



# Spatial Frequency Discrimination: Visual Long-term Memory or Criterion Setting?

MARTIN LAGES, MICHEL TREISMAN\*

*Received 3 January 1997; in revised form 28 April 1997*

**A long-term sensory memory is believed to account for spatial frequency discrimination when reference and test stimuli are separated by long intervals. We test an alternative proposal: that discrimination is determined by the range of test stimuli, through their entrainment of criterion-setting processes. Experiments 1 and 2 show that the 50% point of the psychometric function is largely determined by the midpoint of the stimulus range, not by the reference stimulus. Experiment 3 shows that discrimination of spatial frequencies is similarly affected by orthogonal contextual stimuli and parallel contextual stimuli and that these effects can be explained by criterion-setting processes. These findings support the hypothesis that discrimination over long intervals is explained by the operation of criterion-setting processes rather than by long-term sensory retention of a neural representation of the stimulus. © 1998 Elsevier Science Ltd. All rights reserved.**

Spatial frequency    Discrimination    Sensory memory    Criterion setting    Sequential dependencies

## INTRODUCTION

A sensory memory is traditionally believed to mediate visual discrimination in experiments in which the reference and test stimuli are separated by a time interval. At one time sensory memory was thought of as a continuing trace of the effects of the stimulus. It is now more often thought of as a neural representation of the parameters of the stimulus.

In many discrimination tasks two stimuli are presented in sequence,  $S_A$  followed by  $S_B$ , and the subject must judge  $S_B$  in relation to  $S_A$ . A common assumption is that a representation of  $S_A$  is stored and retained in memory, and the subject makes the judgement by comparing the sensory input from  $S_B$  with the memory trace of  $S_A$ . For example, Regan (1985, p. 619) describes spatial frequency discrimination in terms of four successive neural processes: (1) neurally encode the first grating's spatial frequency; (2) store this representation; (3) encode the second spatial frequency; (4) compare the two neural representations. (For convenience we shall refer to accounts of this type as "sensory memory theory".) Using two-interval forced choice (2IFC) and the method of constant stimuli (MCS), with a range of inter-stimulus intervals (ISIs), Regan measured spatial frequency discrimination for pairs of parallel gratings or for pairs of gratings that were orthogonal. Regan reported no significant deterioration in the spatial frequency discrimi-

nation threshold (the standard deviation of the psychometric function) as interstimulus interval increased from 0.4 to 20 sec, for both parallel and orthogonal gratings. Regan concluded that the stored neural representation undergoes no appreciable decay over this range of intervals.

Magnussen, Greenlee, Asplund, & Dyrnes (1990) confirmed this result: they obtained "perfect" retention in visual short-term memory for up to 30 sec (the longest ISI they tested) of the spatial frequency information in an initial exposure of 60–200 msec. Magnussen and Dyrnes (1994) increased the range over which "perfect" sensory memory extends to 50 hr.

It is surprising that the visual system should be able to store a neural representation of a stimulus with perfect precision over such intervals. However, there is a problem with the argument. The claim for a strong link between the initial stimulus and later judgements rests on the failure to find any effect of change in ISI on performance. But we would have the same result if the reference stimulus had no effect whatever on later performance, and the precision of the latter were due to other causes. The data do not allow us to choose between perfect retention of a long-term sensory memory and a discrimination process that does not involve such a memory at all, except on the prior assumption that it is impossible for good discriminative performance to occur in the absence of a long-term memory mechanism. However, it has long been known (Wever & Zener, 1928; Guilford, 1936; Treisman, 1963) and more recently confirmed (Westheimer & McKee, 1977; Morgan, 1992) that in the absence of a standard, discriminative

\*To whom correspondence may be addressed [Fax: +44 1865 310447; Email: michel.treisman@psy.ox.ac.uk].

Department of Experimental Psychology, Oxford University, South Parks Road, Oxford OX1 3UD, U.K.

judgements may be made that are related to the mean of the stimuli presented. More recently, a theoretical account has been put forward that can explain such observations and provides an alternative to the theory of long-term sensory memory, namely criterion-setting theory (Treisman, 1984, 1985, 1987; Treisman & Faulkner, 1984a,b, 1985, 1987; Treisman & Williams, 1984).

Signal detection theory (SDT; Green & Swets, 1966, 1974) attributes discrimination to a decision process that compares the incoming sensory input, as registered on the decision axis, with a response criterion whose value is determined by parameters such as signal probability and the values and costs of different outcomes. This contrasts with sensory memory theory which attributes discrimination to a comparison of the incoming sensory input with the neural representation of a past sensory input. Criterion-setting theory (CST) is an extension of SDT which proposes that the response criterion is not constant during a session but changes in value from trial to trial under the impact of processes aimed at optimizing performance. These processes also provide an explanation for the occurrence of sequential dependencies, which are understood not as errors but as an expression of the working of the criterion-setting system. As in SDT, CST assumes that discrimination may make use of a long-term reference decision criterion. But it assumes in addition that the sensory system computes trial-to-trial adjustments of the decision criterion, with the object of ensuring that the criterion on each trial is optimal for that decision. Two types of short-term adjustments are made, determined by two criterion-setting processes. These adjustments are the basis of intertrial dependencies.

#### *The stabilization mechanism*

During a period of observation, the flux of sensory inputs that arrive at the sensory decision axis may be badly placed in relation to the position of the decision criterion, or their location may shift over time. For example, the long-term reference criterion might be in a poor position initially, or sensory adaptation might gradually change the magnitudes of the inputs, or the level of internal noise might alter. If this causes the criterion to be too low, say, in relation to the inputs, most inputs will lie above it, so that the responses will convey little information. The purpose of the stabilization mechanism is to maintain the criterion at the most useful location on the decision axis, towards the mean of the flux of sensory inputs; that is, its aim is to stabilize the position of the criterion relative to the changing input. Stabilization is a negative feedback system which prevents the criterion becoming displaced from the location of the incoming inputs, even when the latter drift.

The stabilization mechanism does this by shifting the criterion on each trial in a way that will tend to maintain the criterion at a central position in relation to signal and noise inputs. Following each input, it causes the criterion

to move toward the position of that input. Thus, if the sensory input on a given trial is above the criterion, stabilization raises the criterion. If more inputs fall above the criterion than below it, the net effect is a trend for the criterion to move upwards. This mechanism produces an intertrial dependency. If the input on trial  $t$  falls above the criterion (and so evokes the response HIGH, in a frequency discrimination task), the consequent rise in criterion will reduce the probability of the response HIGH on the next trial. This effect is known as contrast, or preferably (to avoid confusion when discussing visual stimuli) as a negative sequential effect (Treisman & Williams, 1984). The dependency is on preceding stimulus inputs.

#### *The tracking mechanism*

SDT uses the prior probability of the signal, regarded as a fixed parameter of an observation session, as a main determinant of the long-term criterion value. However, in everyday experience signal probability may change continually: in the real world, if the hunter catches a glimpse of the prey, the probability that the prey may be observed a moment later rises. For optimality, the criterion should respond to each probability change. Each observation the subject makes provides fresh evidence about the current probability of observing the stimulus. To take account of this information, the CST tracking mechanism lowers the criterion after a positive response, and raises it after a negative response. The effect is to increase the probability of repeating the same response. In a sequence of laboratory trials the corresponding intertrial effect is a positive sequential effect ("assimilation") that is dependent on past responses. Tracking is a positive feedback system which allows the criterion to adjust to fluctuations in the probability of the stimulus.

The criterion-calculation process uses the information from each short-term adjustment mechanism in a similar way. For example, suppose that a stimulus is presented on trial  $t$  and generates a sensory effect  $e_t$  on the decision axis. The stabilization mechanism records the difference between  $e_t$  and the current value of the criterion on that decision axis,  $e_c$ . This difference,  $e_t - e_c$  (multiplied by a weighting constant), represents the amount by which stabilization requires the criterion to be shifted towards  $e_t$ . This value is stored as a "stabilization indicator memory trace" in a memory store dedicated to the criterion  $e_c$ . What this indicator trace retains in memory is not a representation of the stimulus but a measure (of the disparity between input and criterion) that is required to specify a later adjustment in criterion  $e_c$ . The memory store may also hold similar indicator traces from earlier exposures to the stimulus. On trial  $t + 1$ , the sum of these indicator traces is added to the reference value of the criterion. This gives an effective criterion value for that trial that is shifted toward the mean position of previous sensory inputs. As time passes and more recent trials supervene, older inputs become less relevant and should be given less weight by the criterion-calculation process.

This is achieved by decreasing the (absolute) value of each stabilization indicator trace by a constant  $\delta_s$  on each trial, until this value reaches zero and the trace disappears.

The tracking mechanism operates in a similar way. Following each trial it lays down a response-dependent indicator memory trace whose sign depends on the response. The absolute value of the trace decreases by a constant  $\delta_r$  on each trial, until it disappears. Prior to the decision on each trial the criterion-calculation process adds all the stabilization and tracking indicator traces that are still in memory to the reference criterion to determine the effective criterion for that trial. Tracking indicator traces differ from stabilization traces in that their initial value is constant; the sign of the indicator trace is negative if the response was HIGH or YES (tending to lower the criterion) and positive if the response was LOW or NO; and previous work indicates that they usually decay more rapidly than stabilization traces.

Both short-term mechanisms may be active at the same time, but their effects do not cancel out because of the differences in their parameters.

Sensory memory theory and criterion-setting theory both assume that information is retained in memory, but they differ in what it is they assume is retained. In sensory memory theory it is the neural representation of a sensory input. CST does not make use of a memory of the stimulus as such. (Such traces may exist in a short- or long-term visual memory, but they do not contribute to discrimination.) What is extracted and retained in CST is a specification for modifying the criterion used in later decisions. The reasons for memory decay are different in each theory. In sensory memory theory, memory decay is a failing that results from a limitation in our ability to retain information. In CST the duration of a short-term trace is a matter of design: the life the system assigns to a stabilization indicator trace reflects a compromise between the advantage of larger samples in locating the criterion, and the decreasing relevance of older information.

CST allows long-term retention of information, in the form of a reference criterion value. Such a reference value allows an appropriate criterion to be carried over from one occasion to another, when the same salient or standard or well-practiced discrimination is performed on different occasions, for example, discriminating fusion from flicker, or discriminating departures from the vertical. What is retained here is not in the form of a neural representation of a previously seen display flickering at a very high rate, or a particular vertical line. The information retained is the value of a criterion that has been employed before. However, if a subject is required to discriminate in relation to an arbitrary level on a continuous dimension, such as a particular intensity or frequency, a long-term reference criterion value may have little role, and a criterion may be interpolated on the basis of the first few stimuli.

In CST the mechanism of discrimination is defined by the familiar signal detection model, in which inputs are

compared with response criteria (Green & Swets, 1966, 1974). Sensory memory theory assumes that a comparison is made between two neural representations of stimulus inputs, but what this implies for the mechanism of discrimination is not always defined. In some cases it can be given an interpretation in terms of the SDT model for forced choice; for an interesting model of this sort see Greenlee and Thomas (1993).

There is a crucial difference in the predictions made by sensory memory theory and criterion-setting theory that allows us to test them against each other. The former assumes that a judgement depends on two stimuli only,  $S_A$  (the reference or standard stimulus) and  $S_B$  (the stimulus to be judged). The judgement process is indifferent to other stimuli that may be presented (e.g., Regan, 1985). CST assumes that not only  $S_A$ , if this has been recently presented, but all recently presented stimuli may set up short-term indicator traces that modify the criterion and so affect later judgements. This difference provides a basis for an experimental comparison of the two theories.

## EXPERIMENT 1

The present experimental paradigm can be compared with that employed by Magnussen and Dyrnes (1994). An initial reference stimulus ( $S_A$ ) is presented. This is followed by a fixed retention interval. A series of trials then follows in which stimulus values are selected from a test range in random order, and the subject's task is to judge whether each stimulus is higher or lower than the reference value. The reference stimulus is not repeated on each trial and each test stimulus is given equally frequently, in accordance with the method of single stimuli (MSS).

The sensory memory prediction is that the location of the 50% point of the psychometric function (traditionally known as the point of subjective equality or PSE) will correspond to the value of the reference stimulus, and this will be independent of whatever range of test stimuli is used. The CST prediction is that on any trial the stabilization effects of the random sample of test stimuli presented over preceding trials will tend to centre the criterion and thus the psychometric function at their mean value (the PSE estimates the position of the criterion). The initial presentation of  $S_A$  itself also produces a stabilization indicator trace: this will tend to locate the criterion at the value of  $S_A$ , for stimuli that follow sufficiently soon after  $S_A$ , that is, until this trace has decayed away. Until then it is one member of the set of stabilization traces that determines the location of the psychometric function.

If it were found that the 50% point always coincided with  $S_A$ , whatever the range of test stimuli used, this would indicate that stabilization by the test stimuli is negligible, or that sensory memory theory is correct. If, at the other extreme, the 50% point always coincides with the midpoint of the range of test stimuli, this would indicate that the effect of  $S_A$  is negligible, and would support CST.

### Subjects

Three male and three female graduate students, aged between 22 and 25 years, participated in the experiment. They were members of the subject panel of the Department of Experimental Psychology, and were naive as to the aims of the experiment. All subjects had normal or corrected-to-normal visual acuity. They were paid for attending three hourly sessions on consecutive days, at the same time of day.

### Apparatus

The tasks were programmed in C and run on a Macintosh PowerPC 6100/60 microcomputer with a Macintosh 12 in high-resolution monochrome monitor. The monitor was a cathode-ray tube with (aluminized) PC104 and PC193 phosphor which appeared monochromatic. The monitor was calibrated with a Minolta LS-110 photometer with close-up lens, using routines from VideoToolbox (Pelli & Zhang, 1991). The luminance modulation on screen was improved by using a video attenuator which combined the red, green, and blue output signals from the computer's 8-bit DACs in order to simulate a linear 12-bit monochrome display (Pelli & Zhang, 1991). The screen had a frame rate of 66.7 Hz.

The subject responded by pressing labelled keys on the keyboard to indicate a lower or higher response. Response times were measured using the KeMo utilities (Costin, 1993). This gave an accuracy for measuring keyboard responses of  $\pm 1.7$  msec. The subject was comfortably seated in front of the screen and keyboard in a darkened cubicle with his or her head on a chin-rest (background illumination was of the order of  $6.5 \text{ cd/m}^2$ ). The subject viewed the screen binocularly at a distance of 114 cm.

### Stimuli

The images subtended  $9.6^\circ$  visual angle horizontally and  $5.0^\circ$  vertically ( $19.4 \text{ cm}$  by  $10.0 \text{ cm}$ ). They had an average luminance of  $50 \text{ cd/m}^2$ , and a Michelson contrast of 20%. The reference and test stimuli were vertical sine-wave gratings presented at random spatial phase in a gaussian envelope with a vertical and horizontal SD of 2.0 and  $3.9^\circ$ , respectively. A mask consisted of a random pattern of squares, each of which was  $5 \times 5$  pixels in size and had a luminance of either 40 or  $60 \text{ cd/m}^2$ . The stimuli were presented on a uniform blank background of mean luminance  $50 \text{ cd/m}^2$ , which subtended  $10.6$  by  $8.0^\circ$  visual angle ( $21.4$  by  $16.1 \text{ cm}$ ) and formed a surround for the stimuli and mask to reduce effects of induced contrast (McCourt & Blakeslee, 1994) or possible effects of high voltage droop (Pelli & Zhang, 1991). The uniform background was also displayed between trials.

In each session a practice block was followed by four experimental blocks. The practice block employed a reference stimulus at 1.0 cpd, and presented 22 stimuli ranging from 0.5 to 1.5 cpd. The spatial frequency of the reference grating was 2.5 cpd for the first two experimental blocks and 5.0 cpd for the last two blocks. For the

2.5 cpd reference grating a set of 11 test gratings was generated, covering a range from  $0.75 \times$  midpoint to  $1.25 \times$  midpoint. These stimuli ranged from 1.875 to 3.125 cpd at intervals of 0.125 cpd. (For example, the 11 gratings with midpoint 2.5 cpd had spatial frequencies of 1.875, 2.0, 2.125, 2.25, 2.375, 2.5, 2.625, 2.75, 2.875, 3.0, and 3.125 cpd.) Similar stimulus ranges were generated with the same interval between stimuli but with midpoints of 2.25 and 2.75 cpd. (We shall refer to the stimulus ranges by their midpoints.). For the 5.0 cpd reference stimulus, similar stimulus ranges were generated with midpoints of 4.5, 5.0 or 5.5 cpd, each consisting of 11 stimuli at intervals of 0.25 cpd. Each experimental block began with 10 warm-up trials (these trials covered a range from  $0.5 \times$  midpoint to  $1.5 \times$  midpoint and were not analysed) which were followed by 12 trials for each stimulus value in the test range employed (132 trials).

### Procedure

A session of about an hour was structured as follows: (1) a practice block whose results were not analysed. (2) A vertical grating with a spatial frequency of 2.5 cpd was presented for 10 sec as the reference grating. No fixation point was presented. (3) A retention interval of 30 sec followed. (4) A beep then signalled the onset of 10 warm-up trials, followed by 132 test trials in random order. Each test grating appeared on the monitor for 0.2 sec, followed by a mask which remained present until the response. The mask was intended to exclude any effect of afterimages. On each trial the subject judged whether the test grating had a higher or a lower spatial frequency than the reference grating. When the subject responded the mask disappeared. The uniform blank field stayed visible for 1.5 sec. Then the next trial commenced with the presentation of another test grating randomly drawn from the set of eleven. (5) The block of trials was followed by an interval of about 3–4 min. (6) A second block with 10 warm-up trials and 132 experimental trials was presented. The ranges of test stimuli used in the first two experimental blocks were centred on one of 2.25, 2.5 or 2.75 cpd. For each stimulus value there were 24 experimental trials. (7) There was an interval of about 5 min. (8) A vertical grating with a spatial frequency of 5.0 cpd was presented as the reference grating for 10 sec. (9) A retention interval of 30 sec followed. (10) A third block of 10 warm-up and 132 experimental trials was presented in random order. (11) There was an interval of about 3–4 min. (12) A fourth block of 10 warm-up and 132 experimental trials was presented. The ranges of test stimuli used in the two last blocks were centred on one of 4.5, 5.0 or 5.5 cpd.

The subjects' task was to indicate whether each test stimulus was higher or lower in spatial frequency than the reference grating. No feedback was given.

Each subject did three sessions, with a different sequence of ranges in each. The sequence was balanced in a Graeco-Latin square design (Winer, Brown, & Michels, 1991).

### Results and discussion

The main experimental question is whether the PSE is determined by the reference stimulus independently of the test range, or by the range of test stimuli with little effect of the reference stimulus. For sensory memory theory, the PSE should approximate the reference stimulus. For CST it estimates the position of the response criterion.

Gaussian cumulative distribution functions (cdf) are usually used to fit psychometric functions, but the Weibull function may also be used (Quick, 1974; Greenlee, Georgeson, Magnussen, & Harris, 1991; Magnussen & Dyrmes, 1994). The data for each subject, for each block of each session, were fitted both by a gaussian distribution and by the Weibull distribution:

$$W(x) = 1 - \exp(-(x/\alpha)^\beta). \quad (1)$$

The Weibull distribution was fitted using a maximum likelihood monotonic fit (Watson, 1979; Pelli, 1995). For comparability, a similar procedure was used to fit the gaussian distribution. The mean parameters for these fits for each condition and block are shown in Table 1. To illustrate the overall results, the data were pooled over the six subjects and two blocks, and the results are shown in Fig. 1 with gaussian psychometric functions fitted.

There are a number of features of interest in these data.

(1) The psychometric functions for the different test stimulus ranges are clearly displaced from one another. This is the result predicted by criterion-setting theory. It supports the hypothesis that discrimination performance in the present task can be understood in terms of signal detection processes and criterion-setting.

To test the significance of this finding, an analysis of variance was performed on the gaussian parameter  $\mu$  (the PSE). For this purpose the independent variable was expressed on a common scale for the 2.5 and 5.0 cpd conditions: this was an "arbitrary unit" (AU) scale, given by normalizing each test spatial frequency by division by the corresponding reference frequency. We shall refer to  $\mu/\text{Reference Frequency}$  as  $\mu'$  and midpoint/Reference Frequency (i.e. the midpoint of the range in AUs) as midpt. On the AU scale the reference grating always falls at 1.0 AU, and the lower and upper range midpoints at 0.9

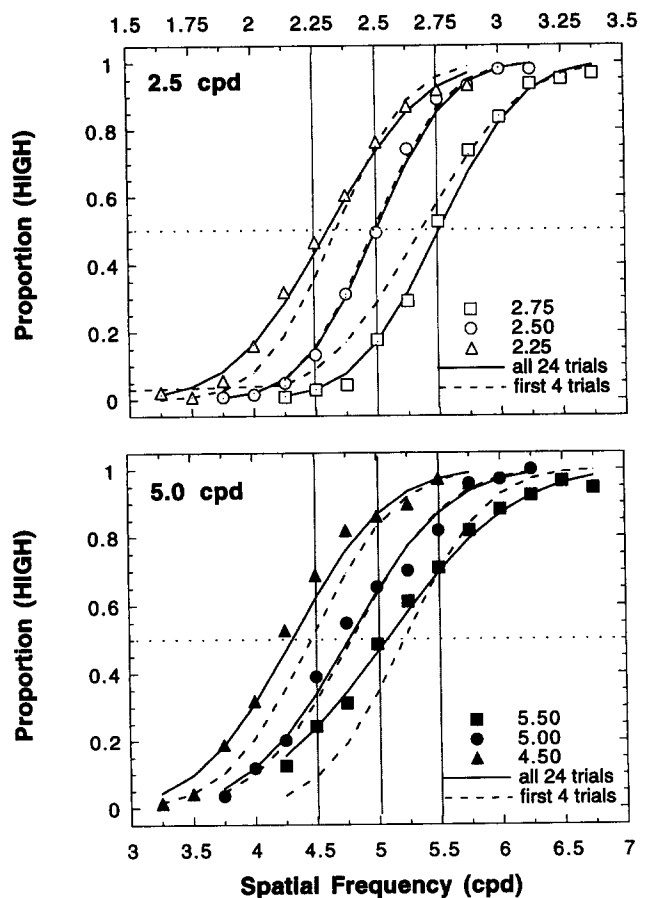


FIGURE 1. Experiment 1: gaussian psychometric functions (continuous lines) fitted to Proportion (HIGH) for two spatial frequency conditions (upper panel, 2.5 cpd; lower panel, 5.0 cpd) and three ranges of stimuli (2.25 or 4.5 cpd, triangles; 2.50 or 5.0 cpd, circles and 2.75 or 5.50 cpd, squares). Range midpoints are indicated by vertical lines. Data are collapsed over six subjects and two blocks. Similar functions fitted to the first four trials at each stimulus value only are plotted as dashed lines.

and 1.1 AU. The analysis of variance examined the factors SF Condition (2.5 and 5.0 cpd), Range (0.9, 1.0 and 1.1 AU), Block (1 and 2) and Subjects.

The analysis showed that Range was highly significant ( $F[2,10] = 41.67$ ,  $P < 0.0001$ ), verifying the displacements illustrated in Fig. 1. SF Condition approached significance ( $F[1,5] = 6.31$ ,  $P = 0.054$ ). No other significant effect was observed.

The effect of Range cannot be explained by sensory memory theory, but the present results do not exclude a possible contribution from a sensory memory of the reference grating as well. If a short-lived sensory memory has some effect on the PSE, this effect should be greater for the first block than for the second. However, neither the Block main effect ( $F[1,5] = 0.04$ , NS) nor the Block  $\times$  Range interaction ( $F[2,10] = 0.07$ , NS) were significant. However, it could be argued that a sensory memory effect may have been present for the first few trials, and lost from sight as further trials accumulated; Magnussen and Dyrmes (1994) recorded only four trials per stimulus value, in contrast to our 24. Accordingly, we repeated our analysis for the first 44 experimental trials

TABLE 1. Experiment 1: Mean parameters and  $\chi^2$  values for Weibull and gaussian psychometric functions

Range	Block	$\alpha$	$\beta$	$\sum df$	$\sum \chi^2$	$\mu$	$\sigma$ (SD)	$\sum \chi^2$
2.25	1	2.39	12.47	40	51.8	2.29	0.22	43.0
	2	2.43	11.29	42	48.3	2.32	0.26	39.0
2.50	1	2.62	13.99	43	49.7	2.52	0.20	30.9
	2	2.57	13.23	43	48.4	2.47	0.21	33.7
2.75	1	2.83	14.19	39	45.7	2.73	0.24	37.2
	2	2.89	17.33	47	39.8	2.80	0.19	26.6
4.50	1	4.57	10.47	44	52.9	4.36	0.49	40.3
	2	4.44	12.38	41	50.7	4.26	0.41	41.7
5.00	1	4.93	14.18	47	33.5	4.76	0.38	30.3
	2	4.94	17.51	48	30.8	4.79	0.32	27.3
5.50	1	5.39	10.78	43	31.1	5.15	0.57	28.7
	2	5.24	9.80	40	48.4	5.02	0.52	34.9

only (four trials per stimulus value) for each SF condition in each session, and the best-fitting gaussian cdfs for these data are superimposed on Fig. 1 as dashed lines.

It is evident that the four-trial data show the same effect as the full data. The cdfs for the three ranges are spaced apart rather than converging on one curve. The spread from the lowest to the highest  $\mu$  is slightly reduced for 2.5 cpd, but unchanged for 5.0 cpd. For the full data set the mean value of the SD of the fitted gaussian functions was 0.22 cpd for the 2.5 cpd condition, and 0.45 cpd for the 5.0 cpd condition. The corresponding values for the four-trial data were 0.25 and 0.39. Thus, psychometric functions have similar slopes at the start of a condition and for the whole condition.

If criterion location is determined by the context of test stimuli, then we should expect the PSEs to regress onto the range midpoints of the stimuli. Figure 2(A) shows the regression of the mean  $\mu'$  on the range midpoints in AUs. The best-fitting lines are given by  $\mu' = 0.91 \text{ midpt} + 0.10$

for the 2.5 cpd condition and  $\mu' = 0.78 \text{ midpt} + 0.17$  for the 5.0 cpd condition.

We conclude that in the present paradigm the position of the criterion is largely determined by the recent context provided by the test stimuli, and not by a neural representation of the reference grating. The retention interval for the reference grating was 30 sec for Block 1. The interval from the reference grating to the onset of Block 2 was 10 min. It appears that over such periods the indicator traces generated by reference gratings have a negligible effect.

(2) The near-significant effect of SF condition corresponds to the displacement of the 5.0 cpd condition psychometric functions towards lower spatial frequencies [see Fig. 2(A)].

(3) The parameters of the gaussian and Weibull fits to the psychometric functions are given in Table 1. The overall goodness of fit for each function can be tested using the sum of the  $\chi^2$  values over subjects and conditions. For the Weibull distribution the summed  $\chi^2$  value is  $\chi^2[517] = 531.1$ ,  $P = 0.326$ . For the gaussian distribution it is  $\chi^2[517] = 413.5$ ,  $P = 0.999$ . Thus, neither distribution can be excluded, but the gaussian distribution appears to give the better fit. We tested this by treating total  $\chi^2(\text{Weibull})/\text{total } \chi^2(\text{gaussian})$  as an  $F$ -ratio. This gives  $F[517, 517] = 1.28$  ( $P < 0.05$ ), indicating that the gaussian fit is significantly better.

(4) The results for the psychometric functions accord with the predictions of CST. This theory further proposes that in the laboratory the effects of the criterion-setting mechanisms may manifest as two types of sequential dependencies, a negative dependency on past stimuli, and a positive dependency on past responses.

The negative dependency results from stabilization, the mechanism that centres the criterion in relation to the range of stimulus inputs. Since we have found a significant effect of stimulus range (Fig. 1), CST requires that the data should also show a negative dependency on past stimuli. They may also show a positive dependency on past responses. To examine this, the data were divided into sets of trials preceded by a high (or low) stimulus at a given lag, or by a HIGH (or LOW) response at a given lag, and a gaussian psychometric function was fitted to each data set. The value of  $\mu$  given by this fit estimates the position of the criterion following a high (or low) stimulus, or HIGH (or LOW) response, at the given lag. For this analysis, preceding stimuli were defined as high or low by dividing each range of test stimuli into two subsets of five stimuli, those that were higher (or lower) in spatial frequency than the midpoint of the set. Preceding stimuli at the midpoint were excluded from this analysis.

(For illustration, the point corresponding to high at lag 3 was obtained by analysing the data for all trials  $t$ , such that on the corresponding trials  $t-3$  the stimulus presented was one of the five highest frequency stimuli. In view of the fact that at all other lags [1, 2, 4, etc.] the stimuli presented were randomly and independently chosen, their net stabilization effect should tend to place the criterion

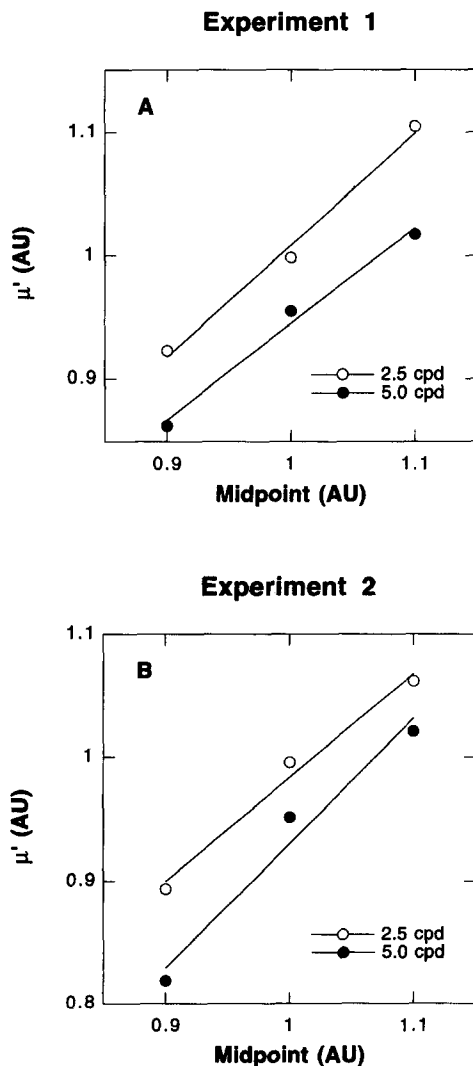


FIGURE 2. Experiment 1 (A) and Experiment 2 (B): Mean parameters  $\mu$  of gaussian psychometric functions (normalized to  $\mu'$  and averaged over six subjects and two blocks) for 2.5 cpd (empty circles) and 5.0 cpd conditions (filled circles) plotted against range midpoints (expressed in AUs).

at its overall mean position. Any deviation of the value of  $\mu$  obtained for high preceding stimuli at lag 3 from the overall mean value will specifically reflect the effect of the stabilization indicator traces generated by high stimuli at trials t-3, independently of the stimuli presented at other lags.) The different preceding stimulus levels, lags, SF conditions and ranges gave 60 data sets per subject. Values of  $\mu$  were normalized before finding the means over SF conditions. Response dependencies were analysed similarly.

Figure 3 plots mean values of  $\mu'$  for preceding high or low stimuli or HIGH or LOW responses at five different lags for three ranges, combined over spatial frequency conditions. The effects appear small. However, there is a problem in assessing the magnitudes of the negative and positive dependencies, since stimuli and the responses that are made to them are correlated, and the two dependencies tend to oppose each other. Thus, when we

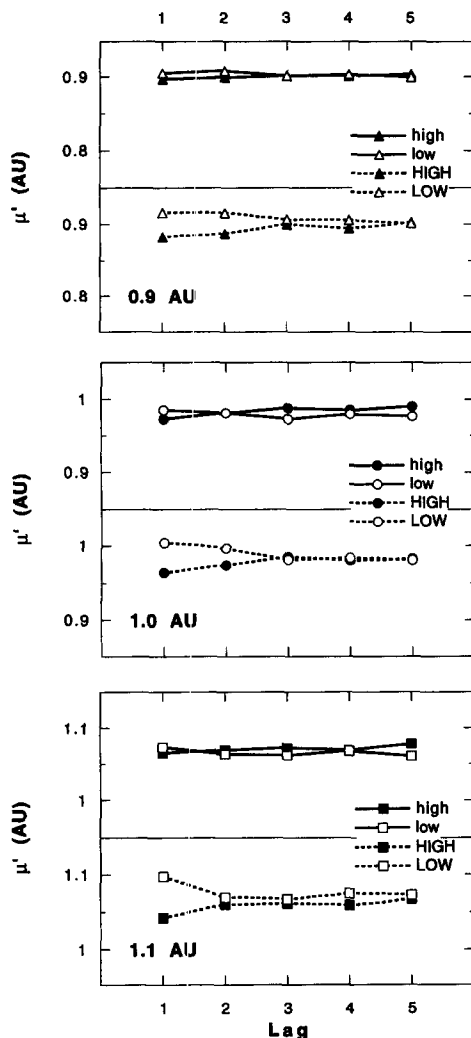


FIGURE 3. Experiment 1. Stimulus dependencies (straight lines) and response dependencies (dashed lines) are shown. Upper panel: ranges with midpoint 0.9 AU (triangles); middle panel: 1.0 AU (circles); lower panel: 1.1 AU (squares). Each point denotes the mean  $\mu'$  for a preceding low stimulus or LOW-response (empty symbols) and a preceding high stimulus or HIGH-response (filled symbols) for five lags. The data are combined over SF conditions and subjects.

examine intertrial dependencies, a positive effect of past responses may tend to hide a negative effect of the stimuli that evoked them, and vice versa. To reveal the underlying effects independently of each other, we need to control for the effects of past responses when examining stimulus dependencies, and the reverse. To do this we plot the results for each combination of preceding stimulus and response separately (see Fig. 4). The data were pooled over test stimulus ranges, SF conditions, and subjects, and gaussian cdfs were then fitted. A comparison of filled and empty symbols of the same shape illustrates the effect of preceding stimulus values, with response controlled. A comparison of circles and squares that are both filled (or both empty) illustrates the effect of preceding responses, with stimulus value controlled.

Figure 4 shows a clear positive effect of past responses at all lags examined. For example, for a preceding high stimulus at lag 1, a LOW response (filled circle) causes  $\mu'$  on the next trial to have the value 1.12 (AU), whereas a HIGH response (filled square) causes it to have the value 0.96. Similarly, for a preceding low stimulus at lag 1, a LOW response (empty circle) causes  $\mu'$  on the next trial to have the value 1.02, whereas a HIGH response (empty square) causes it to have the value 0.84. Thus, whatever the value of the preceding stimulus, a past response causes the criterion to shift in a way that will favor repetition of the same response. The difference between the effects of HIGH and LOW responses at lag 1 on the position of the criterion on the next trial is 0.175 units. At lag 5 it is 0.153 units.

Similarly, there is a clear negative effect of past stimuli. For example, at lag 1, for a preceding HIGH response a low stimulus (empty square) causes  $\mu'$  on the next trial to have the value 0.84, whereas a high stimulus (filled square) causes it to have the value 0.96. Similarly, for a preceding LOW response, a low stimulus (empty circle) and high stimulus (filled circle) give the values 1.02 and 1.12. Thus, with the preceding response

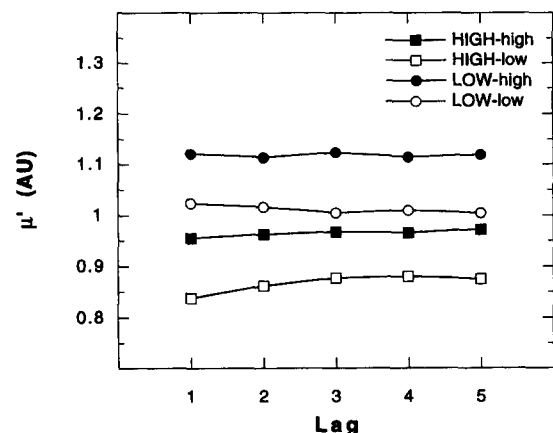


FIGURE 4. Experiment 1: Sequential dependencies as a function of four combinations of preceding stimulus and response. Parameters  $\mu'$  are plotted for each combination of preceding low and high stimulus and LOW and HIGH response for five lags. Parameters were estimated from data collapsed over two blocks, three SF conditions, three ranges, and six subjects.

controlled, a past stimulus causes the criterion to shift in a way that will favour a change in response. The difference between the effects of preceding high and low stimuli at lag 1 is 0.107 units. At lag 5 it is 0.105 units. For comparison, for this experiment the mean value of  $\sigma$ /Reference Frequency is approximately 0.09 AU.

An analysis of variance was performed on these data, for the main effects Preceding Stimulus, Preceding Response and Lag. There were no significant two-way interactions (when tested against the three-way interaction) and so all higher-order interactions were summed to give the error term for testing the main effects. Both Preceding Response ( $F[1,13] = 915.4$ ,  $P < 0.0001$ ) and Preceding Stimulus were highly significant ( $F[1,13] = 429.4$ ,  $P < 0.0001$ ) but the effect of Lag was not significant.

The present results support the CST interpretation that the position of the psychometric function is largely determined by the range of test stimuli, operating through the mechanisms that set the criterion. They do not support the hypothesis that judgements are determined by a process of comparing the test stimuli with a retained neural representation of the initial reference stimulus.

Magnussen and Dyrnes (1994) also tested the hypothesis that "intelligent guessing" (p. 101) interpolated a reference spatial frequency from the test stimuli presented. To test this they ran an experiment in which no reference grating was presented, instead "asking [the subjects] to guess the centre spatial frequency by deciding on each trial whether the test grating was above or below the imagined centre spatial frequency". They obtained very much poorer performance than when a reference grating was presented, and thus reject this hypothesis.

This result appears to present a difficulty for the present explanation. However, a problem with their procedure is that their subjects may have found these instructions more complicated than in the experiments in which a reference stimulus was presented, and this may have disrupted their performance. In Experiment 2 we re-examine the effect of omitting the reference grating.

## EXPERIMENT 2

The design, apparatus and stimuli were as before, except that no reference gratings were used.

### Subjects

Four new subjects, three female and one male, were used, whose ages ranged from 23 to 29 years. One subject had no and the others had some experience with visual discrimination tasks but all subjects were naïve as to the aim of the experiment. Their visual acuity was normal or corrected to normal.

### Procedure

As before, each subject attended three hour-long sessions on consecutive days. The range of test stimuli in the first two blocks of a session was centred on one of 2.25, 2.5 or 2.75 cpd, and in the second two blocks on 4.5,

5.0 or 5.5 cpd. The order of the ranges in each SF condition was different for each subject.

Subjects judged whether the test grating frequencies were LOW or HIGH. Before each block the subjects were instructed "to infer the midpoint from the gratings you are about to see in the following trials. You have to make a guess in your first trial but you can infer a midpoint from the following gratings to give a higher or lower response." As before, the first ten trials in each block were not analysed.

### Results and discussion

The results are shown in Fig. 5 and Table 2. The data were analysed for each subject, for each block in each session, to give the parameters whose mean values are shown in Table 2. The data were pooled over subjects to give the values plotted in Fig. 5. These data were fitted with gaussian cdfs. As before, the first four trials per stimulus value were also analysed separately for each condition.

The important result is that the psychometric functions are almost indistinguishable from those in Fig. 1. The psychometric functions for the three test ranges in each condition are correspondingly shifted. The functions are also similar in steepness to those obtained before: the standard deviations in Table 2 are of the same order as in

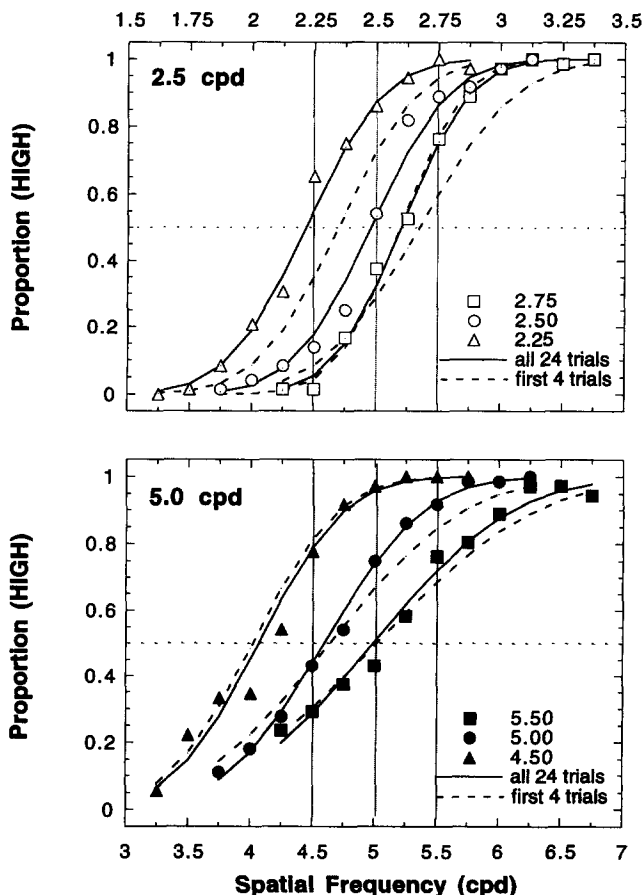


FIGURE 5. Experiment 2: gaussian psychometric functions depicted as in Fig. 1.



TABLE 2. Experiment 2: Mean parameters and  $\chi^2$  values for Weibull and gaussian psychometric functions

Range	Block	$\alpha$	$\beta$	$\sum df$	$\sum \chi^2$	$\mu$	$\sigma$ (SD)	$\sum \chi^2$
2.25	1	2.41	14.90	29	24.4	2.32	0.20	18.9
	2	2.24	12.07	28	36.0	2.15	0.19	23.3
2.50	1	2.65	17.98	25	37.9	2.56	0.20	31.8
	2	2.50	18.47	28	26.6	2.42	0.19	34.1
2.75	1	2.79	14.76	29	22.6	2.68	0.22	16.2
	2	2.73	14.22	29	25.1	2.63	0.21	20.3
4.50	1	4.29	11.89	30	13.5	4.11	0.37	14.4
	2	4.22	15.05	30	23.8	4.08	0.32	22.7
5.00	1	5.01	11.52	33	12.0	4.79	0.50	12.8
	2	4.91	14.90	31	19.5	4.72	0.39	13.4
5.50	1	5.39	13.89	30	27.2	5.20	0.43	18.1
	2	5.20	15.02	32	26.5	5.01	0.44	18.9

Experiment 1, indicating the same level of discrimination. The mean ratio of each SD in Table 2 to the corresponding SD in Table 1 is 0.94. In Experiment 1 the mean Weber fraction (SD/Reference Frequency) for the 2.25, 2.5 and 2.75 conditions was 0.088 and for the 4.5, 5.0 and 5.5 conditions it was 0.090. In Experiment 2 the corresponding values are 0.082 and 0.081. There is no evidence that the absence of a reference grating has in any way impaired performance.

As in the analysis of variance for Experiment 1, the effect of Range on  $\mu'$  is highly significant ( $F[2,6] = 121.2$ ,  $P < 0.0001$ ). No significant effects for SF condition ( $F[1,3] = 2.85$ , NS), Block ( $F[2,6] = 6.5$ ,  $P = 0.08$ ) or any interactions were observed.

The values of  $\mu'$  are plotted against midpoint (in AU) in Fig. 2(B). The best-fitting lines are given by  $\mu' = 0.84 \text{ midpt} + 0.15$  for the 2.5 cpd condition and  $\mu' = 1.01 \text{ midpt} - 0.08$  for the 5.0 cpd condition.

As in Experiment 1, the values of  $\mu'$  for the second pair of blocks in each session (with midpoints 4.5, 5, 5.5) fall below the midpoints of the ranges of stimuli presented.

Once again, both the Weibull distribution (total  $\chi^2[354] = 295.0$ ,  $P = 0.989$ ) and gaussian distribution (total  $\chi^2[354] = 244.7$ ,  $P = 0.999$ ) fit the data quite well. Although the ratio of the total  $\chi^2$  values is not significant ( $F[354,354] = 1.21$ ,  $P < 0.10$ ), it is in the same direction as in Experiment 1.

These results replicate the main findings of Experiment 1 very closely. An initial stimulus identified as a reference value is not required to support discrimination. The effect of the range of stimuli presented is sufficient to locate the criterion at a suitable internal position. The psychometric functions for the first four trials in each condition show the same pattern as the full data set. For all the data, the mean SD was 0.20 cpd for the 2.5 cpd condition, and 0.41 cpd for the 5.0 cpd condition. The corresponding values for the four-trial data were 0.20 and 0.49. Thus, four-trial data and the full data set show similar pictures.

### EXPERIMENT 3

Two experiments have provided evidence that contextual stimuli, operating through the criterion-setting

processes, can account for the position of the PSE in the spatial frequency discrimination task. We turn to the question of how similar two sets of stimuli must be to allow the indicator traces generated by judgements of one set of stimuli to affect the criterion for judgements of the other set of stimuli. Regan and Beverley (1983) showed that adaptation to a stimulus was specific to both its orientation and spatial frequency. This accords with results from masking and post-adaptation experiments at near-threshold contrast that suggest that low-level encoding is orientation selective (Movshon & Blakemore, 1973; Blakemore & Nachmias, 1971). But Burbeck and Regan (1983), using a 2IFC procedure, found that similar levels of spatial frequency discrimination were obtained whether the two gratings presented on each trial were parallel or orthogonal. They conclude that orientation and spatial frequency are independent dimensions in relation to discrimination.

This observation appears to exclude a model in which, say, vertically oriented spatial frequencies are discriminated by one SF-discrimination mechanism, and horizontal SFs are discriminated by a different module, and the two do not interact. Instead it suggests that a single discrimination mechanism accepts SF information from stimuli of different orientation.

However, Burbeck and Regan's (1983) experiment is subject to an alternative explanation based on CST. In their design each trial presented a pair of stimuli that were chosen to be equidistant in opposite directions from a base frequency for that session. CST suggests that the first such pair would have led to the selection of a criterion intermediate between the two stimuli (corresponding to the base frequency) and the indicator traces from the continuing succession of stimuli would be symmetrically disposed about that location and would tend to maintain the criterion there. The second stimulus on each trial would be judged in relation to that criterion; the first stimulus would simply contribute an indicator trace to the set. When, for example, each trial consisted of a horizontal grating followed by a vertical one, then stimuli of both orientations may have contributed to stabilizing the criterion. This would accord with Burbeck and Regan's conclusion. Alternatively, it is possible that the vertical stimuli alone stabilized the criterion that was used in judging the second (vertical) stimulus in each pair. The random sequence of preceding vertical stimuli would have stabilized the criterion in the same location and given a similar result. On this interpretation, Burbeck and Regan's results for orthogonal stimuli do not establish that frequency discrimination is independent of orientation.

This makes it of interest to re-examine the question using a design capable of detecting a positive effect of the orthogonal stimuli. We can do this by testing two useful CST predictions. First, if there is a common mechanism for discriminating spatial frequencies independent of orientation, and the subject's task is to discriminate vertically oriented gratings but the random sequence of trials also includes horizontal gratings, then the criterion

for the vertical gratings will be determined by both the range of the vertical and the range of the horizontal test stimuli presented. Second, as the effect of each stimulus range is mediated by stabilization, negative stimulus dependencies should occur not only within each set of stimuli but also between them. The next experiment tests these predictions.

In this experiment, vertical and horizontal sine-wave gratings were presented for frequency discrimination in random order, using MSS. Two ranges of vertical stimuli and two ranges of horizontal stimuli were factorially combined. If SF discrimination is independent of orientation we expect the horizontal test range to affect the PSE for judgements of the vertical stimuli, and vice versa.

### Subjects

Six new subjects, three male and three female students, aged between 18 to 22 years, with normal or corrected-to-normal visual acuity, participated in the experiment. They were selected from the subject panel of the Department of Experimental Psychology and were naive as to the aims of the experiment. They attended hourly sessions on four consecutive days, at the same time of day, and were paid.

### Apparatus and stimuli

The apparatus and stimuli were the same as in Experiment 1, except where noted. The stimuli were vertical and horizontal sine-wave gratings which subtended 7.2 deg of visual angle vertically and horizontally (14.4 cm by 14.4 cm), in a 2D cosine envelope with a plateau of 5.8 deg visual angle (11.6 cm by 11.6 cm), presented at the centre of the screen.

The reference grating spatial frequencies were 2.5 (5.0) cpd; the test gratings were centred on 2.25 (4.5) or 2.75 (5.5) cpd.

### Procedure

Each subject attended four sessions on consecutive days. Each session started with a practice block whose results were not analysed. This used stimuli with a spatial frequency centred on 1.0 cpd and of lower contrast than in the experimental blocks. Vertical and horizontal 1.0 cpd reference gratings were presented initially, followed by 22 trials in which vertical and horizontal test stimuli were presented in random order.

This was followed by an experimental block for the 2.5 cpd reference grating, followed by a second block for the 5.0 cpd reference grating. Each block consisted of 20 warm-up trials followed by 12 presentations of each of 11 vertically oriented stimulus values and 12 presentations of each of 11 horizontally oriented stimulus values in randomly intermixed order.

In the 2.5 cpd condition, a vertical reference grating with a spatial frequency of 2.5 cpd was presented initially for 10 sec, followed by a mask presented for 0.5 sec. This was followed immediately by a 2.5 cpd horizontal grating also presented for 10 sec, followed by a 0.5 sec mask.

The reference gratings were always given in this order. After a retention interval of 30 sec whose end was signalled by a short beep, the experimental trials began. Following each response, a 1.5 sec interval intervened during which the uniform blank background was visible. Each test grating was presented for 0.2 sec.

In the first block the range of vertical test gratings was centred on either 2.25 or 2.75 cpd, and the range of horizontal test gratings was centred on either 2.25 or 2.75 cpd. The four possible combinations were presented over the four sessions. For the 5.0 cpd condition (second block) the procedure was similar.

The method of single stimuli was used. The subject's task was to decide whether each test grating had a higher or lower spatial frequency than the corresponding vertical or horizontal reference grating. No feedback was given.

The response to each stimulus together with the response time were measured, but the latter will not be discussed here.

### Results and discussion

Figure 6 shows the results for discrimination of vertical gratings, separately for the two SF conditions, and Fig. 7

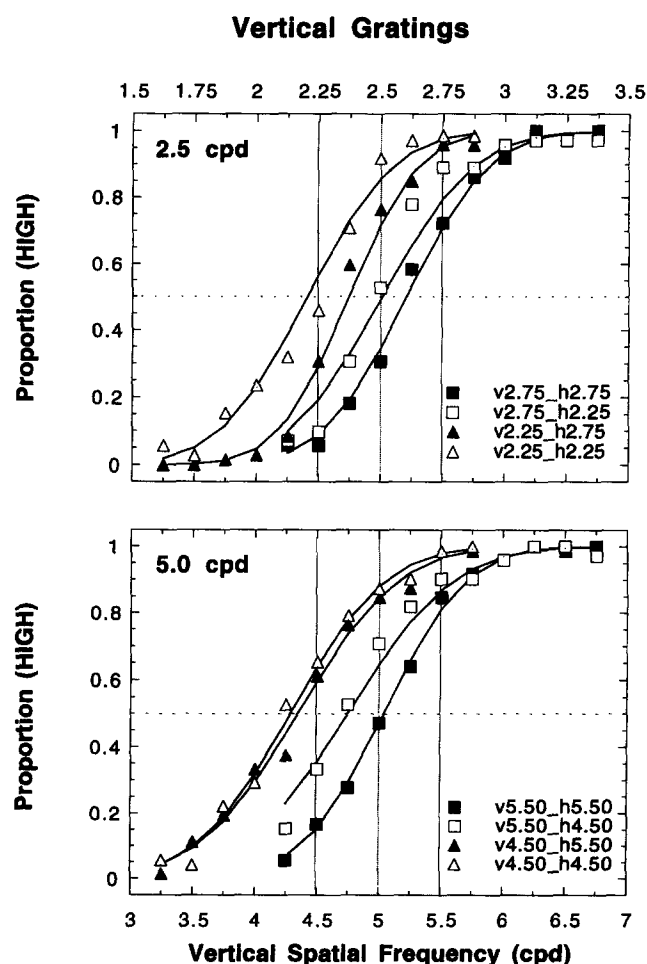


FIGURE 6. Experiment 3: judgements of vertical gratings. Gaussian psychometric functions (upper panel, 2.5 cpd; lower panel, 5.0 cpd) for each combination of horizontal and vertical stimulus ranges. Data are collapsed over six subjects.

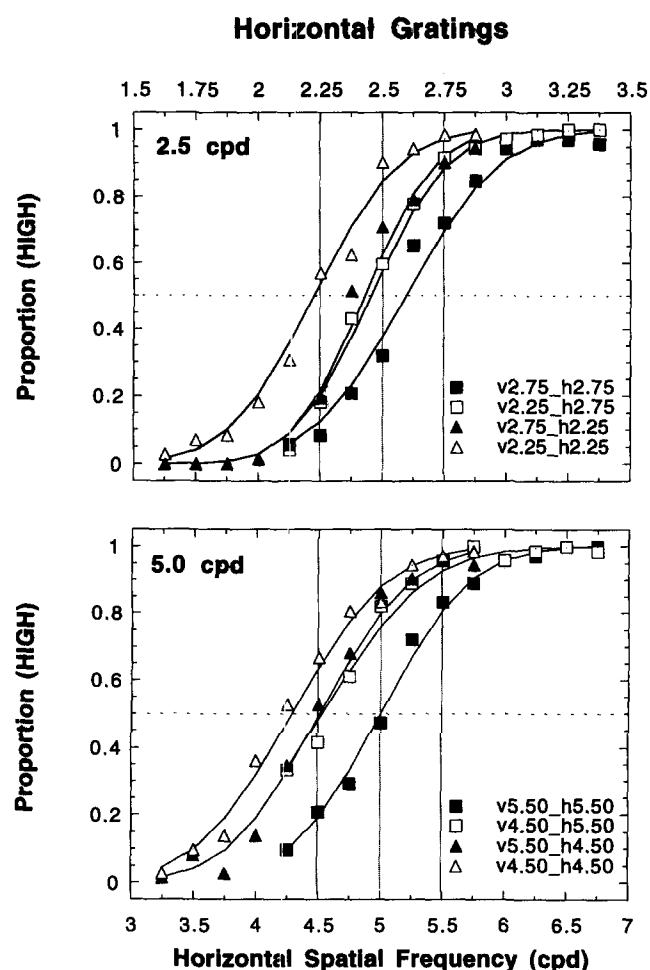


FIGURE 7. Experiment 3: judgements of horizontal gratings, as in Fig. 6.

shows the results for horizontal gratings. In each case the data have been pooled for the six subjects. The curves show gaussian fits for each combination of the range judged and the orthogonal range.

In each case, the difference between filled and empty symbols of the same shape (e.g. filled and empty squares) illustrates the effect of variation in range on the orthogonal dimension, with the range judged held constant. A comparison of squares and triangles, both filled or both empty, illustrates the effect of variation in range on the dimension judged, with the orthogonal range held constant.

For judgements of vertical stimuli in Fig. 6, we note that when the judged range is high (square symbols) the psychometric functions have higher means than when the judged range is low (triangles). This is shown for both SF conditions, and the same effect is seen for judgements of horizontal stimuli in Fig. 7. This is the contextual effect of the test stimulus range that we saw, for stimuli of a single orientation, in Experiments 1 and 2.

If we hold the range judged constant we can examine the effect of the range of orthogonal stimuli that are presented in the same block. A comparison of the filled and empty squares (or triangles) illustrates the effect of variation in the range of the accompanying orthogonal

stimuli. In every case, both the range judged and the orthogonal range tend to shift the psychometric function towards their respective midpoints. Thus, when both ranges are high (filled squares) the curves are highest, when both ranges are low (empty triangles) the curves are lowest, and the mixed cases are intermediate.

Gaussian cdfs were fitted to the data for each subject, session and block, and the values of  $\mu'$  were found. The values obtained for vertical stimuli and for horizontal stimuli were subjected to separate analyses of variance, the factors in each analysis being SF Condition (2.5 or 5.0 cpd), Judged Range (e.g., for judgements of vertical stimuli, the midpoints of the ranges of vertical stimuli), and Orthogonal Range (in this case the midpoints of the accompanying ranges of horizontal stimuli). The analysis for vertical gratings gave a highly significant effect of Judged Range ( $F[1,5] = 37.0$ ,  $P = 0.002$ ), and a significant effect of Orthogonal Range ( $F[1,5] = 9.12$ ,  $P = 0.029$ ). For horizontal gratings both Judged Range ( $F[1,5] = 14.54$ ,  $P = 0.013$ ) and Orthogonal Range ( $F[1,5] = 13.94$ ,  $P = 0.014$ ) were significant. There was also a significant effect of SF Condition ( $F[1,5] = 7.87$ ,  $P = 0.038$ ).

The main finding is that when judging spatial frequency the contextual effects predicted by CST are generated not only by stimuli of the same orientation, but also and to a similar extent by stimuli of the orthogonal orientation. Information from orthogonal test stimuli is not independently processed by separate SF decision mechanisms. It seems that neurophysiological evidence for orientation-selective channels at early stages of processing does not predict the relations between sensory dimensions at the level at which discrimination is determined.

When fitting the individual psychometric functions we found that both the Weibull distribution (total  $\chi^2[709] = 635.3$ ,  $P = 0.975$ ) and the gaussian distribution (total  $\chi^2[709] = 507.2$ ,  $P = 0.999$ ) could be used to fit the data, but when their fits are compared, the ratio of summed  $\chi^2$  values is significant  $F[709,709] = 1.25$  ( $P < 0.05$ ). This confirms the evidence from Experiment 1 that the gaussian distribution gives better fits.

The mean parameter values are given in Table 3. The mean values of the Weber fraction (SD/Reference Frequency) are, for the 2.5 cpd condition, 0.081 (vertical stimuli) and 0.081 (horizontal) and for the 5.0 cpd condition, 0.093 (vertical) and 0.107 (horizontal). This is in accordance with the earlier findings.

In Fig. 8 the mean values of  $\mu'$  are plotted against the ranges judged. A separate line is fitted for each value of the orthogonal range. The vertical displacements on this plot again demonstrate the effect of the orthogonal ranges in shifting  $\mu$ . We may also note that these lines have lower slopes than in the earlier experiments. This is predicted by the criterion-setting model. For example, consider the curve in panel A for discrimination of vertical gratings (ranges 2.25 and 2.75) when accompanied by the horizontal range centred on 2.75 cpd (empty squares). When stimuli in the range of vertical

TABLE 3. Experiment 3: mean parameters and  $\chi^2$  values for Weibull and gaussian psychometric functions

Vertical gratings								
Vert. Range	Horiz. Range	$\alpha$	$\beta$	$\Sigma df$	$\Sigma \chi^2$	$\mu$	$\sigma$ (SD)	$\Sigma \chi^2$
v2.25	h2.25	2.29	14.17	46	38.3	2.20	0.18	18.9
	h2.75	2.46	16.95	45	41.0	2.37	0.18	23.3
v2.75	h2.25	2.61	14.13	45	39.0	2.51	0.22	31.8
	h2.75	2.71	12.48	45	43.9	2.60	0.23	34.1
v4.50	h4.50	4.45	13.62	43	30.4	4.28	0.40	16.2
	h5.50	4.56	11.38	46	35.1	4.35	0.47	20.3
v5.50	h4.50	4.95	9.67	49	29.6	4.72	0.58	14.4
	h5.50	5.24	13.78	47	29.0	5.04	0.42	22.7
Horizontal Gratings								
v2.25	h2.25	2.31	15.12	48	23.6	2.23	0.18	19.3
	h2.75	2.54	16.30	47	20.4	2.45	0.19	21.6
v2.75	h2.25	2.52	14.42	48	49.8	2.43	0.19	31.9
	h2.75	2.71	12.19	42	55.9	2.59	0.24	43.2
v4.50	h4.50	4.49	10.10	37	59.1	4.29	0.46	42.5
	h5.50	4.49	9.16	38	36.4	4.23	0.72	31.1
v5.50	h4.50	4.74	11.58	41	62.5	4.52	0.49	43.8
	h5.50	5.20	12.21	42	41.7	4.99	0.47	35.4

gratings centred on 2.75 are judged, they are accompanied by horizontal stimuli in the same range. Thus, both sets of stimuli should tend to place  $\mu$  at the same location. But when the vertical stimuli centred on 2.25 are judged, they are accompanied by horizontal stimuli in the range centred on 2.75 cpd. The indicator traces from these horizontal stimuli should tend to raise  $\mu$ . This will reduce the difference between the values of  $\mu$  for the two vertical ranges and so flatten the line joining them. Thus stabilization accounts for the flatter slopes in the present experimental design, as compared with the earlier results for single orientations.

The values of  $\mu'$  are generally less than 1.0, especially for the 5.0 cpd condition, indicating a tendency for the psychometric functions to shift downwards. Magnussen and Dyrnes (1994) obtained a similar constant error which they attributed to a difference in the apparent spatial frequency of a stimulus presented for a short time (test stimuli) or a long time (reference stimuli). However, we have found no effect of the reference stimuli, and in Experiment 2 the downward shift is seen in the absence of a reference stimulus. A possible explanation may be that the appropriate scale for the independent variable is a nonlinear function (e.g. reciprocal or logarithmic) of spatial frequency.

### Sequential dependencies

Criterion-setting theory predicts that if an effect of the range of stimuli on  $\mu$  is found, this contextual effect should also manifest as a negative sequential dependency on preceding stimuli. Evidence of a sequential dependency on preceding responses may also be seen.

The sequential dependencies plotted in Fig. 9 relate to those sessions in which the horizontal and vertical ranges had the same midpoint. (This restriction avoids ambiguity in classifying stimuli as high or low.) The data were pooled for judgements of vertical stimuli and horizontal

stimuli and for both SF conditions. They are plotted separately for preceding stimuli with the same orientation as the judged stimulus (upper panel) or for the orthogonal orientation (lower panel). In each case the effects of different preceding stimulus-response combinations are plotted.

As in Experiment 1, the data show a positive dependency on preceding responses and a negative dependency on preceding stimuli, and these effects are similar for both the parallel and orthogonal preceding orientations. For current and preceding stimuli having parallel orientations, the difference in  $\mu'$  values as a function of the preceding response, with stimulus controlled, was 0.124 AU at lag 1, and 0.080 at lag 5. As a function of the preceding stimulus, the differences were 0.076 at lag 1, and 0.070 at lag 5. For orthogonal orientations, the difference in  $\mu'$  values as a function of the preceding response was 0.095 AU at lag 1, and 0.070 at lag 5. As a function of the preceding stimulus, the differences were 0.064 at lag 1, and 0.065 at lag 5. For these data  $\sigma$ /Reference Frequency is approximately 0.09 AU.

A four-way analysis of variance was conducted, the main effects being Preceding Stimulus (high or low), Preceding Response (HIGH or LOW), Preceding Orientation (parallel or orthogonal) and Lag (1-5). The two- and three-way interactions were not significant, so their sums of squares were combined with the four-way interaction to give the error term. The main effects for Preceding Stimulus ( $F[1,32] = 351.0$ ,  $P < 0.0001$ ) and Preceding Response ( $F[1,32] = 550.4$ ,  $P < 0.0001$ ) were highly significant. The effects for Preceding Orientation ( $F[1,32] = 2.61$ , NS) and Lag ( $F[1,32] = 0.78$ , NS) were not significant. Thus, no difference was detected between sequential dependencies on orthogonal stimuli and on stimuli of the same orientation.

### DISCUSSION AND CONCLUSION

Sensory memory theory claims that an earlier stimulus can produce long-term effects on spatial frequency discrimination that are mediated by storage of a neural representation of the initial stimulus in a visual long-term memory. Results (failure of the threshold to increase) interpreted as indicating storage for as long as 50 hr were reported by Magnussen and Dyrnes (1994). In three experiments we have compared predictions from the sensory memory hypothesis with those from an alternative account based on criterion-setting theory, in which performance in procedures such as the MSS is determined by the operation, at the time of testing, of short-term mechanisms that determine the position of the response criterion, with no need to postulate the retention of a representation of the reference stimulus.

These experiments lead to the following conclusions.

First, we have shown that the position of the PSE is determined by the range of stimuli used in testing, even though the reference stimulus is held constant as the test range varies, or is absent. This relation is not compatible with the hypothesis that a sensory memory determines

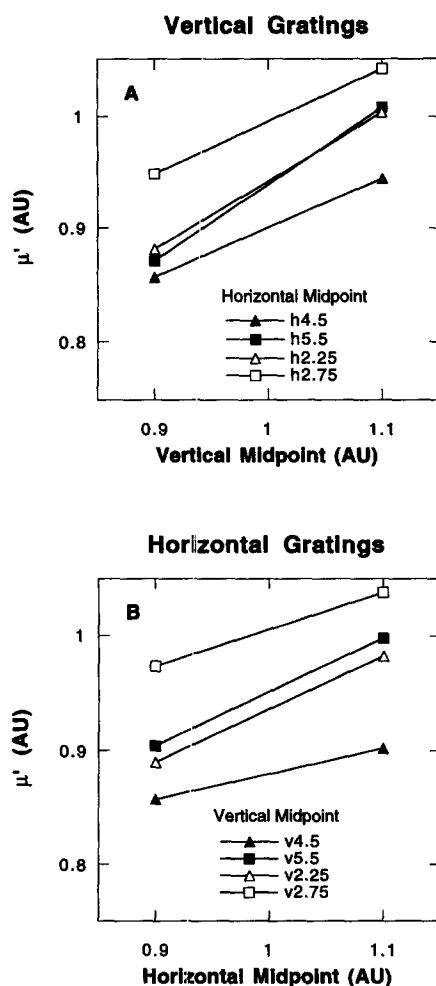


FIGURE 8. Experiment 3. (A) Mean  $\mu'$  for judgements of vertical gratings plotted against vertical range midpoints for the 2.5 (empty symbols) and 5.0 cpd conditions (filled symbols). Separate curves are shown for horizontal range midpoint 2.25 ( $\mu' = 0.61 \text{ midpt} + 0.33$ ); h2.75 ( $\mu' = 0.47 \text{ midpt} + 0.53$ ); h4.5 ( $\mu' = 0.44 \text{ midpt} + 0.46$ ); and h5.5 ( $\mu' = 0.68 \text{ midpt} + 0.25$ ). (B) The corresponding data for horizontal gratings: v2.25 ( $\mu' = 0.46 \text{ midpt} + 0.48$ ); v2.75 ( $\mu' = 0.32 \text{ midpt} + 0.68$ ); v4.5 ( $\mu' = 0.22 \text{ midpt} + 0.66$ —one outlier was excluded); and v5.5 ( $\mu' = 0.47 \text{ midpt} + 0.48$ ).

the location of the PSE. It is a predicted consequence of criterion setting as mediated by the stabilization mechanism. Stabilization predicts corresponding negative sequential stimulus dependencies. Such dependencies were shown.

We conclude that for spatial frequency, in the paradigm we have examined, the PSE of the psychometric function is not determined by a long-term sensory memory. Thus, it is not surprising that the slopes of the psychometric functions in Experiments 1 and 2 are similar, despite the absence of reference stimuli in the second experiment.

We have also shown that recently presented contextual stimuli are not restricted in their effect to the discrimination of stimuli of the same orientation only. The PSE of the psychometric function is modified in a similar way by both orthogonal and parallel contextual stimuli. Correspondingly, sequential dependencies of similar strength

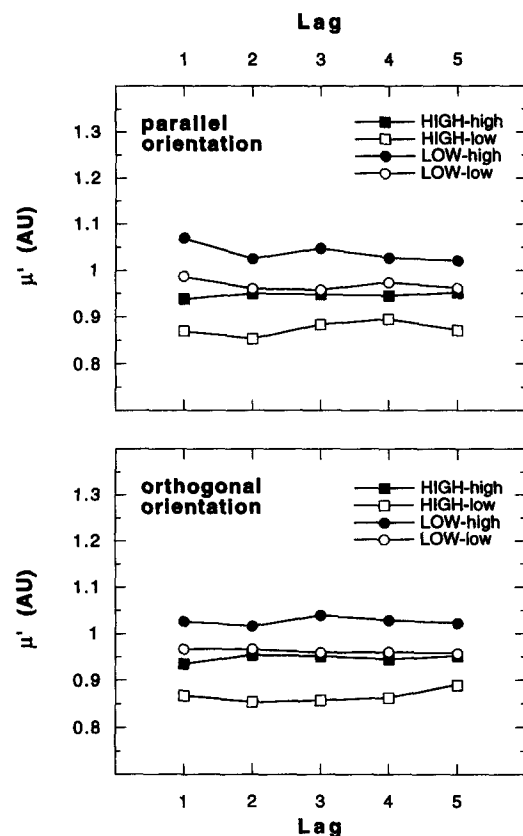


FIGURE 9. Experiment 3: sequential dependencies plotted as a function of four combinations of preceding stimulus and response, as in Fig. 4. The data are plotted separately for preceding stimuli of the same orientation (upper panel) or the orthogonal orientation (lower panel). The parameter  $\mu'$  is plotted for each combination of preceding low or high stimulus and LOW or HIGH response, for five preceding lags. The data come from those sessions in which the horizontal and vertical ranges had the same midpoint and were collapsed over two SF conditions, two ranges, and six subjects.

on both orthogonal and parallel preceding stimuli are seen. This evidence rejects the possibility that separate and independent SF discrimination mechanisms might operate at different orientations, and instead supports a single SF discrimination mechanism that accepts information independent of the orientation of the stimulus. This accords with Burbeck and Regan's (1983) findings. In view of evidence that low level encoding is orientation selective (Movshon & Blakemore, 1973; Blakemore & Nachmias, 1971; Regan & Beverley, 1983), these findings suggest that SF discrimination is located at a later level.

Positive dependencies on preceding responses were demonstrated; these also transferred across stimulus orientation. Analogous results have been obtained in a series of studies on psychophysical scaling by Ward (1982, 1985, 1986, 1990). Giving regularly alternating auditory and visual stimuli, he found reliable positive response-response dependencies both within and between modalities, and a reliable negative stimulus-response dependency within modalities, with occasional weak evidence for an intermodal effect (Ward, 1986). Thus, it appears that the similarity between stimuli may

determine the extent to which negative stimulus dependencies occur between them, but positive response dependencies are not restricted in this way.

We should note that our conclusions apply to the procedure we have studied, and to time intervals equal to or greater than those used here. The four-trial results for Experiment 1 indicate that even at the beginning of the procedure the results accord with CST. However, it is possible that a different outcome might be obtained if our analysis were applied to retention intervals less than 30 sec, and this remains to be investigated.

Both sensory memory theory and CST rely on the concept of memory, but they differ in what they envisage is retained in memory, a representation of the stimulus, or specifications for the response criterion. Sensory memory theory implicitly assumes that retention is limited by unavoidable damage to stimulus traces, by decay or interference. CST assumes that the value of the stabilization decay parameter is set to optimize performance. Thus, stabilization may extend over a few trials or over many. Decay of indicator traces might be a function of elapsed time or trials. The present findings suggest that it may be appropriate to re-examine evidence for sensory memory in the literature, to determine whether it can be accounted for by CST.

A number of studies of spatial frequency discrimination have claimed to find perfect retention of sensory memories (Regan, 1985; Magnussen *et al.*, 1990; Magnussen, Greenlee, Asplund, & Dyrnes, 1991; Magnussen & Dyrnes, 1994). Regan (1985) used two-interval forced choice and the method of constant stimuli, with a range of inter-stimulus intervals, and found no significant deterioration in the discrimination threshold as ISI increased from 0.4 to 20 sec. Magnussen *et al.* (1990) reported "perfect" retention in visual short-term memory for up to 30 sec (the longest ISI they tested) and Magnussen *et al.* (1991) for 10 sec. These observations relate to intervals shorter than those we have studied. Magnussen and Dyrnes (1994) increased the range over which "perfect" sensory memory extends to 50 hr. On the other hand, Harvey, Tran, & Raney (1996) found that memory for spatial frequency decays within 16 sec.

A difficulty with the logic of claiming perfect retention on the basis that the discrimination threshold does not change as time interval increases is that this argument rests on affirming the null hypothesis, which is a doubtful procedure. In Regan's data the threshold was in fact higher at 20 sec than at 0.4 sec for each subject. Regan used a two-tailed *t*-test to reject significance for each subject separately. If, however, we assume that it would not have been reasonable to expect performance to improve as the ISI lengthened, a one-tailed *t*-test is justifiable, and if the probabilities for each subject are combined (Fisher, 1954) the rise in threshold is significant ( $\chi^2 [4] = 11.41, P < 0.025$ ).

Our findings suggest that Magnussen and Dyrnes' results can be attributed to the effect of the reference stimulus in their experiment being negligible, rendering the delay interval irrelevant. In 2IFC discrimination

experiments, CST predicts that whether thresholds increase as ISI increases depends on the rate at which stabilization indicator traces decay. If this decay rate is low relative to the ISIs, little if any increase in threshold will be seen. If the decay rate is higher, then after a sequence of long ISIs, fewer indicator traces may be available to determine the criterion than after short ISIs. Smaller samples of indicator traces will determine a criterion with greater variance, giving higher discrimination thresholds for longer ISIs.

On this account, whether discrimination worsens with delay, and the rate at which it does so, may depend on differences between subjects and dimensions in the stabilization decay parameter. Reasons for variation in the value of  $\delta_s$  may include interindividual variability, and differences in the stability of different dimensions in perceptual experience. For example, the spatial frequencies characterizing a given surface usually remain the same during a period of observation at a constant distance, while the illumination under which it is viewed may vary. This might favor longer retention of indicator traces for frequency than for visual intensity or related measures.

Magnussen, Greenlee, & Thomas (1996) examined sensory memory for both spatial frequency and contrast for ISIs of 1–10 sec, using the same subjects and 2IFC procedures. Over these time intervals, the contrast discrimination threshold increased with ISI. For a single (jittered) reference frequency value the SF threshold was independent of ISI over this range. However, when five reference frequency values were interleaved, the frequency discrimination threshold was higher and increased with ISI. When stimuli varying in contrast and frequency were interleaved, the results for frequency discrimination were similar whether one or five reference contrasts were interleaved, and vice versa.

An interesting experiment by Lee and Harris (1996) also investigates contrast discrimination. They used 2IFC trials with ISIs of 1, 3, or 10 sec, and employed the method of constant stimuli. The reference value was jittered, that is, three values were randomly intermixed in the course of a run. Thus, in one condition the reference stimulus was randomly 9, 15 or 21% contrast, with the test stimulus on each trial sampled from a stimulus range centred on the reference stimulus on that trial. A separate psychometric function was determined for each reference value. Lee and Harris (1996) argued that if subjects construct a representation of the central value of the overall range of stimuli, and make judgements in relation to this value, the three psychometric functions should coincide. If, however, discrimination on each trial is determined by a sensory memory of the first stimulus presented on that trial, the psychometric functions for the three reference values would be correspondingly displaced from one another. They obtained the last result, and conclude in favour of sensory memory.

The hypothesis that subjects employ a representation of the central value of a range relates to the proposal, originally put forward by Helson (1947), that contextual

stimuli may determine an adaptation level or point of reference in relation to which judgements are made. However, adaptation level is defined by a formula as a single fixed value that holds throughout a session. This differs from the CST model in which the stabilization mechanism serves the purpose of optimizing performance, is based on the assumptions of SDT, and produces dynamic trial-by-trial adjustments that cause the criterion to vary about the centre of the range, and so account for sequential dependencies.

Lee and Harris' argument is cogent in relation to the alternatives they considered, sensory memory and a fixed reference level. However, CST offers a different alternative to sensory memory theory, which suggests that a re-evaluation of the experiment in terms of the effects of criterion setting may be useful. Consider trial  $t$ , on which two stimuli are presented, a reference stimulus  $S_r$  and a test stimulus from its associated range. The stimuli presented on previous trials are a random sample from the total range of stimuli in the design. Thus, the indicator traces laid down by those earlier stimuli and that have not yet decayed away will tend to place the criterion at the centre of the total range of stimuli (the three reference stimuli and their associated test ranges). On trial  $t$  the first stimulus,  $S_r$ , adds a further indicator trace that will contribute to determining the value of the criterion against which the second stimulus on that trial will be judged. If  $S_r$  is the lowest of the three reference stimuli (say 9%), its indicator trace will reduce the criterion below its mean value; if it is the highest of the three (21%) it will raise the criterion. Accordingly, the psychometric function for all trials on which the 9% reference stimulus is presented will be displaced downwards; for the 21% stimulus, upwards. On the CST interpretation, these displacements are an example of a sequential stimulus dependency, and are wholly analogous to the stimulus dependencies illustrated in Figs 4 and 9 (at lag 1), which show the displacements in the PSE consequent on the preceding stimulus being high or low, against a random background of preceding stimuli. (A similar argument applies to trials on which  $S_r$  is second.) Thus, there is no need to consider these data as evidence for sensory memory theory.

A second interesting finding is that when the data for the three values of a jittered reference stimulus were combined and a single psychometric function fitted to them, its mean fell at the centre of the stimulus range equally closely whether the interstimulus interval was 1, 3 or 10 sec (see their Fig. 3). If we had the evidence of this observation alone, it might be thought that sensory memory was perfect up to 10 sec. However, the authors also noted that the discrimination threshold increased with interstimulus interval, and so conclude that memory for contrast decays between 1 and 10 sec.

Both observations are consistent with CST. If indicator traces for contrast decay with time, it is likely that more indicator traces determine the criterion for the second stimulus on each trial when the ISI is 1 sec than when it is 10 sec, giving less variation in the criterion. Whether

random samples are large or small, the mean of the sampling distribution will be the same, so the PSE should not differ as a function of ISI.

It is of interest that in the experiments we report here, gaussian distribution functions provide better fits to spatial frequency discrimination functions than do Weibull functions. The reason for this may be that the Weibull distribution is asymmetrical on the linear spatial frequency scale used here.

We have also found that the criterion for spatial frequency discrimination is affected similarly by exposure to spatial frequencies at different orientations. This is an appropriate mode of organizing discrimination for organisms that are regularly exposed to environmental stimuli whose orientations vary. It allows the organism to make use of all environmental information that is relevant to determining the optimal location for the criterion.

A general implication of the CST account is that findings which show an effect of an earlier stimulus on the judgement of a later one should not automatically be taken to be evidence for a sensory memory of the first stimulus. The possibility that the effect is mediated by the mechanisms of criterion setting should be considered.

## REFERENCES

- Blakemore, C. & Nachmias, J. (1971). The orientation specificity of two visual after-effects. *Journal of Physiology*, 213, 157–174.
- Burbeck, C. A. & Regan, D. (1983). Independence of orientation and size in spatial discriminations. *Journal of the Optical Society of America*, 73, 1691–1694.
- Costin, D. (1993) KeMo Reaction Time Utilities. Available from the author; email: dcostin@ucsd.edu.
- Fisher, R. A. (1954). *Statistical methods for research workers*, 12th ed. London: Oliver and Boyd.
- Green, D. M. & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Green, D. M. & Swets, J. A. (1974). *Signal detection theory and psychophysics*. New York: Wiley. Reprinted with corrections. Huntington, NY: Krieger.
- Greenlee, M. W., Georgeson, M. A., Magnussen, S. & Harris, J. P. (1991). The time course of adaptation to spatial contrast. *Vision Research*, 31, 223–236.
- Greenlee, M. W. & Thomas, J. P. (1993). Simultaneous discrimination of the spatial frequency and contrast of periodic stimuli. *Journal of the Optical Society of America A*, 10, 395–404.
- Guilford, J. P. (1936). *Psychometric methods*. New York: McGraw-Hill.
- Harvey, L. O. Jr., Tran, V. X. & Raney, S. M. (1996). Memory channels and memory for spatial frequency. *Investigative Ophthalmology and Visual Science (Suppl.)*, 37, 697.
- Helson, H. (1947). Adaptation-level as a frame of reference for prediction of psychophysical data. *American Journal of Psychology*, 60, 1–29.
- Lee, B. & Harris, J. (1996). Contrast transfer characteristics of visual short-term memory. *Vision Research*, 36, 2159–2166.
- Magnussen, S. & Dyrnes, S. (1994). High-fidelity perceptual long-term memory. *Psychological Science*, 5, 99–102.
- Magnussen, S., Greenlee, M. W., Asplund, R. & Dyrnes, S. (1990). Perfect visual short-term memory for periodic patterns. *European Journal of Cognitive Psychology*, 2, 345–362.
- Magnussen, S., Greenlee, M. W., Asplund, R. & Dyrnes, S. (1991). Stimulus-specific mechanisms of visual short-term memory. *Vision Research*, 31, 1213–1219.
- Magnussen, S., Greenlee, M. W. & Thomas, J. P. (1996). Parallel

- processing in visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 202–212.
- McCourt, M.E. & Blakeslee, B. (1994). Contrast-matching analysis of grating induction and suprathreshold contrast perception. *Journal of the Optical Society of America A*, 11, 14–24.
- Morgan, M. J. (1992). On the scaling of size judgements by orientational cues. *Vision Research*, 32, 1433–1445.
- Movshon, J. A. & Blakemore, C. (1973). Orientation specificity and spatial selectivity in human vision. *Perception*, 2, 53–60.
- Pelli, D. G. (1995). VideoToolbox. Available from the author: Web site: rajskey.psych.nyu.edu/VideoToolbox/Download.html.
- Pelli, D. G. & Zhang, L. (1991). Accurate control of contrast on microcomputer displays. *Vision Research*, 31, 1337–1350.
- Quick, R.F. (1974). A vector magnitude model of contrast detection. *Kybernetik*, 16, 65–67.
- Regan, D. (1985). Storage of spatial-frequency information and spatial-frequency discrimination. *Journal of the Optical Society of America A*, 2, 619–621.
- Regan, D. & Beverley, K. I. (1983). Spatial-frequency discrimination and detection: comparison of postadaptation thresholds. *Journal of the Optical Society of America*, 73, 1684–1690.
- Treisman, M. (1963). Temporal discrimination and the indifference interval: implications for a model of the “internal clock”. *Psychological Monographs*, 77, Whole No. 576, pp. 1–40.
- Treisman, M. (1984). A theory of criterion setting: an alternative to the attention band and response ratio hypotheses in magnitude estimation and cross-modality matching. *Journal of Experimental Psychology: General*, 113, 443–463.
- Treisman, M. (1985). The magical number seven and some other features of category scaling: properties of a model for absolute judgment. *Journal of Mathematical Psychology*, 29, 175–230.
- Treisman, M. (1987). Effects of the setting and adjustment of decision criteria on psychophysical performance. In Roskam, E. E. & Suck, R. (Eds), *Progress in mathematical psychology I* (pp. 253–297). Amsterdam: Elsevier.
- Treisman, M. & Faulkner, A. (1984a) The setting and maintenance of criteria representing levels of confidence. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 119–139.
- Treisman, M. & Faulkner, A. (1984b) The effect of signal probability on the slope of the receiver operating characteristic given by the rating procedure. *British Journal of Mathematical and Statistical Psychology*, 37, 199–215.
- Treisman, M. & Faulkner, A. (1985). Can decision criteria interchange locations? Some positive evidence. *Journal of Experimental Psychology: Human Perception and Performance*, 11, 187–208.
- Treisman, M. & Faulkner, A. (1987). Generation of random sequences by human subjects: Cognitive operations or psychophysical process? *Journal of Experimental Psychology: General*, 116, 337–355.
- Treisman, M. & Williams, T. C. (1984). A theory of criterion setting with an application to sequential dependencies. *Psychological Review*, 91, 68–111.
- Ward, L. M. (1982). Mixed-modality psychophysical scaling: sequential effects and other properties. *Perception and Psychophysics*, 31, 53–62.
- Ward, L. M. (1985). Mixed-modality psychophysical scaling: inter- and intra-modality sequential dependencies as a function of lag. *Perception and Psychophysics*, 38, 512–522.
- Ward, L. M. (1986). Mixed-modality scaling: double cross-modality matching for “difficult” continua. *Perception and Psychophysics*, 39, 407–417.
- Ward, L. M. (1990). Mixed-method mixed-modality psychophysical scaling. *Perception & Psychophysics*, 48, 571–582.
- Watson, A. B. (1979). Probability summation over time. *Vision Research*, 19, 515–522.
- Westheimer, G. & McKee, S.P. (1977). Spatial configurations for visual hyperacuity. *Vision Research*, 17, 941–947.
- Wever, E. G. & Zener, K. E. (1928). The method of absolute judgment in psychophysics. *Psychological Review*, 35, 466–493.
- Winer, B. J., Brown, D. R. & Michels, K. M. (1991). *Statistical principles in experimental design* (3rd edn). New York: McGraw-Hill.

---

*Acknowledgements*—The experiments above were carried out by M. Lages in partial fulfilment of the requirements for the D. Phil. degree, under the supervision of M. Treisman. This work was supported by the EPSRC and BBSRC, respectively. We wish to thank the reviewers for useful comments.