actually presented on .2, .5, or .8 of trials.

Tanner, Haller, and Atkinson (1967) used an identification task in which one of two tones differing slightly in amplitude (2.5 db on average) was presented on each trial. With no feedback, the probability of *LOUD* increased from .52 to .60 as the probability of the louder signal increased from .1 to .9. This increase is so small because both $P(\textit{LOUD}|\textit{loud})$ and

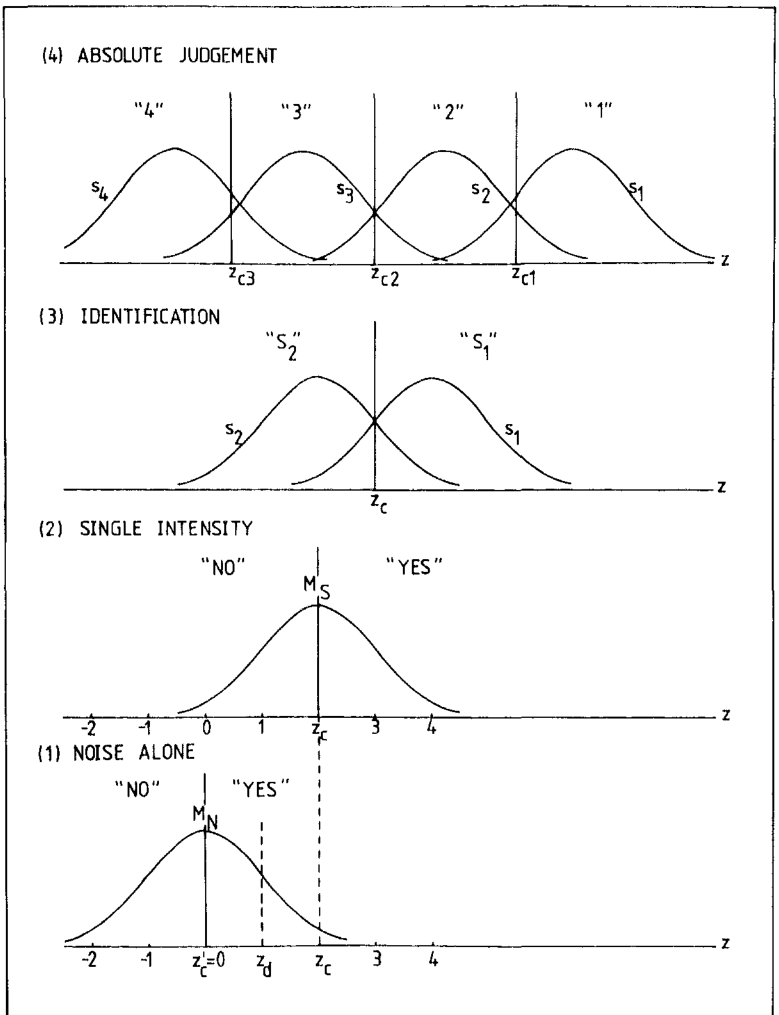


Figure 1. Representations of Thurstonian models for some procedures that have been used to study sequential dependencies. (The symbols are defined in the text.)

$P(\text{LOUD}|soft)$ fell as $P(loud)$ increased: Although there were more *loud* stimuli, LOUD responses did not increase in proportion. Results in which signal probability has so small an effect present a problem for a straightforward application of SDT. For the four conditions of preceding stimulus and response on trial $i - 1$, the probability of LOUD on trial i decreased in the order: $P(\text{LOUD}_i|soft_{i-1}\text{LOUD}_{i-1}) > P(\text{LOUD}_i|loud_{i-1}\text{LOUD}_{i-1}) > P(\text{LOUD}_i|soft_{i-1}\text{SOFT}_{i-1}) > P(\text{LOUD}_i|loud_{i-1}\text{SOFT}_{i-1})$. That is, we see both a positive effect of the last response and a negative effect of the last stimulus. Sandusky (1971) reported that a similar ordering applies to the data of Parducci and Sandusky (1965). Tanner, Rauk, and Atkinson (1970) and Long (1980) have reported similar results. Long (1980) also showed that the effect did not transfer from one modality to the other if visual and auditory judgments were alternated. Haller (1969) studied identification with three intensities, ranging from (an average of) 68.5 db to 70 db, and he too found that subjects tended to repeat the previous response.

A number of explanations for sequential effects have been put forward. Kac (1962) proposed an error-correcting model in which the subject's criterion changes only after an error. Thus, if the subject reports the signal when noise was in fact presented, his criterion rises. Dorfman and Biderman (1971) presented a generalization of this proposal, and Dorfman (1973) applied the term *additive learning model* to accounts of this class. However, to fit their data, Dorfman, Saslow, and Simpson (1975) found it necessary to add to their model the assumption that there are also random drifts in criterion between trials, and Kubovy and Healy (1977), using a numerical decision task as an analog of signal detection, obtained evidence against additive learning models. Thomas (1973) reviewed the literature up to that date and presented a version of the additive learning model in which the criterion is shifted by a quantity Δ following feedback, in a direction that tends to reduce errors, and this shift occurs with a probability determined by the values in a payoff matrix. All these models require that the subject receives feedback. Thomas (1975) examined ways of extending the model to situations without feedback. One useful suggestion was that the criterion shifts towards z_i (the sensory input on trial i), and the extent of the shift is related to the magnitude of $|z_i - z_c|$. Dusoir (1980) has reviewed and tested a number of additive learning models.

Parducci and Sandusky (1965) proposed a *memory state model* for the identification task (see also Sandusky, 1971) that required two criteria. If the input fell in an extreme position (to the left of the left criterion, or the right of the right criterion), the appropriate response would be made. If it fell between the criteria, it was compared with the preceding input: If the latter had been extreme the alternate response would be made, but if both the previous and the current inputs were in the central band, the previous response was repeated with probability v.

A *memory recognition model* has been applied to identification by Haller (1969), Tanner, Haller, and Atkinson (1967), and Tanner et al. (1970). In this case the current input is compared with the trace of the preceding input on every trial. The trace undergoes distortion in the intertrial interval. If the difference between the input and the trace is sufficiently extreme in either direction, the corresponding response is made, otherwise the previous response is repeated. Long (1980) was able to apply both memory state and memory recognition models to his data.

In the *absolute judgment* (category-scaling) task illustrated in Figure 1-4, there are more than two stimuli and more than two categories of response. When the number of response categories is equal to the number of stimuli, the procedure may be referred to as *absolute identification*. Formally, these are extensions of the two-category identification task, and we might expect similar findings. However, a number of analyses of this task have been made that examined dependencies over a number of preceding trials, and these have revealed some interesting features. In particular, the direction of the sequential effect may vary with the lag. Holland and Lockhead (1968) required absolute judgments of the loudness of 10 stimuli that varied in intensity over a range of 25 db, gave feedback, and found that the response on trial i showed assimilation toward the stimulus on trial $i - 1$, but contrasted with the stimuli on trials $i - 2$ to $i - 5$, or further back. They proposed that the stimulus and feedback on trial $i - 1$ were remembered and were used as a standard in judging the presented stimulus. Ward and Lockhead (1970) confirmed this pattern of dependencies in a similar task and also showed that if feedback were omitted, assimilation was now enhanced, extending back to trial $i - 2$, and contrast with stimuli on earlier trials was very much reduced.

In these studies previous stimuli are, of course, correlated with previous responses. Ward and Lockhead (1971) analyzed dependencies separately for previous stimuli and previous responses. In the response analysis, with feedback, assimilation to the response on trial $i - 1$ is similar to or a little greater than that shown for the stimulus analysis, but contrast of the current response with earlier responses appears reduced. When no feedback was given, the response analysis showed assimilation extending back four or five trials and no evidence of contrast. Ward (1973) examined sequential effects in magnitude estimations of the loudness of tones. In his procedure the response was required to represent the subjective ratio of the current stimulus to the one presented on trial $i - 1$. He reported assimilation to stimuli as much as five trials back, the effect tailing away with no appearance of contrast.

Jesteadt, Luce, and Green (1977) also examined absolute judgments and magnitude estimations of loudness. They applied a multiple regression analysis and concluded that sequential effects extend one trial back only. This conclusion has been challenged by Staddon, King, and Lockhead (1980), who argue that a linear model extending over only one trial cannot account for the sequence of contrast and assimilation effects and that mechanisms acting over several trials must be assumed.

Purks, Callahan, Braida, and Durlach (1980) applied a Thurstonian model like that illustrated in Figure 1-4 to absolute judgments of loudness with feedback. They estimated quantities b'_i that give the deviations of the criteria from the positions intermediate between the means of the successive stimulus distributions. Expressed as functions of the preceding stimulus, these indicate a positive sequential dependency: They are shifted to the right when the previous stimulus is less than the current one and to the left when it is greater. Because sensitivity was not improved when the preceding and the current stimuli were similar in intensity, Purks et al. reject the hypothesis that the previous stimulus is used as a standard in judging the current one.

To summarize, detection studies have usually, but not always, produced positive dependencies. Identification studies have shown negative effects when accuracy is considered (higher accuracy for stimulus alternations), but positive dependence on previous responses. Absolute judgments produce both positive and negative dependencies as a function of lag. We are faced with a somewhat bewildering variety of results.

Three Models

Despite a considerable amount of puzzling data and the exploration of several hypotheses, no single explanation has been established for even a major part of the data. One problem is the complexity and variety of the situations with which we are concerned. This suggests that it might be useful to investigate a very simple situation, involving as few independent variables as possible, and to test general, broadly defined models against the data resulting. We look at data obtained with the single intensity procedure of Verplanck et al. (1952) as this has the advantage of excluding both feedback and variation in stimulus intensity.

What models should we test? In choosing models, one may start with requirements for parsimony and generality: It seems desirable to frame simple theories, capable in principle of being extended to the whole range of observations, at least until a contrary necessity is demonstrated. We saw that three hypotheses about response dependencies emerged from the literature reviewed above: the memory state, memory recognition, and additive learning models.

The memory state model is excluded by our requirement that a general model should be capable of applying to a wide range of phenomena. This model, which postulates two criteria and a set of rules for their use, could not easily be adapted to tasks having more than two response categories.

Models, like the memory recognition model, that require the input on trial i to be compared with the trace of the input from trial $i - 1$ are also excluded on the grounds that they would be seriously inefficient. A comparison between a current sensory input and a fixed criterion is subject to the variance attaching to the current input, but a comparison between two successive inputs would be subject to the variance affecting both, doubling the total variance if they are uncorrelated. Only if the sensory system has no alternative should we expect it to perform in this way (Treisman & Leshowitz, 1969). Furthermore, this assumption cannot reasonably be applied to detection: It would predict that a series of presentations of the same stimulus intensity would give rise to a 50% detection rate, whatever the magnitude of the stimulus. It might be argued in reply to this that a fixed criterion may be used in detection when a background or a standard stimulus is available, but comparison with the previous trace when no standard is presented. But this would predict that the standard deviation of the psychometric function should increase by $2^{1/2}$ in the latter case. However, when both the methods of *constant stimuli* and *single stimuli* (in which no standard is presented) have been applied to the discrimination of lifted weights (Fernberger, 1931; Wever & Zener, 1928), taste discrimination (Pfaffmann, 1935), and temporal discrimination (Treisman, 1963), very similar standard deviations have been given by each procedure.

Three models that seemed to possess potential generality were selected for testing. The first two are simple generalizations of the additive learning model of Kac (1962), Dorfman and Biderman (1971), and Thomas (1973).

The Linear Additive Learning (LAL) Model

This model is described as it would apply to a detection or two-choice identification task without feedback. The basic idea is that the response on each trial causes an increment or a decrement to the criterion. We add here the further assumption that there is a reference level for any criterion and that when the current criterion is momentarily increased above or reduced below this reference level, it decays back towards it. To represent this more formally we use the following notation: z_o is the reference level of the detection criterion, $z_c(i)$ is the value of the criterion at the beginning of trial i and determines the response made on that trial, and $z'_c(i)$ is the final value taken by the criterion after it has been incremented or decremented as a function of the response made on that trial. The response on trial i is represented by R_i. If the response is YES, $R_i = 1$; if it is NO, $R_i = 0$. It is convenient also to define $J_i = 2(R_i - 0.5)$; then YES gives $J_i = 1$, NO gives $J_i = -1$. The model has a direction parameter, d, which may take the values 1 or -1. $d = 1$ will cause the model to produce a negative sequential dependency (contrast); $d = -1$ will determine a positive dependency (assimilation). The main assumptions of the model are as follows.

At the commencement of trial i the criterion has the value $z_c(i)$. The response causes a shift of magnitude Δ. The direction of the shift depends on the nature of the response made and the sign of the parameter d:

$$z'_c(i) = z_c(i) + dJ_i\Delta. \tag{1}$$

Thus if d is -1 a positive response will lower the criterion. This will favor repetition of the same response; that is, it will give a positive dependency.

During the intertrial interval the criterion decays toward z_o, but not beyond, at a constant rate δ. Then

$$z_c(i + 1) = \begin{cases} \max(z'_c(i) - \delta, z_o), & z'_c(i) \geq z_o \\ \min(z'_c(i) + \delta, z_o), & z'_c(i) < z_o \end{cases}. \tag{2}$$

That is, if the criterion is less than z_o it will increase by δ, but not beyond z_o; if it is initially greater, the reverse will happen. It follows that no sequential dependencies will be seen if $\Delta \leq \delta$. The model is illustrated in the left panel of Figure 2, for the single intensity procedure with $d = -1$. The responses on successive trials are shown on the ordinate, proceeding downwards. The continuous trace in the left panel shows how the criterion varies over trials and illustrates the effect each successive response has on it. On the initial trial $z_c(1)$ is at z_o, which is taken to be at the mean of the signal distribution. This mean is given the value 0. A positive response is made and this causes $z_c(1)$ to be reduced by $\Delta = 0.5$ (any latency in producing this effect is not shown). The criterion then drifts back toward z_o at a rate of $\delta = 0.2$ per intertrial interval, reaching

$z_c(2) = -0.3$ at the commencement of the second trial. The second response is again positive, the criterion again shifts to the left, then decays back to the reference level at a constant rate, and so on.

The Exponential Additive Learning (EAL) Model

Memory theorists have usually preferred to assume that memory traces decay exponentially (Wickelgren, 1967). Thus it is of interest to develop and examine a model with this feature. The EAL model was designed to be similar to the LAL model except for the mode of decay; it has the same number of parameters. The assumptions of this model are as follows.

The criterion value at the commencement of trial i is $z_c(i)$. The response causes a shift Δ to give

$$z'_c(i) = z_c(i) + dJ_i\Delta = z_o + K_i,$$

where z_o is the reference level of the criterion.

During any interval the deviation from z_o decays toward z_o exponentially. Thus if on trial i the deviation is $K_i = z'_c(i) - z_o$ and no further increments or decrements are added to the criterion, then k trials later this deviation will be reduced to $K_{i+k} = K_i e^{-rk}$, with the term r a constant. Thus, after one intertrial interval, K will be reduced in the proportion e^{-r}. For this model we define $\delta = 1 - e^{-r}$, so that δ

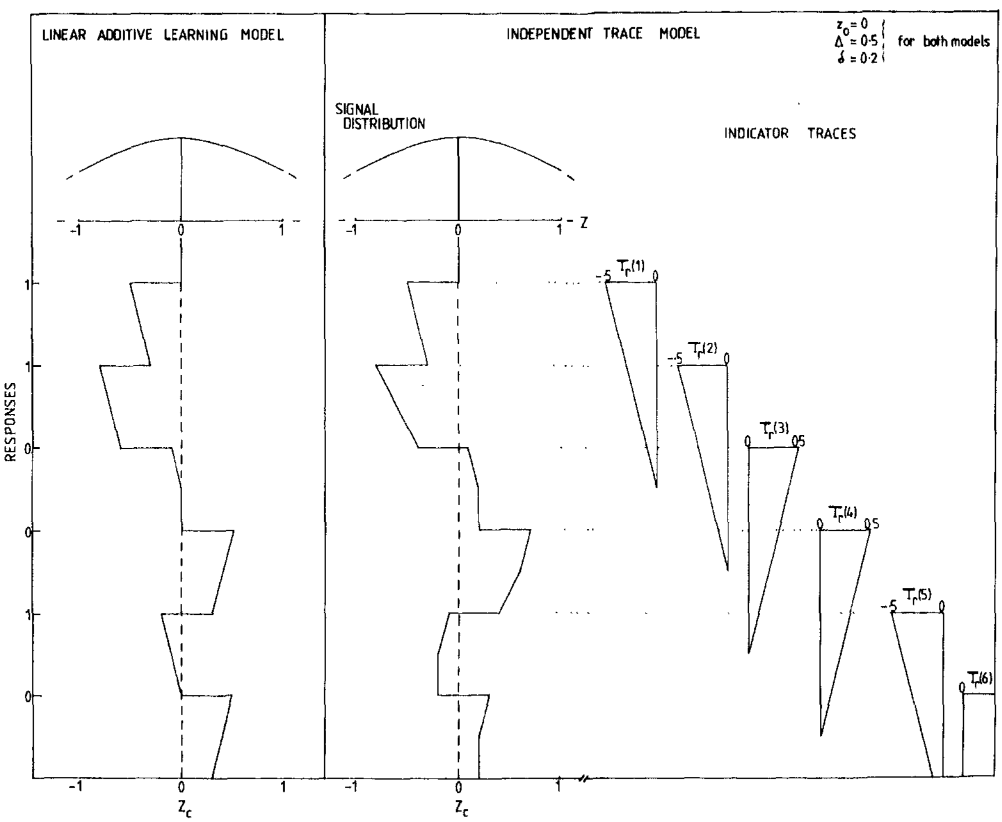


Figure 2. The Linear Additive Learning and Independent Trace models illustrated for the detection or identification procedures. (Changes in the criterion are followed for six trials. For this illustration the parameters have the values $z_o = 0$, $\Delta = 0.5$, and $\delta = 0.2$. The responses are indicated on the ordinate by 1 for YES and 0 for NO, and the criterion z_c commences at 0. For the Linear Additive Learning model each response produces a direct shift in z_c (these are shown as though the events of the trial were instantaneous), which then decays towards z_o at the rate δ. In the Independent Trace model each response produces a separate indicator trace $T_r(i)$ that decays toward 0 at the rate δ, and the value of z_c at any instant is given by the linear sum of the currently nonzero traces.)

represents the proportion of the initial deviation that is lost after one trial.

This model would give a trace similar to that shown for the LAL model in the left panel of Figure 2, save that each deviation decays toward z_o exponentially.

The Independent Trace (IT) Model

The third model shares with the LAL model the assumption that each response causes a shift in criterion and that shifts decay at a constant rate. But it differs in rejecting the assumption that the shift indicated by the response on each trial is immediately embodied in the current criterion. Instead it proposes that each shift is separately recorded and retained. That is, an instruction specifying the shift required by each response is laid down in the form of an independent memory or *indicator trace.* The linear sum of the independently retained traces from previous trials (those that have not yet decayed to zero) determines the value of the criterion at any time. Thus the shift indicated by each response is separately and independently retained and is independently effective, so long as it has not decayed away. These indicator traces are specially constructed quantitative records, which should not be confused with our normal memories of the actual responses. (It is conceivable that a subject may remember, in the sense that he can report, past responses that no longer have an effect on the current criterion, and vice versa.)

The response on trial i will create an indicator trace

$$T_r(i) = dJ_i\Delta. \tag{3}$$

Thus if the response is positive and d is -1, this would indicate a requirement for a reduction in criterion on subsequent trials.

The magnitude of an indicator trace decreases at a constant rate, δ, per intertrial interval, until it reaches 0. k trials after its creation the residual magnitude of the trace will be

$$T_r(i, k) = \begin{cases} \max(T_r(i) - k\delta, 0), & T_r(i) \geqslant 0 \\ \min(T_r(i) + k\delta, 0), & T_r(i) < 0 \end{cases}, \tag{4}$$

where i refers to the trial of creation, and k refers to the number of trials since elapsed. $T_r(i)$ is equivalent to $T_r(i, 0)$.

Thus, the trace will be nonzero for D terms, where

$$D = \begin{cases} \text{int}(\Delta/\delta), & \text{int}(\Delta/\delta) < \Delta/\delta \\ \text{int}(\Delta/\delta) - 1, & \text{int}(\Delta/\delta) = \Delta/\delta \end{cases}, \tag{5}$$

and $\text{int}(x)$ is the integral part of x. D is the number of terms over which the mechanism produces a sequential dependency and is referred to as the *depth* of the dependency.

The criterion on trial i is given by the reference level, z_o, plus the sum of currently nonzero indicator traces deriving from earlier responses:

$$z_c(i) = z_o + \sum_{h=i-D}^{i-1} T_r(h, i - h). \tag{6}$$

The model is illustrated in the right panel of Figure 2. For each of the first six trials the indicator trace formed on that trial is shown on the right, and on the left the resultant criterion given by their linear combination at any moment is traced out.

A Comparison of the Models

The models are as similar as possible in terms of the number of parameters and the assumptions made. In two we also have an assumption of linear decay toward the reference level. The models differ in the way the effects of past events are represented. The additive learning models use a single parameter, the value of the criterion at any time, to represent the net effect of past events. This may appear efficient in terms both of information processing and memory load: After each response there is a single immediate addition to the criterion, and no further information about the events on that trial need be retained. Both the additive learning models are nonlinear. This is evident for the exponential model but is also true for the linear model, because the processing of the effect on the criterion produced on a given trial is not independent of the outcomes on other trials. For example, suppose that on trial i the criterion is at its reference level, z_o. If $d = -1$ and a negative response is made, then the criterion will be shifted upward by Δ. One trial later, the residual effect of the response on trial i will be represented by a shift of $\Delta - \delta$. Suppose, however, that the criterion on trial i is initially at the value $(z_0 + \Delta)$. Following the negative response, the criterion will shift up to $(z_o + 2\Delta)$. One trial later the value of the criterion will have decreased to $(z_o + 2\Delta - \delta)$. This looks as though each component Δ has fallen by $\delta/2$ (or we may divide the fall in any other way between the two components that were summated to give the total shift: The situation is indeterminate). Thus the residual effect of the response R_i varies with its context (the position of the criterion when that response is made). Because the identity of each contribution to the criterion is lost, it also follows that we cannot define the depth, D, of a sequential effect for either additive learning model.

In contrast, the IT model is linear in that each trace decays separately in the same way, whatever the values and number of other traces retained. (The decay is only "piecewise linear", however, because it ceases when the magnitude of the trace reaches zero.) This imposes a burden on memory because the history of each trace must be separately followed. This raises the question: Can the model be formulated in a more simple way? Is it possible to substitute for the maintenance of a set of independent traces a suitably defined additive learning model with a single continuously updated summary trace that decays in accordance with an appropriate function such that these two models are formally equivalent? We can exclude this possibility by showing that no decay function can be defined that is both independent of the past history of the summary trace and can simulate the IT model exactly. If on trial i the summary trace representing the value of the criterion corresponds to a shift from the reference criterion of magnitude K, then any additive learning decay function must specify that this will decay to a new value K' on trial $(i + 1)$, whatever the sequence of preceding responses that built up the initial shift K. By hypothesis, that historical information is not available to an additive learning model. But if the IT model applies, we cannot specify a new value K' on trial $(i + 1)$ simply from our knowledge of the value of K on trial i. The rate of decay of the deviation from the reference criterion is indeterminate if we do not know the number and magnitudes of the indicator traces that sum to give K. If K represents the residual effect of one indicator trace, and if $K > \delta$, then K' will be equal to $K - \delta$, giving a decay of δ for this intertrial interval. If $K < \delta$, the net decay must be K, because $K' = 0$. If K is the sum of the residual effects of two positive indicator traces, each greater than δ, then $K' = K - 2\delta$. If K is the sum of $K_a < \delta$ and $K_b > \delta$, then the net decay will be $K_a + \delta$. Obviously, short of deploying the IT model with its collection of retained memory traces, such effects cannot be reproduced.

There is a resemblance between the IT model and the EAL model in that large deviations of the current criterion from z_o are likely to decay toward z_o more rapidly in both cases. For the IT model, this arises because a large deviation is more likely to represent the sum of a number of indicator traces pointing in the same direction, and the decay of their effect will represent the sum of their independent decay rates. We see this in Figure 2, for the IT model, in the difference in the rate at which the first and second deviations from z_o decay back toward the reference value. This resemblance between the IT and EAL models leads us to expect that they will give more similar matches to data than do the LAL and IT models.

However, the IT and EAL models are not simply discrete and continuous approximations to the same thing. Observe that for the IT model, for both the third and the fifth deviations from z_o shown in Figure 2, net decay is *away* from z_o, something that cannot occur with the EAL model.

The LAL and IT models differ in a way that may be considered an advantage for the latter model. The LAL model in its present generalized form—in which we do not include feedback-based error-correction—is subject to a latent instability. Suppose that the model is set to give a positive dependency ($d = -1$) and the subject responds NO. Then the criterion moves to the right, increasing the probability of further NO responses. There is no limit to the extent to which the criterion may shift from its reference value, given a series of repetitions of the same response, so that we may end with a grossly misplaced criterion likely to stay permanently out of range. This does not seem to happen in practice, which implies that this model would require additional mechanisms to prevent the criterion straying outside an acceptable range. The IT model, however, contains intrinsic bounds on variation in the criterion. Because $D\delta < \Delta \leqslant (D + 1)\delta$, it can readily be shown from Equation 6 that the maximum shift in one direction that can be induced by a series of repetitions of the same response, however long the sequence, cannot be greater than $D(D + 1)\delta/2$. Once the parameters Δ and δ are set, the possible range of variation of the criterion is prescribed.

An advantage of the IT model over both additive models is that it allows different sets of traces, governed by different parameters, to be active at the same time, which may allow the modeling of complex effects. This point is developed later.

A Test of the Models: Howarth and Bulmer's (1956) Data

Howarth and Bulmer (1956) reported results that can be used to compare the models. They used the single intensity procedure: Every 4 sec the same stimulus, a half-sec light flash, was presented at an intensity intended to give approximately 50% detection. Their subjects did not know that the signal was invariant and made detection responses, YES or NO. The authors report response probabilities following different preceding response sequences.

The probability of response R_i on trial i, following a given preceding sequence of three responses, is written $P(R_i|R_{i-3}R_{i-2}R_{i-1})$. Thus $P(1|001)$ is the probability of YES given that the immediately preceding response was YES and that the two responses before that were both NO. We can interpret these data as referring to changes in the position of a criterion (cf. Figure 1-2). The standardized normal deviate corresponding to the conditional probability of the NO response gives the position of the criterion on a standardized decision axis with its origin at zero (corresponding to M_s) and a standard deviation of unity:

$$z_c(R_{i-2}R_{i-1}) = z[1 - P(1|R_{i-2}R_{i-1})] = (E_c(R_{i-2}R_{i-1}) - M_s)/\sigma_s, \quad (7)$$

where $z[P]$ is the value of c such that

$$\int_{-\infty}^{c} N(0, 1)dx = P,$$

E is the unstandardized decision axis, and E_c is the criterion on that axis (Treisman, 1976).

The individual results for Howarth and Bulmer's (1956) 10 subjects were read from their graphs, averaged, and expressed as standardized deviates. These values are plotted in Figure 3.

The preceding sequence obviously produces a large positive dependency. The results for homogeneous preceding sequences, shown on the left, suggest that the further back we define a response, the less effect it has. Thus the difference in z_c between 0 and 00 is greater than that between 00 and 000. But it is not clear from this graph whether or not we can disregard effects extending more than three trials back.

The points on the left are correlated because the shorter sequences include the longer, but those in the right-hand panel, for all possible preceding three-term sequences, are independent. They suggest a tendency to linearity of criterion placement, except for a kink at the center. The relation of these criterion positions to the signal distribution is illustrated in an inset.

Although decay occurs at a constant rate in the linear models, it is limited by a threshold, so in this respect these models are nonlinear. To fit each model, a parameter-estimation program employing the simplex procedure was used to estimate the parameters, z_o, Δ, and δ. In each case the program commenced with z_c equal to z_o and determined successive values of z_c for a sequence of seven trials. It recorded z_c on the seventh trial, contingent on each possible sequence of responses for the three preceding trials, and attempted to minimize the squared deviations from the data. This analysis is adequate for the IT model provided $D \leqslant 6$. For the additive learning models very similar results were given whether four trials or seven trials were considered.

As a measure of goodness of fit, the mean squared deviation between observed and predicted values was found:

$$MSD = \sum_{i=1}^{n} (O_i - P_i)^2/n,$$

and its square root was defined as the standard error of prediction, SE_p. The best fits are shown in the right panel of Figure 3 as a continuous line for the IT model, a dashed line for the LAL model, and a dotted line for the EAL model. The same parameters were used to fit the results for homogeneous series on the left. The parameters and measures obtained are given in Table 1. Although the parameters for the IT and LAL models are quite similar, the IT model gives a markedly better fit. If we take the deviations from this model to represent error variance alone, the ratio between the *MSD*s would give $F(5, 5) = 12.34$, $p < .01$, a significantly worse fit for the LAL model. The EAL model also does worse than the IT model, but not significantly so.

The good fit given by the IT model may seem surprising. It might be expected that an efficient system would use a single parameter to represent the net effect of past events rather than storing each trace separately. While the EAL model does less well, the difference is not great. A further test of the models is given by the following single intensity experiment.

Experiment 1

Method

Subjects performed a detection task in which an auditory signal of the same intensity was presented in white noise on every trial.

Apparatus

An 830-Hz signal from an Advance Instruments J2E Oscillator was mixed with white noise produced at approximately 70-db SPL by a Dawe 419C white noise generator and presented to the subject through Eagle International SE5 headphones. Both signal and noise were presented binaurally. The white noise was continuously present.

In front of the subject was a box on which two lights, one amber and one green, and two response keys, labeled *tone* and *no tone*, were mounted. Behavioural Research and Development (BRD) logic modules were used to con-

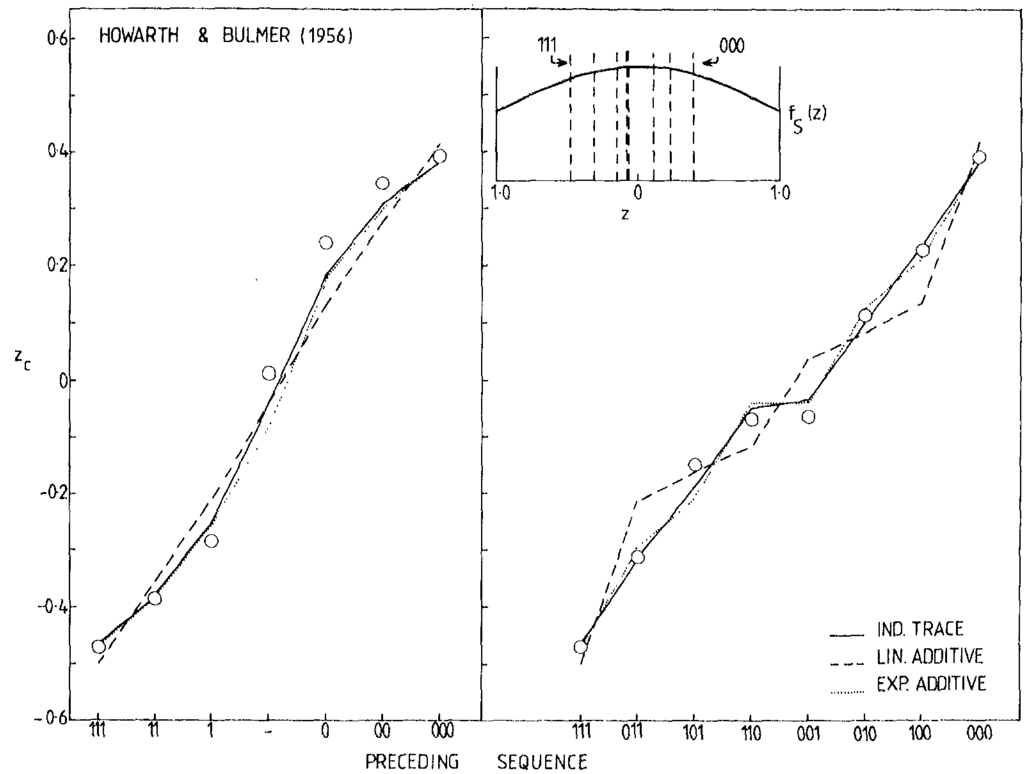


Figure 3. Data from Howarth and Bulmer (1956). Detection probabilities contingent on different sequences of prior responses were converted to the corresponding standardized normal deviates (z_c), and these are plotted as a function of the preceding sequences, arranged to give a monotonic order. (The left panel shows homogeneous preceding sequences of lengths up to three, the right shows all possible three-trial sequences. An inset of the signal distribution, $f_S(z)$, shows the positions of the three-trial criteria on the decision axis z. Theoretical curves have been fitted as described in the text.)

Table 1
Estimated Parameters for Three Models Fitted to the Data of Howarth and Bulmer (1956)

Model	z_0	Δ	δ	D	MSD	SE_P
LAL	−0.035	0.184	0.064	–	0.00421	0.0649
EAL	−0.039	0.352	0.409	–	0.00064	0.0253
IT	−0.041	0.276	0.068	4	0.00034	0.0185

Note. LAL = Linear Additive Learning; EAL = Exponential Additive Learning; IT = Independent Trace.

trol the experimental sequence on each trial and to record responses. A trial began with the amber light coming on. After 0.5 s the green light came on as well, and both lights remained on together for 0.45 s, after which both amber and green lights were extinguished. The signal was gated on 100 ms after the onset of the green light. It lasted 250 ms. There was a 2.98 s intertrial interval, during which the subject made his response, then the next trial commenced.

Procedure

There were two conditions: *signal only* and *noise only.*

Signal only. In this condition every trial had a signal (of the same constant intensity for a given subject) presented in white noise. There were no trials with noise alone. Seven subjects were tested in this condition. Each subject attended individually for one practice and two experimental sessions, on separate days. On the practice day, the intensity of the signal that gave 50% detections for that subject was found, and this intensity was used in the two experimental sessions. In each experimental session the subject received four blocks each of 158 trials, separated by 5 min intervals. The same intensity of the signal was presented on every trial, although the subject did not know this.

The first 8 trials of each block were discarded; 1200 experimental trials were analyzed for each subject.

Noise only. In this condition no pure tone signal was given on the trials later analyzed, although the subject did not know this. On these trials only the white noise was presented. Four subjects were run in this condition. The practice session was conducted as above. On each of the two experimental sessions, each subject received five blocks of 158 trials. The first block, as well as the first eight trials of each of the remaining four blocks, was discarded. In the first block the signal intensity found in the practice session to give $P(Y) = .5$ was presented on every trial. In the remaining blocks, the first trial always contained the signal at the intensity giving $P(Y) = .5$. This intensity was reduced on each of the next four trials, and the signal was absent on all the remaining trials in the block. Thus 1,200 noise-only trials were analyzed for each subject.

Subjects

The subjects were 11 students between the ages of 19 and 33. They were told that they should expect to make a detection on about one half of the trials.

Results

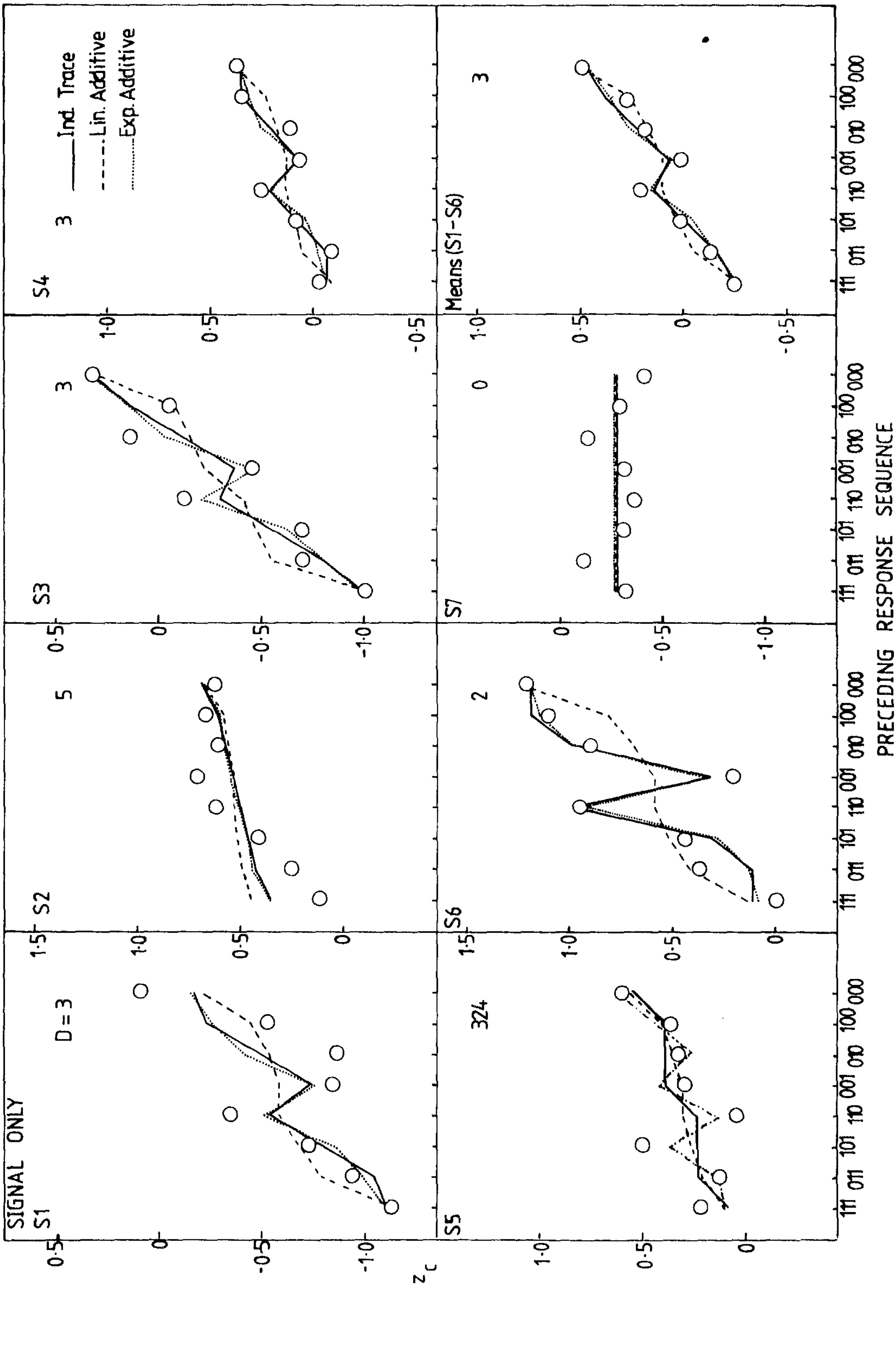
Subjects had no difficulty in performing the tasks. (The results are shown in Figures 4 and 5 as criterion values conditional on preceding response sequences. Thus $z_c = .5$ corresponds to $P(Y) = .31$, etc.) In the signal-only condition the subjects did not discover that the signal was always present at constant intensity; they maintained their criteria at values that determined responses usually not too far from a 50% hit rate overall (see Figure 4). In the noise-only condition the subjects did not realize that the signal had been removed in the experimental blocks and adjusted their criteria to give false positive rates that were not too far from 50% overall for N1, N2, and N3, but somewhat lower for N4 (see Figure 5).

For each subject the conditional criterion values $z_c(R_{i-3}R_{i-2}R_{i-1})$ were found, and the parameter-fitting program was used to fit each of the models to these data. To do so it minimized the squared deviations between observed and predicted values of z_c, each weighted by the number of occurrences of that type of preceding sequence, $N(R_{i-3}R_{i-2}R_{i-1})$. However, the unweighted *MSD* is reported as the measure of fit. Each model was fitted once with $d = 1$ and once with $d = -1$, and the value giving the lowest unweighted *MSD* is reported. The results are summarized in Table 2. Here *Seq* refers to the direction of the sequential dependency: A positive dependency (assimilation) corresponds to $d = -1$. If the best fit was given with parameters generating no sequential dependency, Seq = 0. For each model Subjects S7 and N2—and for the LAL model N1—show no sequential dependency.

The models may be compared in four ways.

1. *Goodness of fit for individual subjects.* We first compare the IT and LAL models. For each model, the best fits for Subjects S7 and N2 show no sequential dependency and are identical. This leaves 9 subjects who may be compared. For 7 of the 9, SE_p is less for the IT model than for the LAL model. This result favors the IT model, though not significantly. Its two-tailed probability is .18. For the IT and the EAL models, there are eight cases in which SE_p differs between them. In six cases it is less for the IT model ($p = .29$). The fits given by the models are shown in Figures 4 and 5.

2. *Consistency of the models when the same*


SIGNAL ONLY
S1
D = 3
S2
5
S3
3
S4
3
Ind. Trace
Lin. Additive
Exp. Additive
S5
324
S6
2
S7
0
Means (S1–S6)
3
z c
PRECEDING RESPONSE SEQUENCE
111 011 101 110 001 010 100 000

data are analyzed in different ways. (By *consistency* we mean that if a model is fitted to data and parameters obtained, the data reordered and the process repeated, a correct model should deliver similar parameters each time. To take a simple-minded instance, suppose that an experimenter knows the individual heights of 10 subjects in Group A and 20 subjects in Group B and the mean for the population, but he or she does not know how these are related. A model is applied in which the sums of the heights in each group are divided by some number; the answer for each group is weighted; and these are added together. The experimenter obtains a good fit if he or she then divides by 10 and 20 for the two groups, respectively, and weights the two answers by ⅓ and ⅔. If the subjects are then randomly reshuffled into new groups of 10 and 20, and the process repeated, the experimenter is likely to obtain the same parameters, subject to random errors of measurement. This model is consistent. But suppose that the experimenter tries to fit a different model in which he or she first finds the products of the heights within each group, then divides each product by a parameter, and then weights and sums the two results. The experimenter will obtain some solution for his or her initial data. But if the subjects are reshuffled and the process repeated, it is highly unlikely that the same parameters will give a fit the second time. The model is wrong and the experimenter will find it to be inconsistent.)

The results analyzed above are the probabilities of a positive response following all possible sequences of three preceding responses. To examine the consistency of each of the present models we can compare the parameters obtained when the models are independently fitted to three- and four-term data. The parameters Δ and δ, as defined for each model and as estimated from three- and four-term data, are given in Table 3 for those subjects showing sequential dependencies with each model.

For the IT model, the three- and four-term estimates of Δ and δ differed by less than 4% in 8 (out of 18) comparisons; for the EAL model, the smallest disparity in 16 comparisons was 4.4%; and for the LAL model it was 21.5%.

To test the significance of these differences, the following procedure was followed. For each subject and model the absolute differences between the three- and four-term parameters $|\Delta_3 - \Delta_4|$ and $|\delta_3 - \delta_4|$ were found. Taking these data together for the IT and LAL models, the median of all 34 values is .090. For the IT model, 15 (out of 18) results were less than this median, and for the LAL model two (out of 16) differences were less than the median. Fisher's exact probability test (Siegel, 1956) applied to these results gives $p = 0.00004$, a decisive rejection of the LAL model. For the IT and EAL models, 12 (out of 18) IT results and 6 (out of 16) EAL results are below the median for the combined data (.0425). This result borders on significance ($p = .0872$).

3. *Nonmonotonicity under certain conditions.* The IT model predicts nonmonotonicity (in relation to the order of conditions that has been used in presenting the three-term data in Figures 4 and 5); the LAL model does not. This provides a further test between these two models. If the IT model holds and the sequential dependency is positive, we may write

$$z_c(110) = z_o + \min(-\Delta + 3\delta, 0) + \min(-\Delta + 2\delta, 0) + \max(\Delta - \delta, 0),$$

and

$$z_c(001) = z_o + \max(\Delta - 3\delta, 0) + \max(\Delta - 2\delta, 0) + \min(-\Delta + \delta, 0), \quad (8)$$

and it is then easy to show that for $\Delta/\delta > 1$,

$$z_c(110) \lesseqgtr z_c(001) \text{ iff } \Delta/\delta \gtreqless 4.$$

For $\Delta/\delta \leqslant 1$, $z_c(110) = z_c(001)$. That is, for D ranging from 1 to 3 only, the ascending curves shown in Figure 4 should dip between

Figure 4. Experiment 1. Individual results for three-term-preceding sequences in the signal-only condition. (The Independent Trace [continuous lines], Linear Additive Learning [dashed lines], and Exponential Additive Learning [dotted lines] models have been fitted in each case, except that the Exponential Additive Learning fit for Subject S5 is not shown (it was very similar to the Independent Trace fit). The dash–dot curve shown for Subject S5 is discussed in the text. Fits to the unweighted mean data for Subjects S1–S6 are also shown. The parameter D is shown for each curve.)

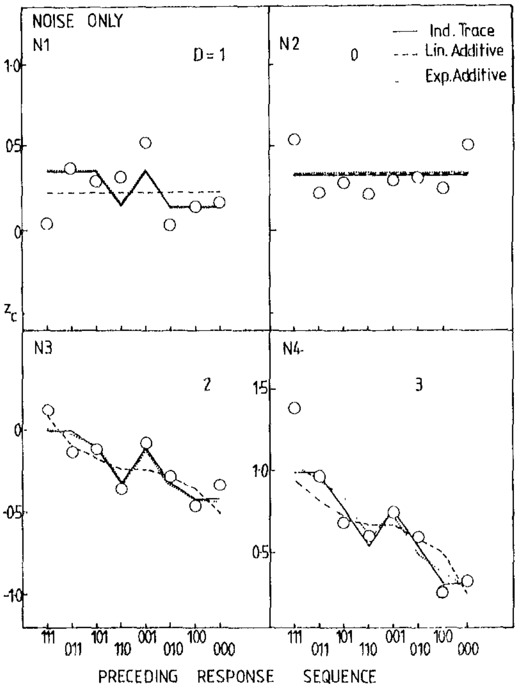


Figure 5. Individual results for the noise-only condition, plotted as in Figure 4.

the conditional criterion values for 110 and 001. If the sequential dependency is negative, as for some subjects in Figure 5, we may invert this prediction.

The IT prediction that $z_c(110)$ will be greater than $z_c(001)$ for $1 \leqslant D \leqslant 3$ for a positive dependency (and the reverse for a negative dependency) is of interest because at this transition the LAL model predicts monotonicity. Of course, it could be argued that departures from monotonicity at this point (as may be the case at other points in the curves where they occur) are due to random noise. However, we can test the IT prediction against the null hypothesis that the relation between $z_c(110)$ and $z_c(001)$ is wholly determined by noise and, therefore, equally likely to go in either direction.

For the 9 subjects who show nonzero sequential dependencies with the IT model, the obtained relation between $z_c(110)$ and $z_c(001)$ is in every case in agreement with the predictions above, for the value of D and the direction of the sequential dependency pertaining to each case. That is, the tendency of the data is nonmonotonic at this point for $D < 4$ (Subjects S1, S3, S4, S6, N1, N3, and N4), monotonic for $D > 4$ (Subjects S2 and S5). The two-tailed probability of so extreme a result occurring as a consequence of chance noise is $p = 0.0039$.

If we restrict ourselves to the seven cases in which the IT model predicts a reversal in tendency ($1 < \Delta/\delta < 4$) and the LAL model does not (as can be seen from the dashed lines), we have a reversal in each case. This has the two-tailed probability $p = .0156$.

As can be seen from Figures 4 and 5, the EAL model, when fitted to these data, is capable of reproducing the nonmonotonicities that are predicted by the IT model. Thus this comparison cannot be used to distinguish between these two models.

4. *Predictability of the numbers of occurrences of each possible response sequence in the data.* We have obtained the parameters that best predict response probabilities conditional on preceding response sequences. A powerful test is given by using these parameters to predict $N(R_{i-3}R_{i-2}R_{i-1})$, the actual number of three-term sequences of each type to be found in each subject's data. These measures were not used in obtaining the parameters. To make these predictions, each model is used to find z_c at each stage for every possible sequence of six responses and from this to calculate the probability of that sequence. These probabilities are combined to give the probabilities of every possible sequence of the final three responses. Chi-square may then be used to compare the predicted and obtained numbers of each sequence type; if the model predicts these numbers correctly, chi-square should not be significant.

The values of chi-square (each with 4 *df*s) are shown in Table 2. They may be summed over all subjects. We then get, for the IT model, $\chi^2(44, N = 11) = 30.76$, which is nonsignificant. For the LAL model we get $\chi^2(44, N = 11) = 216.19, p < .0001$. This rejects the LAL model, but the result is completely consistent with the IT model. This test does not enable us to distinguish between the IT and EAL models. For the latter we get $\chi^2 = 31.60$, which is again nonsignificant.

Discussion

These comparisons give us a clear result for the two models with linear decay. The Howarth and Bulmer (1956) data rejected the LAL model ($p < .01$). Comparisons of goodness of fit for the present data support the IT rather than the LAL model although not significantly ($p = .18$). The test of consistency when fitting three- and four-term data strongly rejects the LAL model ($p = .00004$). The agreement between predicted and obtained nonmonotonicities for the curve segment from $z_c(110)$ to $z_c(001)$ supports the IT model ($p = .0039$).

Table 2
Parameters for the Three Models

Condition and subject	Parameter							
	Seq	z_0	Δ	δ	D	*MSD*	SE_P	χ^2
			Independent Trace					
Signal only								
S1	+	−0.640	0.410	0.127	3	0.0459	0.214	1.512
S2	+	0.510	0.091	0.018	5	0.0177	0.133	5.027
S3	+	−0.334	0.470	0.126	3	0.0212	0.146	9.594
S4	+	0.137	0.209	0.069	3	0.00172	0.0415	0.285
S5	+	0.279	0.065	0.0002	324	0.0191	0.138	5.186
S6	+	0.655	0.777	0.339	2	0.0153	0.124	2.081
S7	0	−0.279	$\Delta < \delta$		0	0.00946	0.0973	1.124
M (S1–S6)	+	0.106	0.273	0.077	3			
Noise only								
N1	–	0.247	0.263	0.159	1	0.0220	0.148	2.089
N2	0	0.348	$\Delta < \delta$		0	0.0147	0.121	2.786
N3	–	−0.221	0.269	0.111	2	0.00579	0.0761	0.034
N4	–	0.651	0.337	0.112	3	0.0220	0.148	1.039
			Linear Additive Learning					
Signal only								
S1	+	−0.587	0.517	0.399	–	0.0453	0.213	33.635
S2	+	0.529	0.227	0.195	–	0.0285	0.169	7.390
S3	+	−0.288	0.268	0.106	–	0.0373	0.193	69.998
S4	+	0.131	0.215	0.150	–	0.00756	0.0870	3.634
S5	+	0.300	0.221	0.156	–	0.0184	0.136	5.047
S6	+	0.577	0.557	0.418	–	0.0551	0.235	67.393
S7	0	−0.279	$\Delta < \delta$		–	0.00946	0.0973	1.124
M (S1–S6)	+	0.100	0.392	0.290	–			
Noise only								
N1	0	0.288	$\Delta < \delta$		–	0.0257	0.160	18.810
N2	0	0.348	$\Delta < \delta$		–	0.0147	0.121	2.786
N3	–	−0.243	0.325	0.223	–	0.0104	0.102	3.101
N4	–	0.680	0.503	0.391	–	0.0380	0.195	3.268
			Exponential Additive Learning					
Signal only								
S1	+	−0.625	0.734	0.597	–	0.0494	0.222	2.248
S2	+	0.494	0.090	0.243	–	0.0182	0.135	5.345
S3	+	−0.324	0.859	0.553	–	0.0112	0.106	7.622
S4	+	0.136	0.394	0.644	–	0.00368	0.0606	0.867
S5	+	0.288	0.080	0.098	–	0.0192	0.139	4.665
S6	+	0.639	1.909	0.773	–	0.0143	0.120	2.861
S7	0	−0.279			–	0.00946	0.0973	1.124
M (S1–S6)	+	0.103	0.419	0.515	–			
Noise only								
N1	–	0.246	22.756	0.995	–	0.219	0.148	2.089
N2	0	0.348	–	–	–	0.0147	0.121	2.786
N3	–	−0.224	0.514	0.690	–	0.00604	0.0777	0.019
N4	–	0.679	0.483	0.541	–	0.0223	0.149	1.971

The prediction of numbers of response sequences strongly rejects LAL ($p < .0001$). This evidence seems sufficient to drop the LAL model from further consideration.

These results suggest that the linear combination of independent indicator traces may provide a basis for understanding sequential effects. It is also of interest to note the similarity between the IT parameters obtained from the present results and those given by the data of Howarth and Bulmer (1956), despite the differences in the experimental situations. Compare the IT parameters Δ and δ fitted to the mean data for Subjects S1 to S6 with those obtained from Howarth and Bulmer's (1956) mean data (Table 1). There is far less similarity for the other two models.

The comparison between the IT and EAL models is far less clear. No individual test is significant. Whichever of the two models is more nearly correct, it is evident that the other model does a very good job of simulating it. However, we shall defer trying to decide between these two models until some further evidence has been presented.

Is Decay Linear or Exponential?

A main difference between the IT and EAL models lies in the assumptions they make

Table 3
A Comparison of Parameters Derived From Three-Term and Four-Term Sequences for Models

Condition and subject	Δ_3	Δ_4	$\lvert\Delta_4 - \Delta_3\rvert$	δ_3	δ_4	$\lvert\delta_4 - \delta_3\rvert$
			Independent Trace			
Signal only						
S1	0.410	0.281	0.129	0.127	0.056	0.071
S2	0.091	0.325	0.234	0.018	0.257	0.239
S3	0.470	0.419	0.051	0.126	0.099	0.027
S4	0.209	0.211	0.002	0.069	0.070	0.001
S5	0.065	0.112	0.047	0.0002	0.021	0.021
S6	0.777	0.793	0.016	0.339	0.352	0.013
Noise only						
N1	0.263	0.287	0.024	0.159	0.183	0.024
N3	0.269	0.271	0.002	0.111	0.111	0
N4	0.337	0.347	0.010	0.112	0.114	0.002
			Linear Additive Learning			
Signal only						
S1	0.517	0.669	0.152	0.399	0.569	0.170
S2	0.227	0.320	0.093	0.195	0.290	0.095
S3	0.268	0.582	0.314	0.106	0.439	0.333
S4	0.215	0.390	0.175	0.150	0.331	0.181
S5	0.221	0.379	0.158	0.156	0.318	0.162
S6	0.557	0.454	0.103	0.418	0.344	0.074
Noise only						
N3	0.325	0.412	0.087	0.223	0.318	0.095
N4	0.503	0.618	0.115	0.391	0.511	0.120
			Exponential Additive Learning			
Signal only						
S1	0.734	0.509	0.225	0.597	0.463	0.134
S2	0.090	0.050	0.040	0.243	0	0.243
S3	0.859	0.808	0.051	0.553	0.534	0.019
S4	0.394	0.329	0.065	0.644	0.580	0.064
S5	0.080	0.103	0.023	0.098	0.227	0.129
S6	1.909	1.821	0.088	0.773	0.763	0.010
Noise only						
N1	22.756	16.755	6.001	0.995	0.994	0.001
N3	0.514	0.559	0.045	0.680	0.712	0.022
N4	0.483	–	–	0.541	–	–

about the loss of information: Is decay linear or exponential? We consider two direct tests.

The first test is based on the present data. For the IT model we rewrite Equation 6 as (for three preceding terms)

$$z_c(i) = z_o + dJ_{i-1}w_{i-1} + dJ_{i-2}w_{i-2} + dJ_{i-3}w_{i-3}, \quad (9)$$

where

$$w_{i-k} = \begin{cases} \Delta - k\delta, & \Delta > k\delta \\ 0, & \Delta \leqslant k\delta \end{cases}. \quad (10)$$

That is, the residual effects of past criterion deviations are represented as weights assigned to the responses that generated them. We can now find z_o and these weights w_{i-k} by multiple regression on sequences of length 3. If decay is linear and the depth of the effect is 3 or greater ($D \geqslant 3$), then the weights should be a decreasing linear function of the lag, k. But if $D < 3$, then the weights should decrease to a floor at zero after one or two intertrial intervals and then stay constant at that level, giving a curve resembling exponential decay.

Thus for the IT model we predict that (a) for subjects with $D \geqslant 3$, the weights will decrease linearly as lag increases from 1 to 3 and (b) for $D < 3$, the weights will decline more sharply as lag increases from 1 to 2 than from 2 to 3.

For any subject we can draw a straight line connecting the weight at lag 1 with the weight at lag 3. If $D \geqslant 3$, the first prediction says that the weight at lag 2 should lie on this line. If $D < 3$, the weight should lie below this line. In testing such a prediction there is a problem in applying the description *on the line* because small random divergences may move points up or down. Because it is points on this line that distinguish between the IT and EAL models, it is a conservative procedure to class all points below the line (however little displaced from it) as *below*, and only those identically on or above the line as *on*.

The alternative to linear decay envisages that each effect produced on the criterion is caught up in an exponential decay process. If so, we should expect that in all cases the current weight assigned to a past response should decrease more rapidly as we go from lag 1 to 2 than from lag 2 to 3. Thus exponential decay predicts that in all cases the weight at lag 2 will be below the line joining w_{i-1} and w_{i-3}.

To test these predictions the best-fitting weights were determined for each subject by multiple regression, with $d = -1$ (a positive dependency) for the signal-only condition and $d = 1$ for noise only. The linear equation $w_{i-k} = \Delta_w - \delta_w k$ was fitted to the weights so found, and the estimated value δ_w was used to transform the weights to

$$w'_{i-k} = w_{i-k}/\delta_w + 3 - w_{i-1}/\delta_w. \quad (11)$$

This linear transformation has the effect that each function now takes the same values at lags 1 and 3, so that the distribution of points at $k = 2$ indicates the general form of the decay function. If it is linear (and $D \geqslant 3$) these points will be scattered about the straight line running through the values at $k = 1$ and $k = 3$. If $D < 3$ or if decay is exponential, these points will lie below the line.

Figure 6 shows these values for all subjects except S7 and N2, who did not show dependencies, and S5, whose weights were markedly nonmonotonic. We see that one half of the points lie above the line and one half lie below the line. Moreover, all the points above the line, but only one below it, have $D \geqslant 3$. This result is markedly inconsistent with the prediction based on exponential decay. How well does it match the IT prediction?

The scattering of points about the line, in Figure 6, could have arisen from a random process, in which case any given weight at Lag 2 would have the same probability of lying above or below the line as any other. Or it could express the pattern predicted by the IT model. To test this possibility, we classify each point that is on or above the line, and for which $D \geqslant 3$, as a hit and each such point that is below the line as a miss, and do the reverse for $D < 3$. This gives seven hits and one miss. The binomial probability of this pattern being produced by chance (one-tailed, because direction is predicted) is $p = .035$.

We now turn to a second test of the nature of trace decay. This rests on the application of the present approach to tasks in which contingent aftereffects occur or in which anchors are used (Treisman, in press).

If we could present a stimulus that caused the criterion to shift and could then observe the criterion as it drifted back to its resting

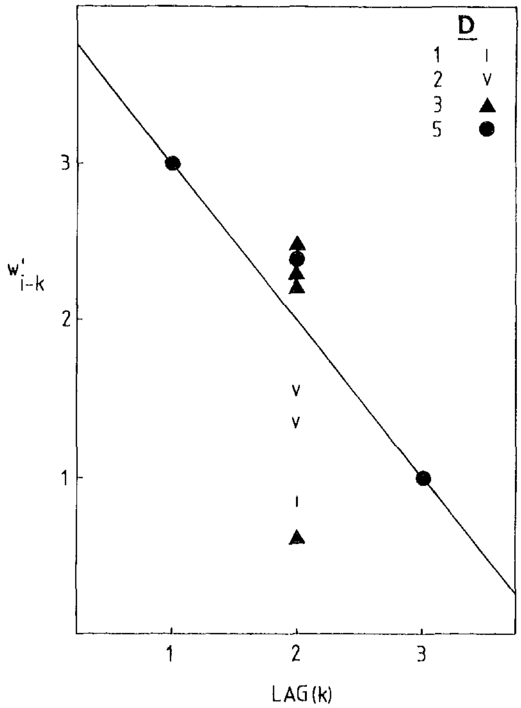


Figure 6. Individual weights attaching to past responses at three lags were calculated for 8 subjects. (These weights were subjected to the linear transformation described in the text, which causes them to lie along the same line and deviate from it only at lag 2. The transformed weights are plotted. The points at $k = 2$ should all lie below the line if decay is exponential [or $D < 3$], near the line if decay is linear and $D \geqslant 3$.)

value, we could examine directly whether the decay is linear or exponential. There is a procedure that may allow us to do this. In this paradigm two auditory stimuli, such as 1000 Hz and 3000 Hz (we refer to these as the *standard stimuli*), are presented in sequence with an interval between them, such as 1.5 s. On each presentation of this cycle an intermediate tone (the *bisector tone*) is also given. This occurs at an interval t_1 after the first standard stimulus. Thus the subject hears a sequence of three brief tones separated by two intervals: "1000 Hz standard–t_1–bisector tone–t_2–3000 Hz standard." The subject's task is to adjust the bisector tone over trials until he or she is satisfied that it bisects the pitch interval between the two standard tones.

This procedure was introduced by Cohen, Hansel, and Sylvester (1954) to investigate an intermodal effect on judgment, the *auditory tau effect*. This effect can be explained in terms of the ideas presented here; the explanation is illustrated in Figure 7. (For a fuller discussion of this effect and the application of the ideas developed in this article to adaptation effects and contingent aftereffects, see Treisman, in press.) The effects of the standards are represented by central dispersions on a decision axis, E_F, and an intermediate criterion, E_c, identifies the point at which the distance between them is subjectively bisected. When a standard is presented, the subject identifies it and sets up a corresponding indicator trace. If the lower standard is presented, this indicator trace prescribes an upward shift, and vice versa. (This represents a positive sequential effect. To anticipate a later development in this article, it corresponds to a tracking strategy.)

In Figure 7 two possible sets of indicator traces are shown, which correspond to different values of the parameters Δ and δ. In Case 1, the indicator trace does not decay to zero until the cycle is complete. Thus if we could observe E_c during the cycle, we would see it return linearly to its initial value, as indicated by the continuous line. In Case 2, the indicator traces return to zero before the cycle is completed. In this case, if we could observe E_c (the dashed line), its course would be difficult to distinguish from exponential decay.

When a bisector tone, coming at a fixed interval, t_1, after the first standard (1000 Hz in the figure), is repeatedly adjusted until the subject considers it to bisect the interval between the standards, we may take it that the central sensory effect of the final value of the bisector approximately coincides with the position of E_c at the interval t_1 after the first standard. (For the signal detection model of the method of adjustment on which this is based, see Treisman, 1969.) Thus if we plot the bisector frequency as a function of time after the first standard, this will provide an indication of the form of decay of the criterion shift induced by the previous standard stimulus. If the IT model is correct this will give us a linear shift in criterion (for Case 1) or an ambiguous result (Case 2). However, if shifts decay exponentially we should not expect to see linear decay.

Relevant information is provided by data obtained by Cohen, Hansel, and Sylvester

(1954) in which observations were made for three values of t_1 and four pairs of standards, and by results of Christensen and Huang (1979) for nine values of t_1. These data are replotted in Figure 7, and the best-fitting straight line is shown for each set of data. Each case illustrates a positive sequential effect. For the data of Cohen et al. (1954) it is immediately evident that in three cases the bisector frequency varies linearly with t_1. The fourth case (first standard 4000 Hz–second standard 2000 Hz) is consistent with Case 2 illustrated above. For the data of Christensen and Huang (1979) the linear trend was significant ($p < .001$), but the quadratic and cubic trends were not.

These observations provide strong evidence that decay is linear and not exponential.

Independent Traces or Exponential Additive Learning?

The IT and EAL models give such similar results in many cases that individual tests may not show a significant difference between them. Nevertheless, there is not one case in our series of tests in which the EAL model does better

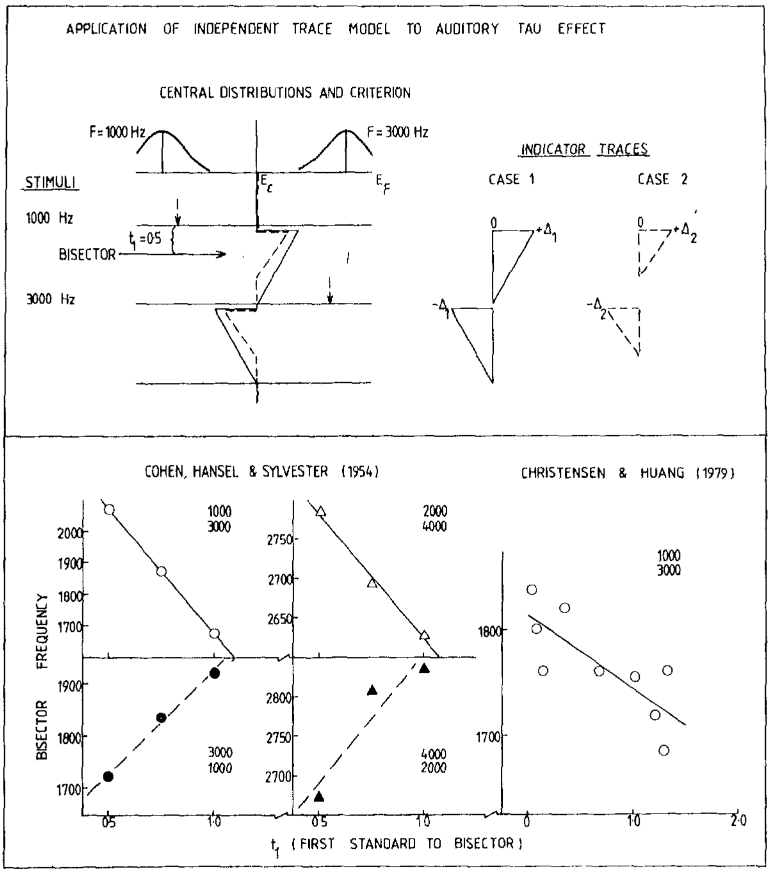


Figure 7. Application of the IT model to the auditory tau effect. (The upper panel applies the IT model to an auditory bisection paradigm: The subject adjusts a bisector tone presented between two successive standards, 1000 and 3000 Hz, occurring at an interval of 1.5 s. The effects of the standards on the intermediate criterion E_c are followed for two trials. Presentation of a standard causes an indicator trace to be set up: Two possible magnitudes of indicator trace, and the corresponding changes in E_c, are shown. Time proceeds downward and the decision axis E_F is repeated for each trial. The lower panel shows data obtained by Cohen, Hansel, and Sylvester (1954) on the left, and Christensen and Huang (1979) on the right. Bisection frequencies are plotted against t_1 in seconds. The first and second standards are shown as parameters for each figure and straight lines are fitted.)

than the IT model. We summarize the evidence below and add some further considerations.

1. Applied to Howarth and Bulmer's (1956) data, the IT model gave a better fit than the EAL model, though not significantly so.

2. Applied to the present data, IT gave a better fit in six out of eight comparisons, but the differences are very small in some cases, and the overall result is not significant (p = .29).

3. When the models are applied to three-term and four-term analyses of the same data, the IT model shows considerably greater consistency, but again the difference is not significant (p = .087).

4. Both the EAL and IT models show the nonmonotonicities at the $z_c(110) - z_c(001)$ segments of the curves in Figures 3, 4, and 5 that are predicted by the IT model. Thus, the satisfaction of these predictions does not exclude EAL. Nevertheless, they present a problem for this model for the following reason. It is clear why the nonmonotonicities occur if the IT model holds. They follow from the mathematical structure of this model, as expressed by Equation 8. But why should they be produced by exponential decay (if the EAL model holds) in just those cases in which they are predicted by the value of a parameter, D, which is estimated by applying the IT model and which has no valid meaning if EAL holds? (D measures the depth of the dependency, a concept that does not apply to exponential decay.) Predictions of this sort are not required by anything in the structure of the EAL model; if the latter is correct, why should it never produce results that violate the IT predictions? If the IT model is correct, however, but the EAL model is applied, then it will demonstrate these predictions simply because parameters can be found for this model that allow it to simulate the IT model very closely, so that it will show nonmonotonicities if the latter does so.

5. Chi-square for the numbers of response sequences predicted and observed is negligibly less for the IT model (30.76 as against 31.60). (It is noted below that a further development of the IT model allows the result for Subject S5 to be re-analyzed to give $D = 3$ in place of the outlier value of 324, and this subject's $\chi^2(1, N = 1)$ is then 0.626.)

6. The regression analysis in Figure 6, which looks directly at trace decay over time, supports the IT model (p = .035).

7. The auditory bisection data (Figure 7) provide direct evidence that traces decay linearly and not exponentially.

The balance of this evidence favors the IT model. But for most tests, differences are not significant and some observers might prefer the EAL model. In many applications this choice may not be important because if the wrong model is chosen, it is nevertheless a very good approximation to the correct one.

We add some general considerations.

8. It might be argued that the EAL model is the more parsimonious. This is dubious. The two models have the same number of parameters, and there is no reason to consider a linear function less parsimonious than an exponential function. In fact, linear decay as a representation of changes in the continuing relevance of past experience to the current environment is theoretically more satisfying. Some stimuli, once present, are likely to persist for some time with little change. Others are transient and fugitive. Others depart at more or less steady rates. Linear decay in the retention of information about past states of the world seems a reasonable approximation to this range of possibilities. But exponential decay implies that the relevance of a past observation falls off rapidly when it is recent but slowly when it is remote. This does not seem to give an intuitively satisfying picture of these possibilities.

9. It may be argued that studies of memory provide evidence of exponential decay (Wickelgren, 1967, 1968). But there is no necessary conflict between the IT model and such results. The assumption of linear decay of indicator traces does not necessarily imply linear decay of other memory traces. And if their decay is linear, this does not exclude the possibility that experiments involving direct retrieval of memories might produce data that appear exponential. For example, suppose that such a task involves two stages: first, retrieval of a trace, with the probability of retrieval being a function of the strength of that trace, and second, identification of the trace with the probability of doing so correctly being again a function of its strength. Then the final probability of a correct response will be determined

by the product of these two probabilities and will resemble an exponential function.

10. The IT model has an important potentiality that will be exploited in the theoretical developments presented in the remainder of this article. In considering the characteristics of assimilation, contrast, and feedback, and their relations, we will show that it is useful to expand the present model by allowing indicator traces from different sources and representing information of different types to cooperate in determining the criterion while being governed by different decay rate parameters. Without such an assumption it is difficult to account for the conjoint occurrence of both assimilation and contrast with different relations to past trials. If we wish to apply the EAL model, we meet the difficulty that no simple equivalent way of expanding it is available. Criterion shifts induced by different factors could determine the addition of different values of Δ to a common EAL criterion. But once this criterion has been shifted, the same decay parameter applies to the merged residual effects of all past shifts, whatever their disparate sources. They cannot be distinguished, and this presents a limitation to the application of this model to some of the problems dealt with below. The greater power of the IT model in this respect is perhaps the single most important reason for preferring it.

We conclude that in some applications the IT and EAL models cannot be reliably distinguished, and it is open to an experimenter to choose either model, in the knowledge that if it is not the correct one it will at least provide a very good approximation to it. However, the balance of the evidence appears to favor the IT model, and in the developments we present below, this model can be applied in a straightforward way; for these reasons we prefer it.

Some further points deserve comment.

a. The signal-only data show a positive sequential dependency, as is usually found with detection. But the direction of the dependency reverses in the noise-only condition. This relation between the direction of the dependency and the strength of the signal deserves discussion, but this will be deferred until later.

b. The presence and magnitude of criterion variance is a topic of theoretical interest (Wickelgren, 1968) that has not been much discussed, perhaps because of the difficulty of estimating it separately from noise produced by physical and sensory causes. But the magnitude of criterion variance can be estimated relatively simply for data of the present type. The values of the conditional criteria reflect trial-to-trial variation arising from sequential effects, though not random trial-to-trial variation, and thus their variance provides a lower bound to the criterion variance. For the Howarth and Bulmer (1956) data the estimated variance of the eight criteria shown in the inset to Figure 3 is $\sigma_c^2 = .079$ (when the sensory variance, σ_s^2 is defined to be 1). For the present experiment, the mean value of σ_c^2 for 11 subjects for the three-trial analyses is .0816, and for the four-trial analyses it is .1011. Thus the variance of the criterion appears to be of the order of 7%–10% of the total variance in a detection task.

The possibility of estimating σ_c^2 raises the question of how best to define d'. We may write $\sigma_S^2 = \sigma_s^2 + \sigma_c^2$, where σ_S^2 represents the total central variance associated with the signal-only condition, and σ_s^2 represents the component deriving from physical and sensory sources of noise; similarly $\sigma_N^2 = \sigma_n^2 + \sigma_c^2$. d' is customarily defined as $d' = (M_s - M_n)/\sigma_N$. Here criterion noise is lumped with noise from other sources, because it is not usually possible to separate them. But if we can use an analysis of response dependencies to partial out criterion noise arising from sequential effects, we could estimate a value of a sensitivity parameter, d'_t, equivalent to d' but intended to be free of criterion noise, $d'_t = (M_s - M_n)/\sigma_n$. Whether it is worth the trouble to estimate such a measure, free of criterion variance, depends on the use to which it will be put. For many practical purposes there would be no especial benefit in finding d'_t rather than d'. d' is intended to be a measure of sensitivity, and, in the long term, sensitivity is restricted by criterion as well as sensory variation. Thus a measure that takes account of both is justifiable. But for some questions, partialing out criterion variance could be of value. For example, if we are interested in comparing detection of a binocular stimulus with that of a monocular stimulus, we may predict different sensory contributions in the two cases but expect the criterion variance to be the same. Or if we wish to test a predicted relation between the variance of two-alternative forced choices and yes–no judgments, it may be more useful to examine σ_s^2 than σ_S^2.

c. If the IT model is correct it may provide a sidelight on some aspects of memory. It provides information on the dynamics of memory traces that is obtained not by the possibly obscuring mediation of the mechanisms of retrieval that underlie verbal report, but by inference from the effects of retained traces on the ongoing process of perception. The models we have tested constitute analogs of two views of memory retention. On the one hand, in the IT model, memory is based on independent differentiated traces that do not interact and that undergo no changes except decay. A very influential alternative view of memory was put forward by Bartlett (1932) in his description of the schema as

> an active organisation of past reactions, or of past experiences . . . which operate . . . as a unitary mass . . . All incoming impulses of a certain kind . . . build up an active, organised setting . . . There is not the slightest reason . . . to suppose that each set of incoming impulses . . . persists as an isolated member of some passive patchwork (p. 201).

The continually modified criterion of the additive learning models is similar in principle to Bartlett's schema. To the extent that the present results support the IT model, they provide some evidence in favor of the "passive patchwork."

d. The IT model allows us to estimate the depth of sequential effects: For the present data, the median value of D is 3. This could represent three trials or 12 s: Time and trials are confounded in the present study. Collier (1954a) found that a measure of intertrial dependency was a decreasing function of intertrial interval, which suggests that time is the relevant variable, but further work is needed.

e. A further question arises about the IT model. We have assumed that indicator traces are determined by responses. But it could equally well be the sensory inputs that determine them. If the input on trial i is z_i, the detection response will be YES if $z_i - z_c \geq 0$ or NO if the reverse is true. Thus sensory input and response are completely confounded in single intensity experiments. Howarth and Bulmer (1956) tried to determine whether there is an effect of sensory input as such by comparing dependencies contingent on responses with long and short reaction times, on the assumption that the latencies reflect the magnitude of $|z_i - z_c|$, and found almost no effect. But a more direct way to determine whether $|z_i - z_c|$ has an effect of its own is to use a task in which the signal alternates randomly with noise. An experiment using this paradigm is reported below.

Experiment 2

Method

Subjects performed a detection task using the rating procedure with four categories of response. A single signal intensity was used, and signal and noise trials were randomly alternated.

Apparatus and Procedure

The following modifications were made to the apparatus and procedure of Experiment 1. The subject had before him four response buttons that corresponded to the responses "certainly yes," "probably yes," "probably no," "certainly no." (These will be referred to as Responses 1, 2, 3, and 4.) The signal was presented on a proportion $P(s)$ of trials, in random order.

In Condition 2.1 each subject did four sessions. The first was a practice session. In this session a signal intensity was found that gave approximately 50% detection. Each of the three experimental sessions consisted of four blocks of 158 trials, with the signal present on each trial with the probability $P(s) = .5$. The first eight trials in each block were discarded. Thus 1,800 trials were analyzed per subject.

Condition 2.2 was a replication with modifications. There was one practice session and four experimental sessions. On two of the latter, $P(s) = .2$, and on the other two, $P(s) = .8$. The order of the sessions varied for different subjects. Thus 1,200 trials were analyzed for each subject for each signal probability.

Subjects

In Condition 2.1, 6 subjects were used. They were members of the departmental subject panel and ranged in age from 19 to 28. Condition 2.2 used 7 new subjects from 18 to 23 years of age, from the panel.

Results

The probabilities of each of the four responses, contingent on the events of the preceding trial, were found for each subject. The preceding events were classified as (s_{i-1}) (preceding signal) or (n_{i-1}) (preceding noise) and YES_{i-1} (Responses 1 or 2) or NO_{i-1} (Responses 3 or 4). The conditional response probabilities were used to estimate the three corresponding criteria, z_{c1}, z_{c2}, and z_{c3}, ordered as in Figure 1-4, in the usual way (Green & Swets, 1966).

If a value of 0 was obtained for a probability (13 cases for $P(s) = .5$; 5 for $P(s) = .2$; 10 for $P(s) = .8$), z_c was assigned the value 2; if the probability was 1.0 (2 points for $P(s) = .2$), it was assigned −2. (The choice of these values is not likely to have affected the results; the conclusions reached are unchanged if these points are treated as missing data and are estimated.)

The criterion estimates are plotted against their nominal values in Figure 8. Analyses of variance were performed on the data from each condition, the factors being the stimulus on the current trial, the nominal criterion, the stimulus on the last trial, the response on the last trial, and subjects. As we would expect, the current stimulus and the nominal criterion have major effects. More surprisingly, $P(s)$ did not produce a significant effect in Condition 2.2. The main effect of the previous response is highly significant ($p < .005$ for 2.1, $p < .025$ for 2.2). This effect is a positive sequential dependency.

Because the average excess of the sensory input above a criterion is greater when the signal is presented than when noise is presented, any effect of the magnitude of the sensory input as such should manifest as an effect of previous stimulus. But this factor had no detectable main effect in the present experiment. For 2.1, $F(1, 5) = 0.000046$; for 2.2, $F(1, 6) = 0.1152$. Although denial of an effect of sensory input would rest on failure to reject the null hypothesis, these results do at least establish that any direct effect of the sensory input in the preceding trial is small. Thus they support the assumption that the source of the indicator traces is the response made, at least in the present type of experiment.

Two interactions were significant. In 2.1 (but not 2.2) there was an interaction ($p < .05$) between current stimulus and criterion, indicating a difference between signal and noise slopes. In 2.1 ($p < .025$) and in 2.2 ($p < .05$) there were significant fourth-order interactions between current stimulus, criterion, previous stimulus, and previous response. We do not attempt to interpret this, but we note that the involvement of previous stimulus does not necessarily imply an effect of sensory magnitude: It may be a consequence of the lumping together of the previous responses, 1 and 2 and 3 and 4, in the analysis. If the extreme responses, 1 and 4, have different or stronger effects than the intermediate responses, it is relevant that the balance between Responses 3 and 4, and between Responses 1 and 2, will be different for signal and for noise.

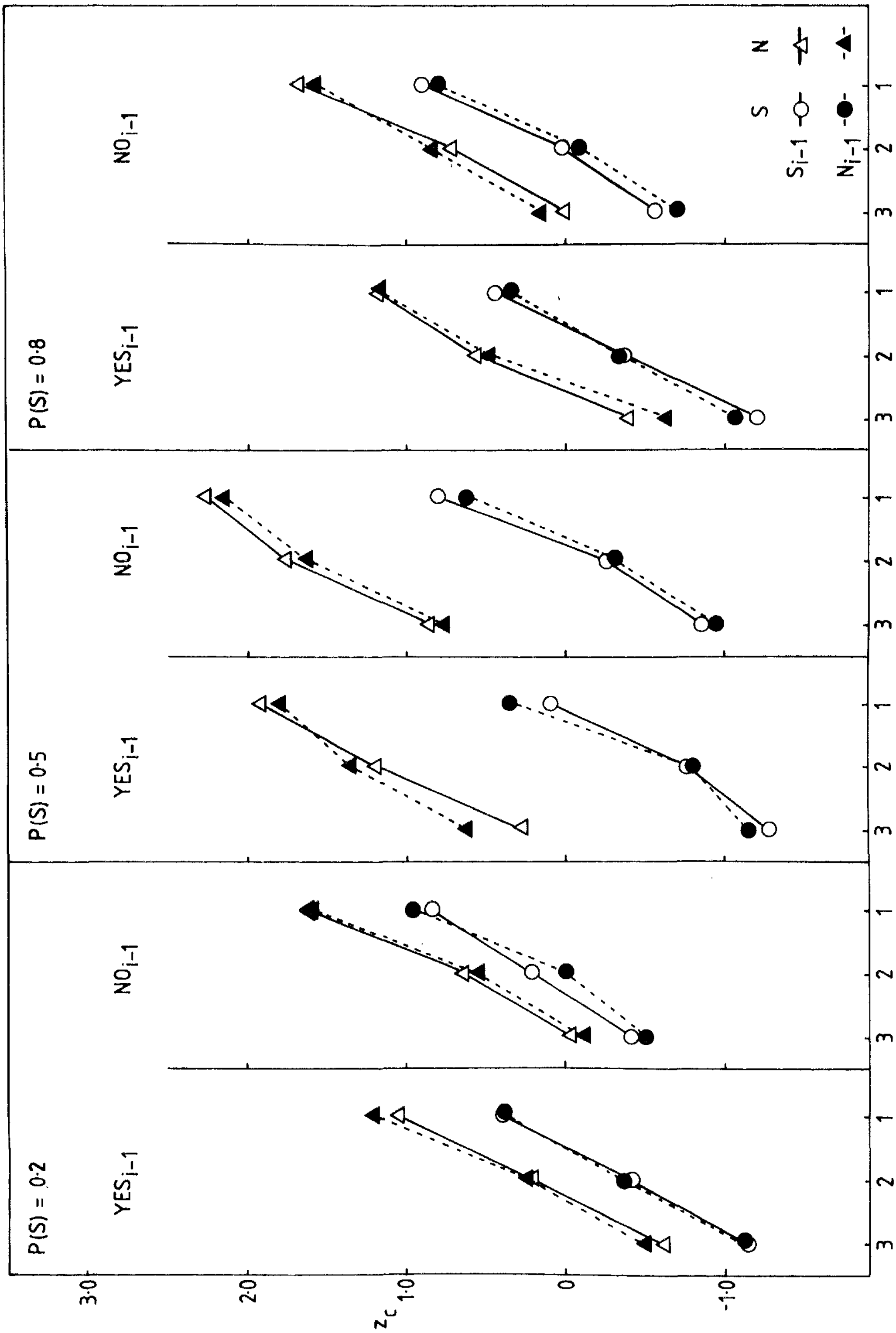
Discussion

Experiment 2 has again produced a positive dependency, with evidence that indicator traces are determined by the responses made, rather than by the magnitudes of the sensory inputs generating them. Further evidence on this point is given by a reexamination of the data of Experiment 1. In a single intensity experiment, the sign of the sensory deviation, the deviation of the input on a given trial from the current criterion, $(z_i - z_c)$, is necessarily confounded with the response made. We also see from Equation 3 that $T_r(i)$ will have the same magnitude whatever the size of the sensory deviation; it is independent of the magnitude of the latter. However, if the indicator trace was directly determined by the sensory deviation itself instead of by the resulting response, we would have, in place of Equation 3, the relation

$$T_s(i) = dJ\Delta_s|z_i - z_c|.$$

That is, in the sensory-determination case the magnitude of the indicator trace is determined by a constant, Δ_s, multiplied by the magnitude of the sensory deviation. In this model the magnitude of the shift is not independent of the sensory deviation. We can test between response determination and sensory-input determination by substituting $T_s(i)$ for $T_r(i)$ in the IT model and seeing which version gives the better fit to the data of Experiment 1.

There is one difficulty. In fitting the response-determined IT model using the simplex procedure, the value of the criterion z_c was followed through seven trials for each possible sequence of responses for each set of parameters examined. At each trial, two responses are possible and have to be considered. But in the sensory-input-determined model, the set of values of $(z_i - z_c)$ that may occur on each trial is infinite. Therefore, in fitting this model the program was made to calculate the expected values of $|z_i - z_c|$ given that $z_i > z_c$, or given that $z_i < z_c$, at each trial, and it used


P(S) = 0·2
P(S) = 0·5
P(S) = 0·8
YES i-1
NO i-1
S
N
S i-1
N i-1
NOMINAL CRITERION
z c
3·0
2·0
1·0
0
-1·0
1
2
3

this approximation in finding $T_s(i)$ for that trial.

For the signal-only condition, the results were unequivocal. For the 6 subjects who showed a positive dependency, SE_P was less for the response-dependent model than for the sensory-input-determined model in five cases, and chi-square was less for the former model for every subject; $\chi^2(24, N = 6) = 23.685$ *ns.* (from Table 2). For the sensory model, χ^2 (24, $N = 6) = 126.349$, $p < .001$, which rejects sensory input determination of the positive dependency.

For the noise-only condition, the results are equivocal. Here 3 subjects showed a negative sequential dependency. For the response-determined model, one subject had a smaller, two had larger values of SE_P than for the sensory model. For the former model, 2 subjects had smaller, one had a larger value of chi-square. For neither model is summed chi-square significant. Thus the choice between determination by response or directly by sensory input remains open for negative sequential dependencies.

A Theory of Criterion Setting

A model based on independent indicator traces derived from past responses appears to give the best account of the present evidence on sequential dependencies. But three important questions remain. First, why is the direction of the effect different in different experiments? In Experiment 1, why did the signal-only condition give positive and the noise-only condition give negative dependencies? Second, why are sequential dependencies ubiquitous in a whole range of tasks ranging from detection to magnitude estimation? Parsimony requires us to seek a single explanation for these observations, if it is possible to find one. Third, why do sequential dependencies occur at all? Do they have a functional role that would explain the observations we have discussed? Dependencies are usually regarded as errors, imperfections in the system, as a nuisance standing in the way of fitting our preferred theoretical accounts. But it would indeed be surprising if a system as complicated as the IT model had come into existence merely to obstruct psychophysicists.

The proposal we now make is that sequential dependencies arise from and reflect the operation of a system that attempts to place criteria at those positions that are optimal at any given moment and to keep them there, and that this same system is at work in all the tasks in which dependencies have been observed. The system consists of a long-term criterion-setting process and two mechanisms that modify criteria by making short-term adjustments. There are also further devices that make use of feedback. We make some basic assumptions about this system and then examine its application to the observations.

The Reference System

Criterion setting involves two stages: first, a general problem of getting in range and second, a requirement for fine adjustment as the environment and past experience change. It is the latter that determines dependencies.

How the long-term value of the criterion is specified is a question that has received considerable theoretical consideration, mainly concerned with its normative prescription (Birdsall, 1955; Green & Swets, 1966; Treisman, 1964). Optimal criteria such as the maximum expected value criterion or the Neyman-Pearson criterion can be defined for detection. In practice, an observer who wishes to employ the maximum expected value criterion needs more information (knowledge of the forms of the signal and noise distributions, $P(s)$, and the payoffs) than is necessary to employ the Neyman-Pearson criterion (the frequency of false alarms). But Boneau and Cole (1967) have shown that, given sufficient feedback, the

Figure 8. Criterion estimates derived from rating responses to signal and noise, expressed as standardized deviates from the means of the signal and noise distributions, respectively. (The results for Experiment 2.1 [$P(s)$ = .5] are shown in the middle panel and for Experiment 2.2 [$P(s)$ = .2 and .8] in the side panels. They are plotted against the nominal criteria, separately for signal and noise trials, for the preceding response either *YES* [Responses 1 or 2] or *NO* [Responses 3 or 4], and for the previous stimulus, either signal or noise. Circles represent responses to the current signal, triangles responses to noise. Empty symbols = previous signal; filled symbols = previous noise.)

minimally informed observer can approximate the first, and it is evident that feedback about false alarms should allow the second to be located (Treisman, 1964). But in many real situations feedback is not available or is not wholly reliable. Even here, however, past experience of similar situations in which feedback has been available may allow reasonable initial values of criteria to be specified. We assume that for any task there is a long-term criterion-setting process that determines initial or reference criteria, such as z_o, on a basis that may be related to some definition of optimality. Thus, in a detection task, this process might aim at a criterion roughly equivalent to the maximum expected value criterion or Neyman-Pearson criterion and it will take account of past experience, that is, the final or mean position of the criterion in any previous similar task, and of global parameters of the current situation, such as the signal probability and payoffs. The setting of the long-term criterion component, z_o, would also be affected by instructions and information from the cognitive system. Once z_o has been specified in this fashion, accumulation of information over a period or change in parameters such as $P(s)$ may cause it to be further modified.

In a task such as absolute judgment, the subject may use his or her initial encounter with the range of sensory inputs to define upper and lower limits and may then place reference criteria at equal intervals, or in accordance with some other scheme, between these limits.

Once the system has specified a value for the reference criterion, this should remain substantially constant over short sequences of trials. But from trial to trial, fine adjustments may be necessary to ensure that the criterion in use is optimally placed. A weighted combination of these short-term adjustments and the reference value will determine the resultant criterion.

There are two sets of considerations that may lead to short-term adjustments. They have opposite implications, and they result in opposite strategies.

The Tracking System

Outside the laboratory, trials are rarely wholly random. A series of detections probably means that there really is something out there at the moment; a series of nondetections may mean the reverse. Given that a generally reliable reference criterion, z_o, has been selected, it may be beneficial to adjust this from time to time in accordance with momentary changes in our assessment of $P(s)$. A predator who has just sighted and failed to capture prey has a chance of making further valid sightings. But if he has seen nothing for some time, a positive response becomes relatively more likely to be a false alarm. In the absence of valid feedback, the best evidence for a raised or lowered $P(s)$ is the occurrence or absence of recent positive judgments. The more recent the evidence, the more highly it should be weighted: The external world changes. Given recent detections the criterion should be lowered; following recent rejections it should be raised.

This strategy of *tracking* the presence or absence of a possible stimulus is just what is achieved when the IT model operates with $d = -1$ to give positive sequential effects dependent on past responses. The LAL model would fail for this purpose: Faced with the continuous presence or continuous absence of a signal, it would send the criterion infinitely low or infinitely high. But the IT model sets sharp upper and lower limits to the range of criterion variation, determined by the values of Δ and δ.

The Stabilizing System

If the long-term position of the criterion is uncertain or unreliable, the first priority must be not to maximize the momentary match to the state of Nature but to use incoming information to position the criterion in an appropriate position and keep it stable. From this point of view, a sustained series of *YES* responses provides information that the criterion is too low, repeated *NOES* indicate that it is too high. The stabilizing strategy requires that a *YES* should be followed by a rise in criterion, a *NO* by a fall in criterion. In the simplest case, this tends to move the criterion to a position such that the probability of *YES* equals the probability of *NO*. Consequently, the stabilizing strategy tends to maximize the information transmitted by the subject's responses. This strategy would be realized by the IT model functioning with $d = 1$ to give a negative dependency.

A simple way to envisage these strategies operating would be to assume that the system may switch from one strategy to the other by setting $d = -1$ or $d = 1$. For example, we may suppose that normally the detection criterion rarely wanders below a lower limit, z_d, and if it does it becomes unstable (cf. Figure 1-1). Then if the criterion does drift below z_d, d would change from -1 to 1, switching into the stabilization mode. This assumption might apply in some simple cases, but it is not adequate in all cases, and it is not the assumption that is made here. Instead we note that it is a strength of the IT model that it can accommodate two strategies that are operating in parallel at the same time with different determinants and that are controlled by different parameters, so as to serve their two purposes simultaneously. We assume that one strategy may be in operation, or the other, or both simultaneously, depending on the requirements of a given task.

Determinants of tracking. Because the subject's recent past decisions constitute the best available indication he has of the momentary state of Nature, these are used to determine the indicator traces for the tracking strategy. This strategy will have parameters Δ_r and δ_r. Because the dependency is normally positive, we drop the parameter d and replace Equation 3 by

$$T_r(i) = -J_i\Delta_r. \qquad (12)$$

Because the most reliable information is the most recent, Δ_r and δ_r are usually relatively large and D is fairly small. This will also have the effect that, although criterion fluctuations produced by tracking may be relatively marked, they will also be fairly evanescent and should not have much effect on the long-term mean position of the detection criterion.

Determinants of stabilization. Not only can a series of *YES* responses indicate that the criterion is too low, but the magnitudes of the differences between the sensory inputs and the criterion $|z_i - z_c(i)|$ provide information that strengthens or weakens this indication. This makes input-criterion differences the best determinants of successive moves in a stabilizing strategy. Furthermore, the larger the sample of sensory indicator traces that determines the current criterion value, the more reliable their mean effect and the less susceptible to momentary changes in $P(s)$. This favors relatively small values of the parameters Δ_s and δ_s, giving relatively slow decay of the traces. Thus we assume that the stabilizing strategy is stimulus dependent and is realized by using indicator traces that are functions of past sensory input-criterion differences. Equation 3 becomes

$$T_s(i) = \begin{cases} \Delta_s(z_i - z_c(i)), & |z_i - z_c(i)| \leq z_U \\ 0, & |z_i - z_c(i)| > z_U \end{cases}. \qquad (13)$$

Here Δ_s is a positive constant of proportionality; z_U is a large positive number. Its function is to exclude the effects of stimuli wholly out of range (where they are not already excluded by forming part of inappropriate perceptual complexes). Thus if clicks are being detected, an unusually loud click may be taken into account, but an intruding door slam would be prohibited from producing a wild rise in criterion.

The indicator trace decays toward 0 at the rate δ_s, so that its value, k trials after its creation, is

$$T_s(i, k) = \begin{cases} \max(T_s(i) - k\delta_s, 0), & T_s(i) \geq 0 \\ \min(T_s(i) + k\delta_s, 0), & T_s(i) < 0 \end{cases}. \qquad (14)$$

It follows that stabilization adjustments are not produced by deviations $|z_i - z_c(i)|$ that are less than an intrinsic lower limit $z_L = \delta_s/\Delta_s$. Deviations less than or equal to z_L produce sensory indicator traces that decay to 0 within one intertrial interval and thus cannot contribute to stabilization. Thus if $\delta_s = \Delta_s$, only deviations more than one standard deviation from the criterion can have an effect.

The two strategies may operate in parallel, their relative effects depending on their parameters. Usually, but not always, we expect response indicator traces to decay rapidly, but larger samples of more slowly decaying sensory indicator traces to determine stabilization. Thus recent trials may produce predominantly positive and more remote trials, predominantly negative dependencies. One or the other strategy may be suppressed entirely by setting $\Delta_r < \delta_r$ or $\Delta_s = 0$.

Because the initial absolute value of the sensory indicator trace is not a constant, unlike $T_r(i)$, the depth of the sequential dependency produced by stabilization is not constant either. A large, remote input-criterion difference may still affect a current criterion when a smaller more recent difference has ceased to be effective.

The discussion above applies to detection. In more complex tasks such as absolute judgment there are several criteria that may be affected by the events on a past trial. Thus if Response 1 is made (cf. Figure 1-4) there are several possibilities. The corresponding tracking indicator traces may (a) shift all the criteria the same distance to the left on the next trial, (b) only decrement the criterion (z_{c1}) that immediately determined the response, or (c) shift z_{c1} and other criteria also, but to a lesser extent. What happens in any case will have to be determined by experiment.

Feedback. The present model provides a basis for a theory of feedback. Feedback may produce effects that vary with subjects and situations. Four possible ways of employing the information provided by feedback are suggested.

1. *Selective stabilization.* If reliable feedback is available, it could be used to select those sensory inputs that should contribute to stabilization. That is, trials on which the feedback FB_i confirms the response R_i made on that trial would be allowed to produce effective sensory indicator traces, because the feedback excludes the possibility that the input-criterion difference arose from a miss or a false alarm. We may represent this effect of feedback by adding a multiplier to Equation 13. We define

$$k_f = 1, \quad FB_i = R_i$$
$$k_f = 0, \quad FB_i \neq R_i \qquad (15)$$

and $k_f = 1$ in the absence of feedback.

Then Equation 13 becomes

$$T_s(i) = \begin{cases} k_f \Delta_s[z_i - z_c(i)], & |z_i - z_c(i)| \leq z_U \\ 0, & |z_i - z_c(i)| > z_U. \end{cases} \qquad (16)$$

2. *Stabilization suppression.* It is possible that feedback, being cognitively mediated, cannot be applied to individual sensory deviations, but only by the reference system. Then the receipt of reliable feedback over a period might be used to estimate global parameters that in turn determine the long-term criterion component z_o—and the weight given to the stabilization procedure would then be reduced, or that mechanism might be suppressed entirely.

3. *Selective tracking.* It is possible that feedback might determine the responses to be given weight in tracking. Then the response R_i would contribute to a positive dependency only if $R_i = FB_i$. This would correspond to multiplying $T_r(i)$ by k_f, as defined by Equation 15. This might parallel simultaneous selective stabilization or stabilization suppression.

4. *Feedback-based tracking.* If the responses made are unreliable, so that they rarely agree with the feedback, it may be advantageous to substitute the feedback for the responses as a basis for tracking. Then FB_i would determine the tracking indicator trace on every trial i, whether or not $FB_i = R_i$.

The resultant criterion. The resultant value on trial i of the criterion produced by the reference, tracking, and stabilization mechanisms may be written as

$$z_c(i) = z_o + \sum_{h=i-D}^{i-1} T_r(h, i-h) + \sum_{h=1}^{i-1} T_s(h, i-h). \qquad (17)$$

The relative weight of each component may be varied by modifying the parameters Δ_r, δ_r, Δ_s, and δ_s.

Because long-term factors (such as prior knowledge of $P(s)$ or payoff values) are embodied in the choice of z_o, we may expect that tasks that are likely to give this criterion component a high weight, such as well-practised detection against a continuous background of familiar noise, will most clearly demonstrate the expected features of optimal criteria as defined by SDT; for example, a fall in criterion with increase in $P(s)$. But if the conditions are unusual or difficult and z_o is uncertain, short-term strategies will dominate, diluting the weight of global factors and perhaps resulting in different relations.

Criterion maintenance when not performing the task. For most of our lives, most of our decision criteria are not in use. In some cases (e.g., deciding which house to buy) the occasions for their application are clearly defined and easily anticipated. But in other cases (e.g., was that a shooting star?), though rarely used, they must be constantly available. How then can criterion setting and maintenance be ensured during periods when the appropriate stimulation for such decision criteria does not occur? For example, consider a detection task and suppose that the system is set to maintain a Neyman-Pearson criterion giving a fixed false alarm rate, ϵ. When stimulation and feedback are available they provide a basis for estimating the false alarm rate and defining a corresponding value of z_o. But how can this criterion be maintained during periods when no stimulation is presented and no overt responses are made?

The simplest possibility is that the previous value of z_o is stored in long-term memory. There are difficulties with this: First, the stored memory of the criterion might decay, causing apparent drift in its position, and there would be no check on this. Second, the background noise level might alter in the interim, changing the false alarm probability given by the old criterion value and requiring a revised value.

A possible strategy is suggested by the observation that during a period without stimulation, false alarms may occur. This tells us that the sensory input (determined by noise) is sampled from time to time. Such regular samples could be allowed to activate the stabilization mechanism, even though no overt responses are made. We refer to this as *latent stabilization.* We assume that regular samples are taken and that each is compared with the current criterion. As a result, an indicator trace is set up that produces an upward shift if the sensory input-criterion difference is positive, or a downward shift if it is negative. Because samples usually fall below the initial detection criterion, however, it is evident that if upward and downward shifts are given the same weight, the criterion must drift down toward the mean of the noise distribution and so abandon the Neyman-Pearson criterion. This difficulty is met by having an asymmetry between upward and downward shifts realized by giving Δ_s different values according to the direction of the shift. Thus we substitute Δ_{su} for Δ_s in Equation 16 when an upward shift is indicated and Δ_{sd} when a downward shift is required, and give Δ_{su} and Δ_{sd} values appropriate to maintaining the desired criterion. If the stabilization strategy is applied under stable conditions, it is reasonable to suppose that the value of the long-term sum of sensory indicator traces varies about an equilibrium point, z_e. We can identify this value by making the supposition that z_e is the point at which the expected absolute value of an upward shift is equal to the expected absolute value of a downward shift. If so, because the mean of the normal distribution truncated at a is given by $f(a)/(1 - (F(a))$, and disregarding the effects of decay and the contribution of z_0, we may write

$$\epsilon\Delta_{su}[f(z_e)/\epsilon - z_e] = (1 - \epsilon)\Delta_{sd}[f(z_e)/(1 - \epsilon) + z_e], \quad (18)$$

where $\epsilon = 1 - F(z_e)$ and gives the Neyman-Pearson false positive rate for suitable Δ_{su} and Δ_{sd}. (We do not have a proof for this conjecture, but it has been tested and is supported by computer simulations.) It follows that

$$\Delta_{sd}/\Delta_{su} = \frac{f(z_e) - \epsilon z_e}{f(z_e) + (1 - \epsilon)z_e}, \quad (19)$$

from which we see that to maintain z_e at the value that will give $\epsilon = 0.05$, for example—despite any drift in the sensory noise level—will require $\Delta_{sd}/\Delta_{su} = 0.0126/1 \doteq 1/80$. As Δ_{sd} increases, so the equilibrium point moves down. Thus if Δ_{sd} and Δ_{su} are set at appropriate initial values, latent stabilization (a negative dependency determined by sensory noise inputs) can ensure appropriate adjustment of the criterion as the sensory state of the subject changes, even in periods of nonstimulation.

Because $T_s(i, k)$ is subject to decay and so is summed over a limited number of terms, its average effect at any time is a partial shift in the direction of z_e that makes a weighted contribution to the resultant criterion. If z_o is unreliable or uncertain, increased weight can be given to latent stabilization by increasing the magnitudes of Δ_{sd} and Δ_{su} without altering their ratio, or by decreasing δ_s, thus increasing the relative weight of z_e as compared with z_o.

Latent stabilization could maintain a Neyman-Pearson criterion in the absence of stimulation and formal feedback. Alternatively, it might be used to approximate a maximum expected value criterion defined for given signal and noise distributions and signal probability, if Δ_{sd} and Δ_{su} are determined as functions of the probabilities of signal and noise, d', and the payoff values.

Some Applications of Criterion-Setting Theory

The theory has now been outlined, mainly as it applies to detection. It will be developed for other applications elsewhere (Treisman, 1983a, 1983b, in press; Treisman & Faulkner 1983a, 1983b, in press.) At this point we stop and ask how well it can account for the better established features of sequential effects. We consider detection and identification tasks first, followed by absolute judgment and magnitude estimation.

Detection and Identification

Further consideration of Experiment 1. The theory suggests that tasks that employ well-established criteria should show positive sequential dependencies, but more difficult or unfamiliar tasks should be characterized by increased salience of stabilization, resulting in negative dependencies. It is of interest to reconsider Experiment 1 in relation to these possibilities. The signal-only condition is represented by Figure 1-2. Here, the decision criterion in use is likely to be the subject's prior and long-established detection criterion. In this experiment, an initial practice session was used to find the signal intensity that gave $P(Y) =$ .5. Such a signal will have a sensory distribution whose mean M_s is at or near z_c, the subject's preestablished criterion. So this condition required the subject to employ an habitual criterion, leaving him free to give priority to tracking. This argument predicts—and it will be remembered that we found—positive sequential dependencies in this condition.

The noise-only condition is represented by Figure 1-1. The absence of a signal, and the instruction to aim at $P(Y) = .5$, should force the subject to shift his criterion far to the left of its familiar position. It will move to a new lower value z'_c that coincides with the mean of the noise distribution. We have supposed that normally the criterion rarely wanders below a lower limit, z_d, and that if it does so it becomes unstable. So we expect stabilization to be the prime concern in this condition, giving negative sequential dependencies, and this was what we found.

These results support the assumption that the salience of tracking or stabilization will depend on the conditions of the task. The assumption that stabilization may be suppressed when the criterion is well established and input-criterion differences are small also finds support in the lack of a stimulus effect in Experiment 2, a detection task with a constant noise background and a single signal alternating with noise.

Is it possible that both strategies were operating in Experiment 1 but that the less salient strategy was overlooked? To examine this question, the results were re-analyzed. For each subject the data were fitted to Equation 17, the parameter-fitting program being used to find five parameters, z_o, Δ_r, δ_r, Δ_s, and δ_s, for seven trial sequences. There was no clear evidence of an improved fit for any subject. However, for Subject S5 the assumption that the two mechanisms were in simultaneous operation gave more reasonable parameters than tracking alone had done: D was reduced from 324 to 4.

Another possibility that was considered is the following. Because subjects were receiving an intensity at the 50% detection level alone, in one condition, and no signal at all in the other, it is possible that a latent stabilization strategy might continue to operate during the experiment. To examine this, a model was fitted that combined tracking and stabilization with separate parameters Δ_{su} and Δ_{sd}. Again, this produced no clear improvement over the single strategy results for any subject except S5. However, for this subject, the parameters $z_o = 0.092$, $\Delta_r = 0.330$, $\delta_r = 0.090$, $\Delta_{su} = 1.184$, $\Delta_{sd} = 0$, and $\delta_s = 0.431$ give $MSD = 0.00747$, $SE_p = 0.0864$, and $\chi^2(1, N = 1) = 0.626$. This is considerably better than the result for a single value of Δ_s, $\chi^2(2, N = 1) = 8.830$, and D now has the reasonable value 3. The improved fit is shown as a dashed and dotted line in Figure 4.

Task conditions, especially novelty or familiarity of the task. The predicted relation between the conditions of the task and the direction of the dependency is also supported by the data of Sandusky and Ahumada (1971). In place of the positive dependency usually seen with detection against a constant noise background, these authors found that if the signal and the noise are gated on and off together, there is a negative sequential effect. They relate this to the absence of the framework of reference constituted by a constant noise background. In terms of our assumptions, the absence of this framework has its effect by reducing the reliability of placement of z_o and thus requiring more active stabilization.

An unfamiliar identification task, in which only a single stimulus is presented on each trial, should induce uncertainty about the criterion and therefore should require stabilization. Parducci and Sandusky (1965) provide data on such a task. A light was shown briefly to the subject in one of two positions, left or right, in the dark. A reference light was present and no feedback was given. They found that the probability of a correct response was increased for a stimulus alternation as compared with a repetition. This indicates that the criterion had moved toward the previous stimulus, illustrating a negative stimulus dependency. Other experiments (e.g., Tanner et al., 1967) that support this result are discussed below.

Probability of the signal. Criterion-setting theory predicts that a stable criterion largely determined by z_o will show the changes predicted by SDT in response to variation in $P(s)$, but that a criterion largely determined by the short-term strategies will not. There is no lack of detection studies in which well-practised subjects exposed to familiar noise backgrounds vary their criteria in response to changes in $P(s)$ in the directions predicted by SDT (Green & Swets, 1966). We may contrast this situation with the identification task, in which the subject may be exposed to randomly alternating intermittent presentations of, say, tones at 68.5 db or 70 db and must categorize them. This is a much less familiar task than detection in daily life, and the subject is likely to have special difficulty in maintaining a fixed criterion for it. Here the stabilization mechanism should be important in determining the position of the criterion, that is, z_e will be weighted large. If so, the effects of change in $P(s)$ on z_o should be less evident in the behavior of the resultant criterion.

What effect should $P(s)$ have on z_e? To answer this question, we extend Equation 18 to apply to the two distributions used in a typical identification task. The stimulus s_1 (see Figure 1-3) is assumed to occur with probability $P(s_1)$, and s_2 with probability $1 - P(s_1)$. If we take the equilibrium value of the criterion, z_e, as the point on a standardized scale at which the expected magnitude of shift in either direction is the same, let the mean of the s_1 distribution be d' and the mean of the s_2 distribution be zero, with a standard deviation of one in each case, and define

$$\epsilon_2 = \int_{z_e}^{\infty} = N(0, 1)dz$$

for s_2, and

$$\epsilon_1 = \int_{z_e-d'}^{\infty} N(0, 1)dz$$

for s_1, we have

$$\begin{aligned}\Delta_{su}\{&(1 - P(s_1))\epsilon_2[f(z_e)/\epsilon_2 - z_e] \\ &+ P(s_1)\epsilon_1[f(z_e - d')/\epsilon_1 - (z_e - d')]\} \\ = \Delta_{sd}\{&(1 - P(s_1))(1 - \epsilon_2)[f(z_e)/(1 - \epsilon_2) \\ &+ z_e] + P(s_1)(1 - \epsilon_1)[f(z_e - d')/ \\ &(1 - \epsilon_1) + z_e - d']\}. \quad (20)\end{aligned}$$

(This is supported by computer simulation.) For $\Delta_{sd} = \Delta_{su} = \Delta_s$, this simplifies to

$$z_e = P(s_1)d'. \quad (21)$$

The implications of this relationship are illustrated by the variation in the probability of the response S_1, $P(S_1)$, as $P(s_1)$ varies, which is shown for two values of d' in Table 4.

Most experiments of this sort are run with d' of the order of one. Evidently, as $P(s_1)$ changes, $P(S_1)$ will remain almost constant at about 50%, to the extent that z_e determines the criterion in use.

Relevant evidence is provided by Parducci and Sandusky's (1965) position identification experiment, in which the accuracy data showed a negative dependency, indicating active stabilization. For three probabilities of the left stimulus, .2, .5, and .8, the probability of the response LEFT was .48, .52, and .53, respectively. This indicates that there is a shift of the criterion toward s_1 as $P(s_1)$ increases, as can be seen in Figure 2 of Sandusky (1971). Tanner et al. (1967) performed an identification experiment using two 1000-Hz 100-msec tones, the louder at 70 db, the softer reduced to a level at which about 70% of responses were correct. There was no feedback, and the effect of increasing *P*(*loud*) from .1

Table 4
Identification Task

	$P(s_1)$	
$P(s_1)$	$d' = 1$	$d' = 2$
0.1	0.496	0.475
0.2	0.494	0.465
0.5	0.500	0.500
0.8	0.506	0.535
0.9	0.504	0.525

Note. Values of the overall probability of the response S_1, $P(S_1)$, for different values of $P(s_1)$, the prior probability of presenting the stimulus s_1, predicted by Equation 21.

to .9 was to give a small increase in P(LOUD), from .52 to .60. Tanner et al. (1970) performed a similar experiment. In an initial no-feedback condition, *P*(LOUD) increased from .54 to .56 as *P*(*loud*) varied from .2 to .8. But after further experience (12 sessions with feedback), a repetition of this condition resulted in P(LOUD) increasing from .43 to .69, indicating an increased weighting for z_o as a result of this lengthy training. (With feedback, *P*(LOUD) increased from .18 to .81 as *P*(*loud*) went from .2 to .8, which shows that feedback increased the weight given to z_o. This supports the hypothesis that feedback causes *stabilization suppression* rather than *selective stabilization.*) Parducci and Sandusky (1970) interpolated an intermediate tone between successive presentations of their experimental loud or soft tones. In one condition the interpolated tone was described as a *standard,* in the other, as a *warning.* The probability of the loud stimulus was .2, .5, or .8. In the standard condition the corresponding values of *P*(LOUD) were .50, .46, and .50. In the warning condition they were .56, .54, and .54. These various results accord with the prediction derived from Equation 21.

Further support is provided by Sandusky and Ahumada (1971). In a detection task in which both signal and noise were gated on and off together, the criterion shifted to the right when *P*(*s*) increased from .25 to .75.

Criterion-setting theory assumes that both stabilizing and tracking mechanisms may operate at the same time. Several lines of evidence support this claim: Some of the data are illustrated in Figure 9.

Collier and Verplanck (1958) used the single intensity detection procedure, with a stimulus of fixed luminance. Because the stimulus intensity is the same on every trial, the criterion is likely to be near M_s (Figure 1-2). Some of their data are replotted in the first panel of Figure 9. The points labeled *single intensity* give *P*(*Y*) following all sequences of two preceding responses, such as two YESES (11), and so forth. These points show a positive dependency on both the preceding response and the penultimate response. In a second condition, occasional stimuli were interpolated that were either 0.5 log units greater or 0.5 log units less than the fixed luminance. Trials on which an immediately preceding interpolated strong stimulus elicited YES or a weak stimulus NO, grouped also according to the penultimate response (to the standard luminance), are shown under the heading *Forced Response.* Here again *P*(*Y*) is higher for an immediately preceding (forced) YES, and higher for an (unforced) YES on the trial before that. But this positive dependency on the preceding response is now reduced as compared with natural two-trial sequences: 11 exceeds 10, and 01 exceeds 00 to a smaller extent when the final 1 corresponds to an enhanced stimulus, the final 0 to a dim stimulus. This demonstrates that an operation that has increased the expected values of $|z_i - z_c(i)|$ has amplified an underlying negative effect of the stimulus that subtracts from the positive effect of the response.

Tanner et al.'s (1967) results for identification of the loudness of a tone, without feedback, are illustrated in Figure 9 for *P*(*loud*) = 0.5. They obtained similar results for four other values of *P*(*loud*). Similar results are reported by Tanner et al. (1970), and the same picture emerges from Sandusky's (1971) analysis of the data of Parducci and Sandusky (1965). It is evident that *P*(LOUD) is greater on the current trial if the preceding response was LOUD than if it was SOFT, a positive dependency on the last response, but *P*(LOUD) is less if the preceding stimulus was *loud* than if it was *soft*, a negative dependency on the past stimulus.

In an identification experiment by Haller (1969), three closely spaced tones were identified in two categories, LOUD and SOFT, by 10 subjects. s_1 was 70 db, s_3 was at an average of 68.5 db, and s_2 was intermediate. There were four conditions, which differed in the distributions of probability over the three

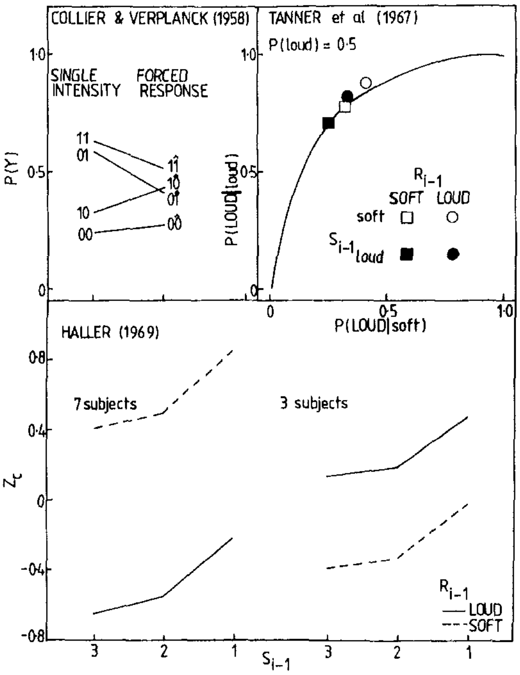


Figure 9. Some experimental results on sequential dependencies. (a) Collier and Verplanck (1958): The single intensity procedure. The data replotted here compare $P(Y)$ following a naturally occurring pair of responses with $P(Y)$ following a pair in which the first response (1 for YES or 0 for NO) was made to the standard intensity; the second response was forced by a reduced or increased intensity stimulus (forced response). (b) Tanner, Haller, & Atkinson (1967): The probability of a LOUD response to a *loud* stimulus is plotted against *P*(LOUD|*soft*) for the four possible conjunctions on the previous trial. (c) Haller (1969): The position of the identification criterion is plotted against the stimulus on the previous trial (where 1 = *loud* and 3 = *soft*). The continuous lines represent a previous LOUD, the dashed lines a previous SOFT response.

stimuli. All three stimuli were employed in three of the conditions. The data for these three conditions have been re-analyzed as follows: The standard deviate corresponding to each probability of LOUD was used as an estimate of z_c on a scale with its origin at the mean of the corresponding stimulus, and mean values of z_c for each combination of preceding stimulus and response was found. The subjects fell into two groups: 7 subjects' mean scores are plotted on the left; 3 subjects' mean scores, on the right.

In both cases, the stronger the previous stimulus, the higher the criterion—a negative effect. For 7 subjects, the previous response had the reverse effect, giving a positive dependency. For these subjects, an analysis of variance gave highly significant (allowing for two missing data points) main effects (probability condition, $p < .025$; current signal, $p < .001$; preceding response, $p < .025$; preceding signal, $p < .001$), but no interaction, at any level, was significant. This accords with our assumption that the effects of previous stimuli and of previous responses combine by linear addition. For the remaining 3 subjects, the previous stimulus and response both produced negative dependencies, although these were not significant.

For all the subjects, z_c increased when the relative probability of s_i increased and decreased when the probability of s_3 increased, which accords with Equation 21. (Although this relation is produced by the stabilization mechanism, which is also responsible for the effect of preceding signal in the analysis of variance, it arises as the resultant effect of a series of previous trials, and thus probability condition manifests as a separate additive factor.)

John (1973) and McNicol (1980) also provide evidence that the current response is assimilated to the preceding response and contrasted with the preceding signal in identification tasks.

The effects of feedback. Four possible strategies for utilizing feedback were derived from the theory. The first two proposed that stabilization may either be selective, or it may be suppressed in favor of a higher weighting for the long-term component of the criterion. Evidence from a number of sources may allow us to choose between these two alternatives, for detection and identification tasks.

If feedback produces selective stabilization (Equations 15 and 16), it should cause a negative dependency similar to, but more marked than, that given by stabilization when there is no feedback. Stabilization suppression would, of course, eliminate the dependency. Parducci and Sandusky (1965) report a negative dependency on the previous stimulus when there is no feedback, but not when feedback is given. This indicates suppression. Support for this is provided by Haller's (1969) comparison of data obtained by Kinchla (1966; as reported by Haller, 1969) and by Tanner et al. (1967); both studies examined identification using the same stimuli and apparatus. As we have already

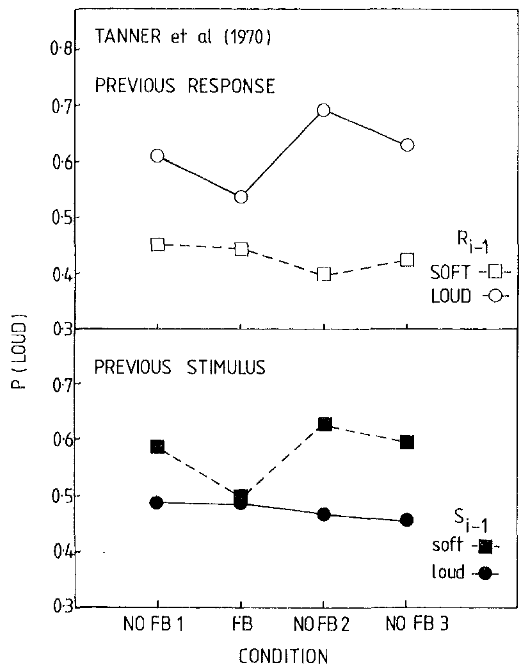


Figure 10. Data from an identification experiment by Tanner, Rauk, & Atkinson (1970) in which subjects judged a tone as *LOUD* or *SOFT*. (In the upper panel the probability of the *LOUD* response is shown as a function of the response on the preceding trial; in the lower panel it is shown as a function of the preceding stimulus. There were three conditions with no feedback and one with feedback.)

seen, *P*(*LOUD*) changed very little as *P*(*loud*) increased from 0.1 to 0.9 in the latter study, in which no feedback was given and a negative stimulus dependency was shown. But *P*(*LOUD*) closely matched *P*(*loud*) in Kinchla's (1966) experiment, in which feedback was given. So it seems that (at least for identification) feedback may suppress stabilization and enhance reliance on z_o, which is determined by *P*(*s*) in accordance with SDT, and so will tend to give probability matching. Dorfman, Saslow, & Simpson (1975) also report probability matching when feedback is given.

Tanner, et al. (1970) ran four identification conditions in succession with the same subjects. Feedback was given in the second condition but not in the other three. From their Table 1 we may calculate the average values of *P*(*LOUD*) contingent on each possible stimulus, or response, on the previous trial. These values are plotted in Figure 10. In every condition, we see a positive dependency on the previous response and, except in the feedback condition, a negative dependency on the previous stimulus. In the feedback condition, the probabilities contingent on the previous stimulus are almost identical, indicating that stabilization was suppressed. It is in keeping with this interpretation that *P*(*LOUD*) matched *P*(*loud*) closely in the feedback condition but, when feedback was absent, not at all in naive subjects and much less so in experienced subjects.

If feedback in a single-interval detection task suppresses a negative dependency on the previous stimulus, this should have the effect of reducing criterion variance and so should increase a measure of sensitivity. Gundy (1961) reported small effects of feedback on d_s, not all in this direction. However, because he used a constant noise background, any negative dependency may have been small to start with. Carterette, Friedman, & Wyman (1966) found a lower *P*(*C*) in a 2IFC procedure with continuous noise background when feedback was sometimes incorrect than when it was always correct or not given. It is possible that the incorrect feedback induced an increase in criterion variance. Friedman, Carterette, Nakatani, & Ahumada (1968) examined single-interval auditory detection against a continuous background, with *P*(*s*) varying from .25 to .75. In both their low- and high-signal-intensity conditions, *P*(*hit*) and *P*(*false alarm*) both varied over larger ranges as a function of *P*(*s*), in the direction of probability matching, when correct feedback was given than in a no-feedback or in a random-feedback condition (which suggests that the random feedback may have been disregarded).

Friedman et al. (1968) also found that if feedback was random on signal trials but correct on noise trials, the criterion was very high, and in the reverse condition it was very low. This is an intriguing finding that provides further evidence against the possibility that feedback produces selective stabilization in detection. Consider random feedback on signal trials, with correct feedback on noise trials. Assuming that selective stabilization occurs when feedback and input-criterion differences agree, the main effect would be to nullify about half of the differences arising on the signal trials (on which the greater proportion of the sensory deviations would be positive), and this would cause an imbalance in favor of the negative dependency produced on the noise trials. Consequently, the criterion would move down. But in fact it did not. An explanation for this result is provided by the third feedback strategy, selective tracking (which may have been combined with stabilization suppression). If feedback in disagreement with a response invalidates it for tracking, then in the signal-random, noise-correct case, the major contribution to tracking shifts would have been made by responses to noise. As most of these would be NO, which causes the criterion to rise, this would explain the high criterion value found. Selective tracking can also explain the relative reduction in the positive dependency on the previous response that is seen in the feedback condition of Tanner et al.'s (1970) identification experiment, as illustrated in the upper panel of Figure 10.

Overall, the evidence from these tasks supports the hypothesis that reliable feedback can substitute for short-term stabilization and can provide a basis for maintaining confidence in the long-term criterion, z_0. There is no evidence that it can be used to give selective stabilization. But there is support for the occurrence of selective tracking.

Absolute Judgment and Magnitude Estimation

Absolute judgment results obtained by Ward and Lockhead (1971) are illustrated in the upper four panels of Figure 11. Their subjects judged 10 tones separated by 1 db, and the responses were the numbers 1 to 10. For each stimulus, the average departure from its nominal value was found. In the figure, the average error for all stimuli on trial i is plotted for the case when the stimulus on trial $i - 1$ was 9 or 10 (lag 1), when the stimulus on trial $i - 2$ was 9 or 10, whatever the stimulus on $i - 1$ (lag 2), and so on. A similar curve is shown for trials preceded by Stimuli 1 or 2 at the various lags. The curves for other preceding stimuli, which are not shown in Figure 11, were largely intermediate between those that are shown. On the right, the same data are plotted in relation to preceding responses. The upper panels show a condition with, and the middle panels a condition without, feedback. Similar results have been obtained for other sensory modalities, and using magnitude estimation with different response ranges (King & Lockhead, 1981). Lockhead and King (1983) used a successive ratio-judgment task in which subjects were asked to respond to the current stimulus I_n by giving the ratio of its loudness to the loudness of the immediately preceding stimulus (I_{n-1}). They derived the equivalent magnitude estimation data from these ratios: These show sequential dependencies very similar to the absolute judgment results in Figure 11. Other authors have reported similar findings (e.g., Jesteadt et al., 1977; John, 1973).

These data show features that demand an explanation. Considering the feedback condition first, we see a positive dependency on the immediately preceding trial and a negative dependency on earlier trials. Because stimuli and responses are correlated, the overall picture is similar whether plotted for past stimuli or for past responses. These findings can be explained by the assumptions that (a) both of the short-term mechanisms are active, (b) the stabilizing strategy establishes a negative dependency extending over several trials, and (c) the tracking mechanism adds a positive dependency that is more marked but is limited to more recent responses. These assumptions are supported by the observations that in these data the negative dependency appears greater when the curves are classified by preceding stimuli than by preceding responses, and the positive dependency is more evident when the data are analyzed in relation to past responses than in relation to past stimuli.

When feedback is omitted, we see a marked increase in the positive dependency that is especially evident when the data are classified by past responses (note the doubled scale units for the right middle panel of Figure 11). This may be explained by the third strategy for handling feedback information, selective tracking. When this strategy is applied, only if $R_i = FB_i$ does the response affect the criterion: Complete or close agreement with feedback is required if the response is to contribute to tracking. If so, when feedback is given, many responses will be ineffective for tracking, allowing the effects of stabilization to bulk relatively larger. But in the absence of feedback, all responses will be effective, producing a larger positive dependency. This "release of tracking" is what we see in the no-feedback panels. It is, of course, more evident when plotted against past responses than against past stimuli.

The evidence that stabilization occurs both with and without feedback shows that in absolute judgment (unlike detection or identification), stabilization is not suppressed by feedback. Because the absolute-judgment task requires many more criteria than detection or two-category identification (nine in this case), it must be correspondingly more difficult to use feedback alone to establish and maintain a set of stable long-term reference criteria, and so the stabilization strategy remains essential. In detection, feedback may provide an estimate of $P(s)$ that can be used to determine the best value for z_o. It is not evident that it can play a similar role in absolute judgment, in which it is the distribution of the stimuli, not the probability of a single signal, that is important for locating the criteria. The apparent increase in the negative dependency when feedback is given can probably be attributed to the relative reduction of the positive dependency on responses produced by selective tracking.

To demonstrate that Equation 17 is capable of producing patterns of this sort, a computer simulation of the no-feedback case was run. It was not intended to fit the Ward and Lockhead (1971) data parametrically—to do this would require more information than we have at present about the effects of a stimulus or

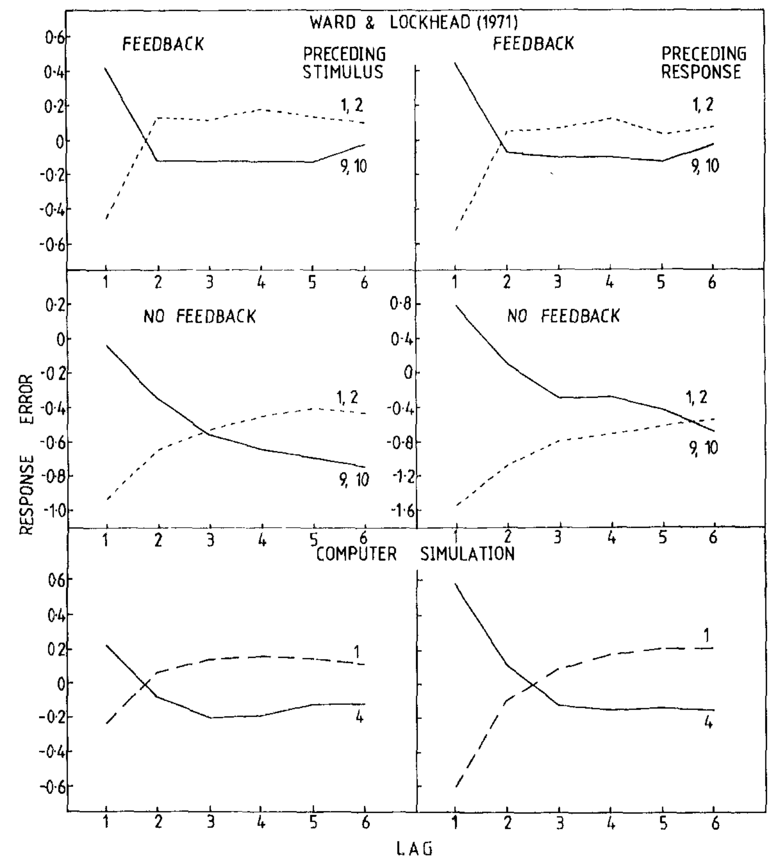


Figure 11. Absolute judgments. (The four top panels show data from Ward & Lockhead, 1971, for absolute judgments of 10 equally spaced stimuli, the softest being represented by 1 and the loudest by 10. The average departure from the nominal value of each stimulus on trial *i* is plotted against the stimuli presented or responses made on previous trials, for conditions with or without feedback. The curves when preceding stimuli at various lags are 1 or 2, or 9 or 10, are shown. Computer simulations of a four-category absolute judgment experiment without feedback are shown in the lower panels. Analyses in relation to preceding stimuli are shown on the left, in relation to preceding responses are shown on the right.)

response on criteria remote from those determining the response—but to examine whether the present theory would produce a qualitatively similar pattern of results. The simulation was based on the following assumptions: Four equally spaced stimuli with means 1, 2, 3, and 4, and producing normally distributed effects ($SD = 0.5$) on a central scale are presented in random order on successive trials. The sensory input on each trial is compared with criteria with initial values $z_{c1} = 0.75$, $z_{c2} = 2.5$, and $z_{c3} = 4.25$, to give one of four responses. (If $z_i < z_{c1}$ then 1, if $z_{c1} \leqslant z_i < z_{c2}$ then 2, etc.) The stabilization and tracking parameters are $\Delta_s = 0.5$, $\delta_s = 0.05$, $\Delta_r = 2.0$, and $\delta_r = 1.0$, giving $D = 1$. The short-term mechanisms are assumed to act on every trial.

At the beginning of each trial the criteria are updated. For this purpose rules are needed to prescribe adjustments to remote criteria following a given input and response. For the purpose of the simulation, the following rules were employed. (a) A past response affects every criterion equally. Thus, if the response on trial $i - k$ was 2, then on trial i, z_{c1} is decremented by $\max(\Delta_r - k\delta_r, 0)$, and z_{c2} and z_{c3} are both raised by this quantity. (b) A previous input-criterion difference affects neighboring criteria more than distant ones, the effect decreasing inversely with distance. Thus, if the response on trial $i - k$ was 2, we must have had $z_{c1} \leqslant z_{i-k} < z_{c2}$ on that trial, giving two input-criterion differences of absolute magnitudes $Dif_u = \Delta_s(z_{i-k} - z_{c1})$ and $Dif_d = \Delta_s(z_{c2} - z_{i-k})$. Then on trial i, z_{c1} is raised by $\max(Dif_u - k\delta_s, 0)$, z_{c2} is decreased by $\max(Dif_d - k\delta_s, 0)$, and z_{c3} is decreased by this quantity divided by 2. For a third criterion the divisor would be 3. In updating criteria the program considered only the previous eight trials. Three thousand trials were simulated and the results are shown in Figure 11.

These results show a general parallel to Ward and Lockhead's (1971) no-feedback condition. It is likely that changes in parameters and in the rules for the effects on remote criteria would allow still greater agreement.

Lockhead and King (1983) made two interesting observations using their successive ratio-judgment task. First, if two successive stimuli, I_{n-1} and I_n, are identical in auditory intensity they do not necessarily give an average ratio judgment of one. Furthermore, the result given is affected by preceding stimulus presentations. Some of their results for stimulus pairs of the same intensity are shown in the left panel of Figure 12: As the stimulus pair increases in intensity, the subjective ratio of the second presentation to the first increases. And if I_{n-2} was a low stimulus the judgment is higher than if I_{n-2} was high. We have examined whether our model would give a similar result. Similar ratios were calculated from the simulation illustrated in Figure 11. Mean ratios of the responses given to successive stimuli of the same nominal value are shown in the right-hand panel of Figure 12; the parameter for each curve is the nominal value of the stimulus preceding the pair. Except for a decrease in two of the curves at the highest stimulus value, the results parallel those of the experiment: The ratios increase with stimulus intensity, and the values are higher if I_{n-2} was low than if it was high.

There is not much evidence on the extent to which and the range over which responses or sensory inputs may produce shifts in remote parallel criteria. Purks et al. (1980, Figure 3) show shifts in absolute judgment criteria z_{cj} away from the preceding stimulus. These shifts appear to increase initially as the separation of the criterion from the previous stimulus increases, but return to zero at greater distances. On the present interpretation, this positive effect is a function of the preceding response (not stimulus), and the initial increase in the magnitude of the shift with distance may simply reflect the increasing proportion of the responses to the preceding stimulus that lie on the same side of z_{cj} as it becomes more remote. Holland and Lockhead's (1968) data also suggest that the induced positive shift may be greater for nearer criteria, but Ward and Lockhead (1970) provide evidence that the shift is the same for all criteria, suggesting that response-induced tracking may have a similar effect on the whole scale.

Some further points deserve comment. A phenomenon identified by Parducci (1965) may be explained by the stabilization strategy. He showed that in category judgments there is a tendency for category boundaries to shift so as to equate the frequencies with which different responses are used. Such an effect could be caused by stabilization. Consider three criteria distributed as in Figure 1-4, and

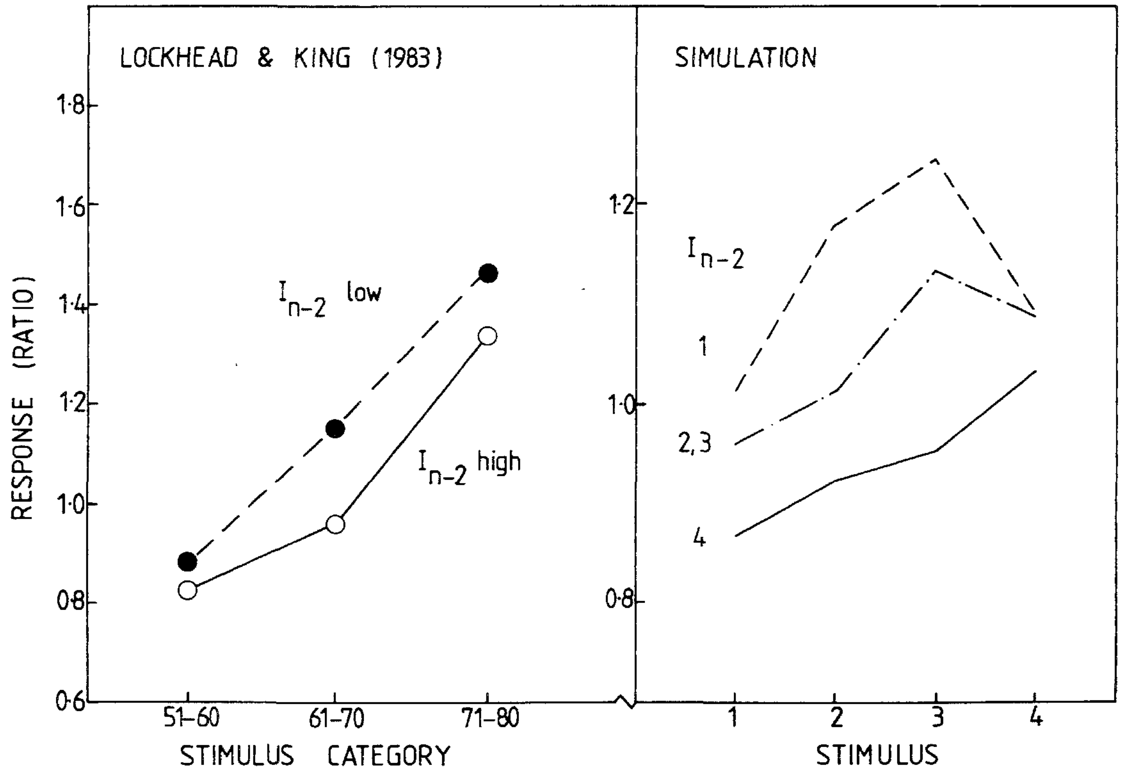


Figure 12. Successive ratio judgments. (Data from Lockhead & King, 1983, are shown on the left. Thirty stimuli ranging from 51 to 80 db in 1 db steps were presented to subjects who judged the subjective ratio of the current stimulus (I_n) to the preceding one (I_{n-1}). For cases in which $I_n = I_{n-1}$, the subjective ratios obtained are plotted for low, intermediate and high stimuli, and for I_{n-2} low or high. On the right similar ratios obtained from the simulation of absolute judgments of four stimuli illustrated in Figure 11 are shown. Here 1 is a low and 4, a high stimulus.)

suppose that the stimulus s_3 is presented more frequently than the stimulus s_2. Then more sensory inputs will lie in the interval $z_{c3} < z_i < z_{c2}$, generating the response 3, than will lie in the interval $z_{c2} < z_i < z_{c1}$. The effect of each sensory input falling in the first interval will be a tendency to shift z_{c3} upwards and z_{c2} down, which will extend over several subsequent trials: A similar effect will apply for inputs falling in the second interval. Because s_3 is more frequent, so that more sensory inputs fall between z_{c3} and z_{c2}, the net effect will be to move z_{c2} downwards, thus decreasing the number of occasions when inputs are classified as 3 and increasing the number eliciting the response 2. If s_3 is more frequent than s_4 there will be a similar net upward shift of z_{c3}, narrowing the interval from z_{c3} to z_{c2}. The overall effect will be to reduce an initial excess of 3 responses.

Ward and Lockhead (1970) reported that shifts in the stimulus range from one day to the next result in constant errors related to the previous day's range. This illustrates the relative long-term stability of reference criterion values. Ward and Lockhead (1971) performed a mock absolute judgment experiment in which no stimuli were presented, subjects were required to guess, and feedback was given. The main sequential dependency shown was a strong positive effect of the previous stimulus (i.e., the feedback given), with no effect of previous responses (guesses). The absence of a response effect, with a strong positive effect of feedback is what would be expected if feedback-based tracking were substituted for response tracking: When responses are highly unreliable only feedback inputs can usefully be used to generate tracking shifts. The same authors obtained almost identical results in a very difficult line-length absolute judgment task with feedback. The perceptual difficulty of the task implies that the sensory differences were small and so any negative dependency would be small, and only a low proportion of responses would be validated by feedback. The absence of any effect of responses, with a positive effect of feedback, implies that subjects employed the feedback-based tracking strategy here, as in the guessing experiment.

Further evidence on how information fed back to the subject may be used is provided by an experiment of Wagner and Baird (1981) that is similar to Ward and Lockhead's (1971) guessing experiment. Again, subjects guessed integers and were given feedback. In this case, however, they were told that they were participating in an extrasensory perception experiment. These results also demonstrate a strong positive effect of feedback at lag 1, but there is also a somewhat weaker positive effect of the response made, extending back for several trials. It appears that here—perhaps because the subjects' responses were dignified as possible ESP rather than guesses—positive dependencies were induced both by recent feedback and by past responses, the two tracking mechanisms having different parameters. If this is the correct account, feedback may not only be substituted for responses, it may also serve in its own right as a parallel basis for a third system of short-term adjustments.

In both the guessing experiments the analyses in terms of feedback show a small contrast effect. The feedback, being a number, is identical in form to the missing stimulus (an unknown integer) that the subject is attempting to divine in each case. It is possible that it was processed as such a stimulus, producing a stabilization effect.

Magnitude estimations of tones obtained by Ward (1973) and plotted by Jesteadt et al. (1977) in relation to preceding stimuli show sequential effects. There is a marked positive dependency extending back for at least five trials, and no evidence of contrast at any lag. A multiple regression analysis of their own magnitude-estimation data by Jesteadt et al. (1977) shows a large positive regression coefficient on the previous response (0.38) and a much smaller negative regression coefficient on the previous stimulus (−0.05). Such results accord with the account we have given. As there is no feedback, all responses are effective in producing a positive dependency, which may therefore be considerable. As responses are correlated with stimuli, it is possible to see this positive dependency in an analysis against past stimuli. The following considerations suggest that stabilization is negligible in this procedure. The defining feature of magnitude estimation is that the subject has unlimited scope to choose the form of his response. We may view the effect of this on the input-criterion differences as follows. The numerical system with which the subject compares his sensory inputs corresponds to the set of criteria in an absolute-judgment task, with the difference that in magnitude estimation the subject may employ numerous criteria and can interpolate additional, more closely spaced criteria (based on finer subdivisions of the numerical scale) whenever there is any significant separation between a sensory input and the nearest criterion initially employed. This will ensure that input-criterion differences are small and so remove or greatly diminish the basis for a negative dependency. Thus the absence of such a dependency in Ward's (1973) data and its small size in the data of Jesteadt et al. (1977) is readily understandable.

Conclusion

We have described and reviewed experimental evidence that provides support for a model of criterion setting in which the criterion is given by a linear combination of a long-term reference value and indicator traces determined by previous stimuli and responses and in some cases by feedback. This theory has been shown to provide an account for many features of sequential dependencies.

If the arguments presented here are correct, then sequential effects should no longer be viewed as bothersome errors arising from instabilities or inadequacies in our imperfectly designed detection and categorization systems, hindrances to the experimenter. Instead they provide evidence for the existence of a sophisticated control system that is able to combine information from a number of sources to determine the best position of the decision criterion from moment to moment. Factors such as past experience, knowledge of signal probability, and perhaps privileged sensory inputs such as stimuli denominated as anchors or a constant background, determine an initial long-term reference criterion. This may be differently weighted according to its reliability in different tasks. Best determinations of the current state of the world embodied in the subject's most recent responses, and weighted by their immediate relevance, produce shifts in criteria calculated to facilitate repeating those observations that appear currently valid: the positive sequential effect. We reject the idea that this represents a confusion of the present stimulus with, or assimilation of it to, a past input. Rather, it results from a rational response to the short-term nonrandomness of the real world. And because this nonrandomness is short-term—the longer we wait the less likely are we to find the same things still about us—the effect normally decays quite rapidly.

Additionally, evidence about the prevailing flux of sensory inputs, usually sampled over longer intervals to increase its reliability, is used to place criteria centrally in relation to this flux, so stabilizing criteria at positions that favor maximum information transmission. Thus Equation 17 combines past experience, an assessment of the current state of the world, and a requirement for optimal information transmission.

If this interpretation is right, sequential effects reveal the intelligence of the sensory system in operation. Application of the model to observations in the literature has revealed no serious difficulties and has suggested a number of refinements. There is a small amount of evidence suggesting the possibility that the direction of tracking may be reversed, so that responses may generate negative dependencies, perhaps because of particular difficulty in stabilizing the criterion—which we may understand as the suppression of tracking when every resource must be recruited for stabilization. Although the negative response dependency shown by three subjects in Haller's (1969) experiment (see Figure 9) did not reach significance, similar effects can sometimes be found in other data. For example, two of Sandusky's (1971) subjects showed a positive and two a negative response dependency. As we have seen, we cannot say whether the negative dependencies for the noise-only condition of Experiment 1 represent reversed tracking or normal stabilization. But there is no evidence from any source to suggest that stabilization can reverse its direction.

The theory suggested ways in which feedback information might be used, and the literature review has provided relevant evidence. We have seen that in detection and identification, the presence of feedback may cause stabilization to be suppressed, so that the long-term criterion (which takes account of the information, such as $P(s)$, that feedback supplies) is weighted more highly. But stabilization suppression does not occur when a number of response categories must be used. In no case has evidence been found for selective stabilization. Tracking, however, can be modified in detail: Selective tracking may occur when feedback information can be used to improve it by invalidating those responses with which it conflicts. Feedback-based tracking may substitute entirely for response tracking when the latter is highly unreliable, and both response and feedback-based tracking may coexist. Selective tracking or feedback-based tracking may explain such varied effects as the detection criterion shifts seen when only signal (or noise) trials receive random feedback (Friedman et al., 1968), the release of tracking in no-feedback conditions in absolute judgment, and the positive dependency on the previous stimulus (i.e., on the feedback) but not on the previous response in absolute judgment with extremely difficult or nonexistent stimuli (Ward & Lockhead, 1971). The transition from positive dependencies at short lags to negative dependencies at long lags seen in absolute judgment provides a demonstration of the conjoint operation of the different mechanisms, and the negligible or absent negative dependency in magnitude estimation is also explained by criterion-setting theory.

If it is correct that selective tracking can occur, but not selective stabilization, this has interesting implications. It suggests that stabilization may occur at an early level in a hierarchy of sensory processing at which it is susceptible to suppression by higher mechanisms, or alteration in its parameters, but not to step by step control, but that response and fed back information are made use of at a higher level in the hierarchy that is susceptible to modification on a trial-by-trial basis.

The problem of the maintenance of a criterion during periods when the appropriate stimulation does not occur was discussed. A latent stabilization strategy may explain this. Some support for this strategy is given by the observation that the data for Subject S5 in Experiment 1 were best fitted on the assumption that for him latent stabilization continued during the experiment.

Alternatives to the IT model were considered. A linear additive learning model can be rejected, but an additive learning model with exponential decay does appreciably better, although marginally not as well as the IT model. But it does provide a reasonable approximation to the latter. However, it does not account for the direct evidence of linear decay, and it is not evident that it could be elaborated in any simple way to handle possibilities such as the simultaneous operation of different strategies, including latent stabilization and response-based and feedback-based tracking, each with different decay rates.

If the IT model is valid, its implications for memory are of interest. The indicator traces record the implications for decision, not the occurrence, of past sensory inputs or responses. They decay, apparently linearly, at rates that may differ for different types of trace. These decay rates do not represent an unavoidable loss of information from the system but are parameters that weight the information provided by a past event by its current significance for decision and that it may be open to the system to modify.

Many points remain open that require further investigation. Criterion-setting theory may be of importance not only in relation to the features of sequential dependencies that we have discussed. Papers in preparation consider its application to a number of further problems in psychophysics (Treisman, 1983a, 1983b; Treisman & Faulkner, 1983a, 1983b, in press).

References

Bartlett, F. C. (1932). *Remembering.* Cambridge, England: Cambridge University Press.

Birdsall, T. G. (1955). The theory of signal detectability. In H. Quastler (Ed.) *Information Theory in Psychology.* Glencoe, IL: Free Press.

Boneau, C. A., & Cole, J. L. (1967). Decision theory, the pigeon, and the psychophysical function. *Psychological Review, 74,* 123–135.

Carterette, E. C., Friedman, M. P., & Wyman, M. J. (1966). Feedback and psychophysical variables in signal detection. *Journal of the Acoustical Society of America, 39,* 1051–1055.

Christensen, I. P., & Huang, Y. L. (1979). The auditory tau effect and memory for pitch. *Perception and Psychophysics, 26,* 489–494.

Cohen, J., Hansel, C. E. M., & Sylvester, J. D. (1954). Interdependence of temporal and auditory judgments. *Nature (Lond.), 174,* 642–644.

Collier, G. (1954a). Intertrial association at the visual threshold as a function of intertrial interval. *Journal of Experimental Psychology, 48,* 330–334.

Collier, G. (1954b). Probability of response and intertrial association as functions of monocular and binocular stimulation. *Journal of Experimental Psychology, 47,* 75–83.

Collier, G., & Verplanck, W. S. (1958). Nonindependence of successive responses at the visual threshold as a function of interpolated stimuli. *Journal of Experimental Psychology, 55,* 429–437.

Dorfman, D. D. (1973). The likelihood function of additive learning models: sufficient conditions for strict log-concavity and uniqueness of maximum. *Journal of Mathematical Psychology, 10,* 73–85.

Dorfman, D. D., & Biderman, M. (1971). A learning model for a continuum of sensory states. *Journal of Mathematical Psychology, 8,* 264–284.

Dorfman, D. D., Saslow, C. F., & Simpson, J. C. (1975). Learning models for a continuum of sensory states reexamined. *Journal of Mathematical Psychology, 12,* 178–211.

Dusoir, A. E. (1975). Treatments of bias in detection and recognition models: a review. *Perception and Psychophysics, 17,* 167–178.

Dusoir, A. E. (1980). Some evidence on additive learning models. *Perception and Psychophysics, 27,* 163–175.

Fernberger, S. W. (1931). On absolute and relative judgments in lifted weight experiments. *American Journal of Psychology, 43,* 560–578.

Friedman, M. P., Carterette, E. C., Nakatani, L., & Ahumada, A. (1968). Comparisons of some learning models for response bias in signal detection. *Perception and Psychophysics, 3,* 5–11.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Gundy, R. F. (1961). Auditory detection of an unspecified signal. *Journal of the Acoustical Society of America, 33,* 1008–1012.

Haller, R. W. (1969). *The role of information feedback in signal recognition.* Unpublished dissertation, Stanford University, Department of Psychology.

Holland, M. K., & Lockhead, G. R. (1968). Sequential effects in absolute judgments of loudness. *Perception and Psychophysics, 3,* 409–414.

Howarth, C. I., & Bulmer, M. G. (1956). Non-random sequences in visual threshold experiments. *Quarterly Journal of Experimental Psychology, 8,* 163–171.

Jesteadt, W., Luce, R. D., & Green, D. M. (1977). Sequential effects in judgments of loudness. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 92–104.

John, I. D. (1973). Sequential effects in absolute judgments of loudness without feedback. *In Attention and Performance IV* (pp. 313–326). New York: Academic Press.

Kac, M. (1962). A note on learning signal detection. *IRE Transactions on Information Theory, IT-8,* 126–128.

Kinchla, R. A. (1966). A comparison of sequential effects in detection and recognition (Tech. Rep. No. 1) New York: New York University, Experimental Psychology Series, Psychology Department.

King, M. C., & Lockhead, G. R. (1981) Response scales and sequential effects in judgment. *Perception and Psychophysics, 30,* 599–603.

Kubovy, M., & Healy, A. F. (1977). The decision rule in probabilistic categorization: What it is and how it is learned. *Journal of Experimental Psychology: General, 106,* 427–446.

Levitt, H. (1967). Testing for sequential dependencies. *Journal of the Acoustical Society of America, 43,* 65–69.

Lockhead, G. R., & King, M. C. (1983). A memory model of sequential effects in scaling tasks. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 461–473.

Long, J. (1980). Effect of prior context on two-choice absolute judgments without feedback. In R. S. Nickerson (Ed.), *Attention and Performance VIII,* Hillsdale, NJ: Erlbaum.

McNicol, D. (1980). Multiple regression analyses of sequential effects in loudness judgements. In R. S. Nickerson (Ed.), *Attention and Performance VIII* (pp. 179–196). Hillsdale, NJ: Erlbaum.

Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review, 72,* 407–418.

Parducci, A., & Sandusky, A. (1965). Distribution and sequence effects in judgment. *Journal of Experimental Psychology, 69,* 450–459.

Parducci, A., & Sandusky, A. J. (1970). Limits on the applicability of signal detection theories. *Perception and Psychophysics, 7,* 63–64.

Pfaffmann, C. (1935). An experimental comparison of the method of single stimuli and the method of constant stimuli in gustation. *American Journal of Psychology, 47,* 470–476.

Purks, S. R., Callahan, D. J., Braida, L. D., & Durlach, N. I. (1980). Intensity perception. X. Effect of preceding stimulus on identification performance. *Journal of the Acoustical Society of America, 67,* 634–637.

Sandusky, A. (1971). Signal recognition models compared for random and Markov presentation sequences. *Perception and Psychophysics, 10,* 339–347.

Sandusky, A., & Ahumada, A. (1971). Contrast in detection with gated noise. *Journal of the Acoustical Society of America, 49,* 1790–1794.

Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill.

Staddon, J. E. R., King, M., & Lockhead, G. R. (1980). On sequential effects in absolute judgment experiments. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 290–301.

Tanner, T. A., Haller, R. W., & Atkinson, R. C. (1967). Signal recognition as influenced by presentation schedules. *Perception and Psychophysics, 2,* 349–358.

Tanner, T. A., Rauk, J. A., & Atkinson, R. C. (1970). Signal recognition as influenced by information feedback. *Journal of Mathematical Psychology, 7,* 259–274.

Thomas, E. A. C. (1973). On a class of additive learning models: Error-correcting and probability matching. *Journal of Mathematical Psychology, 10,* 241–264.

Thomas, E. A. C. (1975). Criterion adjustment and probability matching. *Perception and Psychophysics, 18,* 158–162.

Thomas, E. A. C., & Legge, D. (1970). Probability matching as a basis for detection and recognition decisions. *Psychological Review, 77,* 65–72.

Treisman, M. (1963). Temporal discrimination and the indifference interval: Implications for a model of the "internal clock." *Psychological Monographs, 77,* (13, Whole No. 576).

Treisman, M. (1964). The effect of one stimulus on the threshold for another: an application of signal detectability theory. *British Journal of Statistical Psychology, 17,* 15–35.

Treisman, M. (1969). Measurement of sensory discrimination. In A. R. Meetham & R. A. Hudson (Eds.),

Encyclopaedia of Linguistics, Information and Control. Oxford: Pergamon.
Treisman, M. (1976). On the use and misuse of psychophysical terms. *Psychological Review, 83,* 246–256.
Treisman, M. (1983a). *The magical number seven and some other features of absolute judgment: An application of criterion setting theory.* Manuscript submitted for publication.
Treisman, M. (1983b). *A theory of criterion setting: An alternative to the attention band and response ratio hypotheses in magnitude estimation and cross-modality matching.* Manuscript submitted for publication.
Treisman, M. (in press). Contingent aftereffects and situationally-coded criteria. *Annals of the New York Academy of Sciences.*
Treisman, M., & Faulkner, A. (1983a). *Criteria can interchange locations.* Manuscript submitted for publication.
Treisman, M., & Faulkner, A. (1983b). *The generation of random sequences by human subjects: A model based on a theory of criterion setting.* Manuscript submitted for publication.
Treisman, M., & Faulkner, A. (in press). The setting and maintenance of criteria representing levels of confidence. *Journal of Experimental Psychology: Human Perception and Performance.*
Treisman, M., & Leshowitz, B. (1969). The effects of duration, area, and background intensity on the visual intensity difference threshold given by the forced-choice procedure: Derivations from a statistical decision model for sensory discrimination. *Perception and Psychophysics, 6,* 281–296.
Treisman, M., & Watts, T. R. (1966). Relation between signal detectability theory and the traditional procedures for measuring sensory thresholds: Estimating d' from results given by the method of constant stimuli. *Psychological Bulletin, 66,* 438–454.
Verplanck, W. S., Collier, G. H., & Cotton, J. W. (1952). Nonindependence of successive responses in measurements of the visual threshold. *Journal of Experimental Psychology, 44,* 273–282.
Wagner, M., & Baird, J. C. (1981). A quantitative analysis of sequential effects with numeric stimuli. *Perception and Psychophysics, 29,* 359–364.
Ward, L. M. (1973). Repeated magnitude estimations with a variable standard: sequential effects and other properties. *Perception and Psychophysics, 13,* 193–200.
Ward, L. M., & Lockhead, G. R. (1970). Sequential effects and memory in category judgments. *Journal of Experimental Psychology, 84,* 27–34.
Ward, L. M., & Lockhead, G. R. (1971). Response system processes in absolute judgment. *Perception and Psychophysics, 9,* 73–78.
Wever, E. G., & Zener, K. E. (1928). The method of absolute judgment in psychophysics. *Psychological Review, 35,* 466–493.
Wickelgren, W. A. (1967). Exponential decay and independence from irrevelant associations in short-term recognition memory for serial order. *Journal of Experimental Psychology, 73,* 165–171.
Wickelgren, W. A. (1968). Unidimensional strength theory and component analysis of noise in absolute and comparative judgments. *Journal of Mathematical Psychology, 5,* 102–122.

Received March 31, 1982
Revision received August 30, 1983 ■

Search Opens for Editor of *Contemporary Psychology*

The Publications and Communications Board has opened nominations for the editorship of *Contemporary Psychology* for the years 1986–1991. Donald Foss is the incumbent editor. Candidates must be members of APA and should be available to start receiving and processing books for review in early 1985 to prepare for issues published in 1986. To nominate candidates, prepare a statement of one page or less in support of each candidate. Submit nominations no later than February 1, 1984, to:

Anne D. Pick
ATTN: Journal Search Committee
Institute for Child Development
University of Minnesota
51 East River Road
Minneapolis, Minnesota 55455