

# Decision strategies and visual-field asymmetries in *same-different* judgments of word meaning

MARGARET A. FRANCIS and R. JOHN IRWIN  
*University of Auckland, Auckland, New Zealand*

The accuracy with which observers judged whether two words belonged to the same semantic category was determined from a detection-theoretic analysis of *same-different* judgments. In Experiment 1, one word was presented centrally and the other word in either the left visual field (LVF) or the right visual field (RVF); in Experiment 2, both words were presented to either the LVF or the RVF. In order to obtain receiver-operating characteristics (ROCs) of performance, observers were asked to rate their confidence that the two words belonged to the same semantic category. Two models of the decision strategy were fitted to the obtained characteristics: a differencing model, in which the decision variable was the difference between the two observations; and an optimal model, in which each observation was judged in relation to a criterion. In both experiments, the optimal model provided a better fit than the differencing model to the obtained characteristics. Maximum-likelihood estimates of both the criterion-free parameter,  $d'$ , and the area under the operating characteristic,  $p(A)$ , were greater for words presented in the RVF than for those presented in the LVF.

The *same-different* task has enjoyed wide usage in investigations of cognitive functions, including investigations of hemispheric asymmetries, perceptual matching, and the processes involved in stimulus comparisons. The predominant focus in cognitive psychology has been on reaction times rather than accuracy. Typically, errors in *same-different* studies are of secondary interest, since the conditions are usually such that few errors occur (Farell, 1985), and most investigators attempt to minimize them.

In contrast to this approach, we report here on the accuracy of *same-different* judgments rather than on reaction time. We present findings concerning the accuracy with which people can judge whether two words belong to the same semantic category or not, and also concerning the question of whether the accuracy of their judgments depends on the visual field in which the words are displayed. By so doing, we hope to provide a different perspective on the processes involved in *same-different* judgments of this kind. Thus, we propose that an appropriate measure of accuracy can reveal the underlying decision processes in judgments of identity or difference, and that these processes are fundamental to understanding such judgments. One way in which this can be accomplished is by analyzing the *same-different* receiver-operating characteristic (ROC), which provides a

measure of accuracy, because its shape and location depend upon the decision strategy that an observer adopts in the task (Irwin, Stillman, Hautus, & Huddleston, 1993; Macmillan & Creelman, 1991; Noreen, 1981).

The purpose of the work reported here was to illustrate the potential usefulness of this psychophysical approach to a specific problem in cognition—namely, the problem of determining the decision strategy adopted by an observer in judging whether two things are the same or different (see Algom, 1992, for other psychophysical approaches to cognition). First, however, we briefly review some possible indices of accuracy for the *same-different* task and attempt to justify why we have chosen to present an index based on the ROC.

## MEASURES OF ACCURACY FOR THE *SAME-DIFFERENT* TASK

The most natural measure of accuracy for the *same-different* task is proportion correct but, for several reasons, we believe this is rarely a satisfactory measure. The *same-different* task, like the simpler yes-no task, is prone to response biases, such that the proportion of correct judgments that a particular observer achieves will depend upon that observer's bias. Furthermore, a given proportion correct attained even by an unbiased observer has a different meaning in the *same-different* task from its meaning in other tasks, such as forced-choice tasks, in the sense that it represents a different level of performance (Macmillan & Creelman, 1991).

An alternative index is the detection-theory index,  $d'$ , which possesses the desirable property that it represents a uniform level of performance for any task. As is well known,  $d'$  is the distance between the means of the two underlying Gaussian processes associated with each

We thank Michael Corballis, Carole Ernest, and Michael Hautus for helpful comments on an earlier draft of this paper, and Michael Corballis for advice on Experiment 2. We also thank the editor and three anonymous reviewers for their useful and challenging comments. Experiment 1 is included in part of a doctoral dissertation by M.A.F. Requests for reprints should be sent to R. J. Irwin, Department of Psychology, The University of Auckland, Private Bag 92019, Auckland, New Zealand (e-mail: rji@auckland.ac.nz).

event. Its unit is the standard deviation (*SD*) of those processes, which is assumed to be equal for each event. However, even the use of this index can be problematical, for a number of reasons. First, it is not appropriate to treat *same-different* data as though they came from a yes-no experiment, and thus to compute  $d'$  from the standard yes-no tables. For one thing, the *same-different* task contains two observation intervals (or two observation places), whereas the yes-no task contains only one. Even so, a single point in ROC space—whether from a single-interval or a *same-different* task—is not sufficient to provide an unequivocal measure of accuracy, because the path of the ROC cannot be determined unambiguously from a single point (see Swets & Pickett, 1982, p. 29).

A more subtle problem with using the detection-theory index arises from the fact that there are at least two decision variables that can be used to make a *same-different* judgment, and these decision variables give rise to different outcomes and therefore to different estimations of  $d'$ . A major purpose of our work was to determine which decision variable (i.e., which decision strategy) the observers adopted in deciding whether two words belonged to the same or different semantic categories. We attempted to do this by examining the shape of the ROC for the task. A subsidiary purpose was to illustrate the benefits of this analysis by applying it to a standard problem in cognition—that of determining hemispheric asymmetries in the processing of semantic information. In the following section, we review the psychophysics of the *same-different* task, and then consider its application to hemispheric processing of semantic information.

### PSYCHOPHYSICAL ANALYSES OF THE *SAME-DIFFERENT* TASK

In early psychophysical analyses of the *same-different* task (Macmillan, Kaplan, & Creelman, 1977; Sorkin, 1962), it was supposed that the decision variable was the difference between the experiences produced by each observation. If that difference exceeded a criterion value, the two stimuli were judged as being different; otherwise, they were judged as being the same. This has been called a *differencing* strategy. An alternative strategy, however, uses a fixed criterion value, against which each of the two experiences is compared. Only if the two experiences fall on different sides of this criterion are they judged as being different; otherwise, they are judged to be the same. It can be shown that this is in fact an *optimal* strategy (Johnson, 1980; Macmillan & Creelman, 1991; Noreen, 1981).

For readers who are not familiar with this analysis of the *same-different* task, a simple example similar to one presented elsewhere (Hautus, Irwin, & Sutherland, 1994; Noreen, 1981) may clarify the distinction between the two models. Suppose that an observer has to decide whether or not two apples are of the same variety, but that the only information available about the apples is

their weight. Further suppose that the mean weight of one variety is 100 g, while that of the other variety is 90 g, and that the weight of each variety is distributed normally with an *SD* of 10 g (thus,  $d' = 1$ ). Consider a trial in which the first observation was a weight of 80 g and the second was a weight of 110 g. An observer who adopted the differencing strategy would compute the difference in weight of the two observations (i.e., 30 g), and if that difference exceeded the criterion difference (of, say, 20 g), would decide that the two observations stemmed from different varieties of apples. An alternative strategy would be to set an absolute criterion (e.g., 90 g) for the decision. In the example just considered, the two observations fall on different sides of that criterion, and so this strategy, like the differencing strategy, would also lead to a decision that the two samples came from different varieties. However, suppose that one observation was 100 g and the other was 130 g. In this case, although the differencing strategy would still lead to the decision that the samples came from different varieties, the optimal strategy would lead to the decision that they came from the same variety.

This example, however, oversimplifies the analysis—it may give the impression that the criterion is a point on the decision axis, just as it is in the fundamental yes-no task, but this is not so. The standard analysis of the yes-no experiment assumes that the events being judged give rise to Gaussian distributions of equal variance, and even though we make the same assumptions for the *same-different* experiment, the analysis leads to a very different result from that of the more familiar yes-no, or single-interval, case. Thus, in contrast to the unidimensional decision axis of the single-interval experiment, the *same-different* experiment entails a two-dimensional decision space, which arises because each observation of a trial is represented by an independent random variable, and the outcome of a trial is represented by a point on this space. An experiment comprising many trials gives rise not to two univariate normal distributions, as in the standard case, but to four bivariate normal distributions, one for each of the four possible kinds of trials (see Appendix). Rather than being a point on a unidimensional decision axis, as in the single-interval task, a criterion on this space is, then, a boundary that divides the decision space into regions for accepting that the two observations stemmed from the same or from different events. Different decision strategies divide the space up in different ways and can therefore lead, as our example shows, to different decisions based on the same evidence.

The key to our method for inferring which decision strategy is adopted in an experiment is provided by the shape of the resulting ROC, since it can be shown that the differencing strategy gives rise to an ROC that is asymmetrical about the negative diagonal of the ROC square, whereas the optimal strategy gives rise to an ROC that is symmetrical about that diagonal. In the appendix we present a proof, which is not to our knowledge available elsewhere, that the optimal strategy gives

rise to a symmetrical ROC. The properties of the differencing strategy are better known—the asymmetrical nature of its ROC was first demonstrated by Sorokin (1962), and it has subsequently been delineated by Irwin et al. (1993) and Macmillan et al. (1977)—and we therefore do not derive its shape again here.

### OTHER VIRTUES OF ROC ANALYSIS

It is obviously desirable to reduce or eliminate the response biases to which *same-different* experiments are especially prone (Macmillan & Creelman, 1991). One way of accomplishing this, at least for measures of accuracy, is to obtain a full ROC that yields measures uncontaminated by response biases. The rating method of detection theory is an efficient means of obtaining a full ROC (McNicol, 1972). When applied to the *same-different* task, the rating method requires observers to rate on a multipoint scale their confidence that two stimuli were either the same or different. An observer's performance can then be determined from analysis of the rating ROC. This gives a more complete picture of performance than can be obtained from a single point in the ROC space (as from a yes-no task), and in addition, the shape of the *same-different* ROC allows an inference to be made, as we have stated, about the decision strategy that the observer adopted.

### EXPERIMENTAL EVIDENCE ON SAME-DIFFERENT DECISION STRATEGIES

A few studies have attempted to decide between these decision strategies in simple psychophysical tasks. Vogels and Orban (1986) concluded that the differencing strategy was adopted in judging whether two lines were the same or different in orientation because the results then agreed with those from a forced-choice experiment in which the same observers judged the same stimuli. This somewhat indirect method for inferring which decision strategy was adopted is based upon the following reasoning: As we have noted, a proper measure of discrimination ought to be independent of the particular task that measures it, and by assuming that observers adopted a differencing strategy, Vogels and Orban obtained a similar result from both tasks (i.e., the *same-different* task and the forced-choice task). Had they assumed that the observers adopted an optimal strategy, they would have found that the two tasks yielded different estimates of an observer's ability to judge the orientation of lines—an ability that should not depend upon the task used to measure it. Similarly, Irwin et al. (1993) concluded from the asymmetrical shape of the ROC that judgments as to whether two drinks were the same or different in concentration were based on a differencing strategy. In another study of simple stimuli, Hautus et al. (1994) reported asymmetrical ROCs for judgments of whether two sounds were the same or different in amplitude. They, too, therefore concluded that the differ-

encing strategy was adopted in deciding whether two things were the same or different.

### VISUAL-FIELD ASYMMETRIES IN SAME-DIFFERENT JUDGMENTS OF WORD MEANING

We have applied this psychophysical analysis of the *same-different* task to the familiar problem of determining visual-field asymmetries in processing verbal material. We therefore briefly review some earlier findings. The stimuli used in the majority of previous verbal *same-different* studies have required physical comparisons not comparisons of meanings, of either characters (e.g., Bagnara, Boles, Simion, & Umiltà, 1983; Egeth & Epstein, 1972) or words (e.g., Barron & Henderson, 1977; Barron & Pittenger, 1974; Gibson, Dimond, & Gazzaniga, 1972). These judgments can be made on the basis of a few physical features. On the other hand, when semantic comparisons involving words are called for, the comparisons of meanings must occur at an abstract level. In one such study, Gross (1972) had observers judge whether two 3-letter words belonged to the same or different semantic categories. She found that observers responded faster when the words were in the right visual field (RVF) than when they were in the left visual field (LVF). Later, Urcuioli, Klein, and Day (1981) reported both reaction times and error rates in a similar task. Urcuioli et al. presented two 3-letter words, one centrally and the other peripherally, and measured the time observers took to decide whether the words belonged to the same semantic category. Like Gross, they found that observers responded more quickly when the peripheral word was in the RVF than they did when it was in the LVF. Because they reported error rates (false positives and false negatives) for the *same-different* task, accuracy and response bias can be computed from their results. From their Table 2, the hit rate (the complement of their false negatives) in the matching task was .929, and the false-alarm rate (their false negatives) was .050 for the LVF (averaged over both groups). The corresponding rates for the RVF were .953 and .075. These values yield only one point in ROC space, so that it is not possible to infer what decision strategy might have given rise to that point. We therefore computed the accuracy measure,  $d'$ , and the criterion measure,  $\ln(\beta)$ , as though these points had come from a yes-no experiment. The accuracy in each visual field was identical ( $d' = 3.11$ ). However, the criterion was more lax for the RVF [ $\ln(\beta) = -0.37$ ] than it was for the LVF [ $\ln(\beta) = 0.28$ ]; thus, the more lax criterion adopted when responding to the RVF may partly account for the faster responses to words in that visual field.

### SEMANTIC CATEGORIZATION OF WORDS

The present experiments required more than simple judgments of stimulus magnitude. The *same-different* judgments in these experiments had to be semantically

based; that is, observers were required to judge whether the members of pairs of words belonged to the same or different semantic categories—a comparison that cannot be made visually. Of interest here was the question of whether the differencing strategy would be adopted for this problem, too. Also, we wished to examine whether visual-field asymmetries observed in previous *same-different* studies would be observed when the stimuli required semantic comparisons and when a criterion-free measure of accuracy was available. On the basis of the usual findings from reaction time experiments (Bradshaw, 1989), a distinct advantage for the left (language-dominant) hemisphere was anticipated for these semantic judgments.

### EXPERIMENT 1

In order to obtain ROCs for *same-different* judgments of word meaning, we made the task more difficult than in the usual reaction time experiment, in which few errors occur. We accomplished this by presenting the stimuli for brief durations to parafoveal vision. We assessed hemispheric asymmetries by presenting one member of each pair of words to either the LVF or the RVF. In addition, we examined the possibility, suggested by Chiarello (1988), that visual-field asymmetries may depend upon word length.

#### Method

**Observers.** Seven women and three men volunteered to participate in the experiment. They were all students of psychology, and all described themselves as right-handed.

**Apparatus.** IBM-compatible AT-class computers equipped with standard 14-in. color VGA displays were used. Observers were seated with their chin upon a chinrest 40 cm from the computer screen. All timing (accurate to 1 msec) was controlled by the computer.

**Stimuli.** The stimuli consisted of two categories of words—artificial and natural. Artificial words were defined as words corresponding to any object or structure, whether constructed or manufactured, and regardless of size (e.g., *gun*). Natural words were defined as those corresponding to any organism, or part of an organism, whether plant or animal (e.g., *whale*). Three different word lengths were presented from each category—namely, three-letter words, four-letter words, and five-letter words. There were 56 words of each length in each category, making 336 words in all. The natural and artificial words in the experimental session were selected in order of frequency from Thorndike and Lorge's (1944) list. In addition, several words corresponding to animals and trees that are familiar to the average New Zealander were used (e.g., *kiwi*). All words were capitalized and displayed in the default font (8 × 8 pixels) of Turbo Pascal (Borland, 1990), using medium-resolution graphics (640 × 350 pixels).

**Procedure.** The observers fixated upon a central cross (+), which was then replaced by two words—a central word and a side word. The onset of the two words was simultaneous, but the central word was displayed for 150 msec, and the side word for 30 msec. The side word was located so that its inner edge was aligned 3° to either the left or right of center. Following the presentation of the words, the observer pressed either the "B" key to categorize the central word as natural or the "N" key to categorize it as artificial. The observers were informed that both reaction time and accuracy were important for this part of the task.<sup>1</sup> On any trial, each of the following four possible combinations had an

equal probability of occurring: (1) both words were artificial; (2) both words were natural; (3) the central word was natural and the side word artificial; and (4) the central word was artificial and the side word was natural. After a response had been made to the central word, a six-point rating scale appeared on the screen, and the observers, who were provided with written instructions on how to use this scale, were prompted to rate their confidence that the side word belonged to the same category as the central word. Category 1 corresponded to "very confident not the same," and Category 6 to "very confident the same." Responses in this part of the task were not timed. Every experimental session began with six warm-up trials in which the words were drawn from a separate pool of stimuli. To initiate a trial, the observer pressed a key. The experimental trials were equally divided in terms of whether the words were presented in the LVF or the RVF, whether they were composed of words of the same or of different categories, and whether they consisted of three-, four-, or five-letter words. On any trial, the two words were always of the same length.

Before commencing an experimental session, all observers completed two blocks of 36 practice trials to familiarize themselves with the rating scale. The task was the same in the practice session as in the experimental session, except that six-letter words were presented, and feedback on rating responses was given after each trial and after each block of trials. The block feedback showed the observers their usage of the rating scale and their accuracy. The observers were informed of their progress at the end of each block, and where necessary (e.g., when only a few points on the scale had been used), the instructions for using the rating scale were repeated.

#### Results

The reaction times to the central word are not analyzed here, as they are not germane to our main purpose, which is concerned with *same-different* judgments.<sup>2</sup> In order to carry out an ROC analysis of the accuracy of the *same-different* responses, we defined hits as those trials on which an observer judged the two words as belonging to the same category when in fact they did, and false alarms as those trials on which an observer judged the two words as belonging to the same category when in fact they did *not*. An ROC could then be constructed from the ratings by plotting the hit rate against the false-alarm rate in the usual way (McNicol, 1972). In the analyses that follow, all the trials are included, regardless of whether the response to the central word was correct (as it was on 85% of the trials).

**Individual observers.** Each observer's ratings were analyzed separately. The maximum-likelihood values of the parameter,  $d'$ , were estimated for each model for each observer. Fitting the optimal model to data involves intensive numerical computation, but an excellent approximation is also available (Hautus et al., 1994; Irwin & Hautus, 1993), and was in fact used here. Because there were only 168 trials per observer, the variance associated with each point of the ROCs can be expected to be large; as a result, this procedure was not carried out for the still smaller number of trials that were presented to each visual field.

The results of this analysis are shown in Table 1 for 7 of the 10 observers. That 3 observers (H.S., O.T., and G.C.) could not judge whether the words belonged to the same semantic category or not is shown by the value of  $p(A)$  for these 3 observers, in the last column of the

**Table 1**  
**Maximum-Likelihood Estimates ( $d'$ ) of the Best-Fitting**  
**Parameters of the Differencing and Optimal Models**  
**for Each Observer in Experiment 1<sup>a</sup>**

Observer	Differencing Model		Optimal Model		$p(A)$
	$d'$	$\chi^2$	$d'$	$\chi^2$	
H.S.					.46
F.A.	1.98	2.562	1.61	1.860	.72
R.V.	0.76	3.111	0.78	1.464	.57
E.B.	0.98	7.301	0.98	4.324	.60
S.S.	1.22	10.251*	1.18	4.953	.63
M.M.	2.66	6.503	2.12	1.987	.83
O.T.					.50
V.H.	1.07	4.025	0.92	3.721	.58
Q.D.	1.35	9.156	1.18	6.398	.64
G.C.					.50
Total		42.906*		24.707	

<sup>a</sup>Also given are the goodness-of-fit statistic,  $\chi^2$ , and the area measure,  $p(A)$ . \* $p < .05$ .

table. The index  $p(A)$  is the proportion of the area under the ROC when the experimental points are joined by straight lines, and because this quantity makes no distributional assumptions, it is said to be nonparametric (McNicol, 1972). The value  $p(A) = 0.5$  obtained by O.T. and G.C. stems from an ROC lying along the major diagonal of the ROC square, and such an ROC represents chance performance. We have not fitted the models to the data from these observers because, in this limiting case, both models predict the same result, so that data of this kind cannot discriminate between them.

Table 1 also shows the goodness-of-fit statistic, chi-square, of each model for the data of each of the remaining 7 observers, together with the summed chi-square for those observers. Although there are probably too few trials to allow the data of each observer to decide between the models, the summed chi-squares, which have 28  $df$  (4 for each observer), show that the data departed significantly at the .05 level from the differencing model ( $p = .036$ ), but not from the optimal model ( $p = .644$ ). According to this analysis, therefore, the optimal model provides the better fit to the ratings.

We also tested which of the two models better described the resulting ROCs, by fitting the single-interval unequal-variance model to each observer's ROC. Although, as explained earlier, this is not the correct model for the *same-different* task, we reasoned that one of the parameters of that model (the ratio of the  $SDs$  of the two normal distributions or the slope of the ROC on  $z$ -coordinates) should signify which model better described the results. Because of its asymmetrical ROC, the differencing model predicts a slope parameter that is greater than 1, whereas the optimal model, because of its symmetrical ROC, predicts a slope parameter of exactly 1. The results of this test yielded a mean slope of 0.937, with a standard error of 0.059, for the 7 observers. This value is not significantly different from 1 [ $t(6) = 1.050$ , n.s.], a result that agrees with the predictions of the optimal model.

**Pooled data.** To provide information on any differences between the two visual fields, an analysis was made of the pooled ratings from all 10 observers. To pool the data, we added the frequencies with which each rating was made by each observer; the hit and false-alarm rates of the pooled ratings were computed from these summed frequencies (see Swets & Pickett, 1982). The parameters of the best-fitting population ROC were then estimated from these data, using the jackknifing technique since simple averaging of ROCs can yield misleading results. In this case, use of this technique involved a resampling plan combined with maximum-likelihood estimation (for a similar procedure, see Dorfman & Berbaum, 1986).

The best-fitting population ROCs based on the differencing and optimal models are shown in Figure 1. The jackknifed estimates of the ROC parameter,  $d'$ , obtained for each model are shown in Table 2 for each visual field, as well as for both fields combined, for all 10 observers. The calculated values of  $p(A)$ , ranging from 0.56 to 0.65, suggest that on average, the observers were able to judge whether the words belonged to the same or different categories. However, it should be appreciated that a given value of  $p(A)$  represents a more impressive level of performance on the *same-different* task than it does on, for example, the single-interval rating task (Macmillan & Creelman, 1991). On the other hand, the detection-theory measure of accuracy shown in Table 2 ( $d'$ ) has the same meaning in this task as it does in other tasks. Visual inspection of the fitted curves for the pooled ratings (see Figure 1) supports the conclusion based on the summed chi-squares—namely, that the differencing model does not provide a good fit to the data. The optimal model, by contrast, provides an excellent fit. This comparison has to be made by eye because at present, there is no goodness-of-fit statistic available for jackknifed *same-different* ROCs.

A two-way analysis of variance (ANOVA) of the values of  $p(A)$  for each of the 10 observers, with visual field and word length as factors, showed that accuracy was significantly better for words in the RVF [ $p(A) = 0.64$ ] than it was for words in the LVF [ $p(A) = 0.55$ ,  $F(1,9) = 9.23$ ,  $p < .025$ ]. There was no significant difference in the accuracy with which three-, four-, and five-letter words were judged, nor was the interaction between visual field and word length significant. There was a general, but not significant, decrease in accuracy with increasing word length. Table 3 shows the average value of  $p(A)$  for each word length, and the corresponding estimates of  $d'$  (from the jackknifing method) for each model. All of these measures of accuracy declined with increasing word length.

## Discussion

The symmetrical shape of the obtained ROCs is consistent with the adoption of an optimal strategy, rather than a differencing strategy, for these judgments of semantic category. The category of the central word was

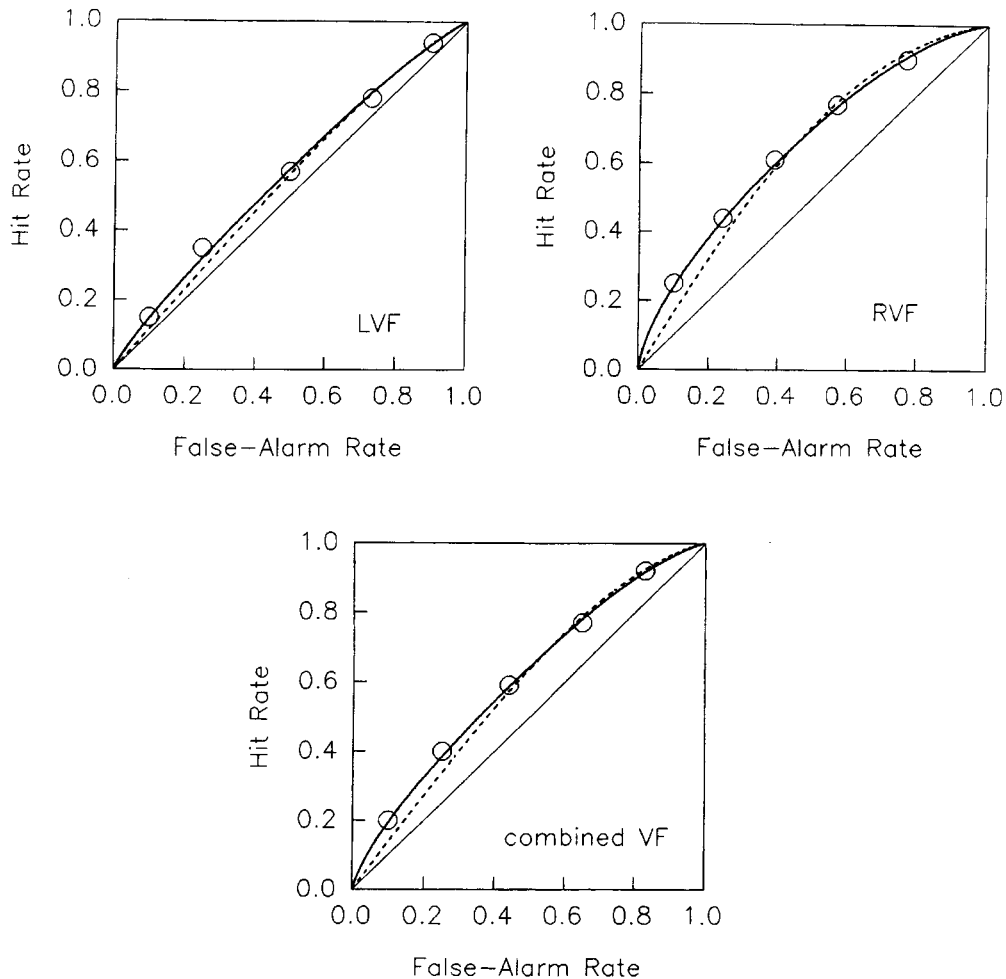


Figure 1. Best-fitting jackknifed ROCs of 10 observers in Experiment 1, for the LVF, the RVF, and the combined visual fields. The theoretical curves are for the differencing model (broken lines) and the optimal model (solid lines).

relatively easy to determine because it was attended to, whereas that of the side word would have been much more difficult to determine because of its brief exposure in the parafovea. The relative ease with which the central word could be categorized might have promoted the adoption of an optimal strategy in this experiment. The procedure might have encouraged observers to classify the central word and to judge the side word independently of the central one. The final decision would then be based on whether those two classifications agreed. In a second experiment, therefore, we avoided this arrangement in order to provide a fairer test of the two models.

The ROC parameters,  $d'$  and  $p(A)$ , demonstrate a clear RVF advantage for the accuracy of *same-different* judgments of word meaning. This advantage might reflect a tendency of the observers to scan from left to right (Heron, 1957), although this is perhaps unlikely, because the side words were presented for only 30 msec. Such a mechanism, at least with the English language, might favor words displayed in the RVF, but there is some debate about the contribution of this process to visual-field

asymmetries (Bradshaw, 1989). Another possibility is that the RVF advantage stems from the dominance of the left hemisphere in processing verbal stimuli. Still another possible explanation is that the difference between the two fields stems from an attentional asymmetry favoring the left hemisphere (Kinsbourne, 1970). The conditions in our experiment may have encouraged this attentional asymmetry. The need to categorize the central word first may have evoked greater activation in the left hemisphere, causing a shift of attention to the RVF. This, in turn, may have amplified any RVF advantage. The

Table 2  
Estimates of Parameters of the Pooled *Same-Different* ROCs When the Side Words in Experiment 1 Were Presented to the LVF, to the RVF, and for Combined Visual Fields

Parameter	LVF	RVF	Combined Fields
$p(A)$	0.56	0.65	0.61
$d'$ (differencing)	0.77	1.43	1.14
$d'$ (optimal)	0.71	1.21	1.00

**Table 3**  
**Estimates of Parameters of the Pooled Same-Different ROCs for Three-, Four-, and Five-Letter Words in Experiment 1**

Parameter	Word Length (Letters)		
	3	4	5
$p(A)$	0.62	0.61	0.58
$d'$ (differencing)	1.35	1.07	0.57
$d'$ (optimal)	1.11	1.01	0.83

origin of the RVF advantage in this experiment is, therefore, uncertain.

## EXPERIMENT 2

In a second experiment, we attempted to obtain well-defined ROCs for individual observers so that goodness-of-fit statistics could be obtained for the RVF and the LVF, both separately and combined, for each observer. The observers therefore all undertook a larger number of trials than in Experiment 1. The trials were presented in three separate sessions. Another difference was that in Experiment 2, the words were presented unilaterally, so that any putative advantage for the optimal strategy that stemmed from the observers' being able to see one of the words relatively easily might be minimized. By this means, too, we hoped to clarify the comparison between the performance of the two visual fields, since unilateral presentation overcomes the problem of controlling fixation (Bryden, 1988). As in Experiment 1, we also studied whether visual-field asymmetries depended on word length.

### Method

**Observers.** Three men and four women, all psychology students volunteered to participate; six described themselves as right-handed, and one (K.W.) as mixed-handed. One other observer completed the first session but not the subsequent sessions, and two others (not included above) were excluded (one neglected to follow the instructions concerning central fixation, and the other was visually impaired).

**Apparatus and Stimuli.** The same equipment, words, and category definitions were employed as in Experiment 1. A separate pool of six-letter words was available for the practice session.

**Procedure.** The observers fixated upon a central cross (+). Two words were then presented simultaneously for 150 msec, one above the other, with their inner edge 3° to either the left or the right of center. Their vertical separation was 0.7°. Following the presentation of a pair of words, a six-point rating scale appeared on the screen, and the observers, who were provided with written instructions on how to use the rating scale, were prompted to rate their confidence that the pair of words belonged either to the same semantic category or to different categories. Reaction times were not recorded.

As in Experiment 1, the trials were self paced and no feedback was offered. The observers were instructed at the beginning of the experiment to look only at the center of the screen during a trial, and to use all the categories of the rating scale. Each observer undertook one practice session of 72 trials, followed by three experimental sessions, each of which comprised two blocks of 84 trials. Within the experimental session, there were six warm-up trials with a separate stimulus pool. The first experimental session fol-

lowed the practice session, and the remaining two sessions were performed on separate occasions (except for those of J.W., sessions were conducted on separate days). The words presented on a trial were chosen from the set at random, but with the constraint that they were of equal length, and they had an equal probability of belonging to the same category as of belonging to different categories. An equal number of trials presented the words to the LVF as to the RVF, and trials were equally divided, too, in terms of whether they contained three-, four-, or five-letter words.

### Results

**Individual observers.** ROCs were constructed for each observer's LVF, RVF, and combined visual fields. The differencing and optimal models were then fitted to these ROCs. The maximum-likelihood estimates of the parameter  $d'$  are shown in Table 4 for the two models for each observer, together with their associated goodness-of-fit statistic, chi-square. With the exception of the LVF data of Observer L.H., the optimal model fits the data of every observer for each visual field and for the combined visual fields ( $p > .05$ ). The differencing model does not fit the combined-visual-field data for 5 of the 7 observers. Both models fit the RVF data. One observer (K.W.) was noticeably less accurate than the others and performed at chance level for words in the LVF. In gen-

**Table 4**  
**Maximum-Likelihood Estimates of Best-Fitting Parameters of the Differencing and Optimal Models for the Combined Fields, the LVF, and the RVF, for Each Observer in Experiment 2**

Observer	Model				$p(A)$
	Differencing		Optimal		
	$d'$	$\chi^2$	$d'$	$\chi^2$	
Combined Fields					
A.F.	1.37	17.407†	1.23	6.414	.64
B.M.	1.59	15.525†	1.35	8.681	.67
L.H.	1.60	11.247*	1.33	8.583	.66
K.W.	0.63	5.084	0.63	3.248	.55
D.W.	2.28	10.102*	1.84	1.499	.78
J.W.	2.38	5.429	1.93	2.836	.79
I.C.	1.95	17.097†	1.62	4.153	.73
Total		81.891†		35.414	
LVF					
A.F.	1.43	14.357†	1.28	8.052	.65
B.M.	1.50	9.731*	1.30	5.687	.65
L.H.	1.47	20.134†	1.23	19.210†	.64
K.W.					.50
D.W.	1.98	6.562	1.59	4.777	.73
J.W.	2.10	5.494	1.71	3.733	.74
I.C.	1.59	11.540*	1.37	6.721	.67
Total		67.818†		48.180†	
RVF					
A.F.	1.34	7.603	1.19	3.024	.63
B.M.	1.69	7.625	1.42	4.626	.68
L.H.	1.75	2.331	1.43	0.693	.69
K.W.	0.93	5.569	0.85	3.574	.58
D.W.	2.61	8.718	2.09	2.833	.82
J.W.	2.65	1.347	2.15	0.674	.82
I.C.	2.28	8.846	1.85	1.542	.78
Total		42.039*		16.966	

\* $p < .05$ . † $p < .01$ .

eral, the summed chi-squares with 28  $df$  lead to the same conclusion as the one based on the individual goodness-of-fit statistics; that is, they show that the data are unlikely to have stemmed from a differencing strategy, but that they could well have resulted from an optimal strategy. The exception is the summed chi-square for the LVF, but that sum is heavily influenced by the bad fit of the LVF data of Observer L.H. to the optimal model.

As in Experiment 1, we estimated the slope of the resulting ROCs on  $z$ -coordinates by fitting the single-interval unequal-variance model to each observer's ROC for the combined visual fields. The mean slope was 1.001, with a standard error of 0.066, a result in accordance with the predictions of the optimal model.

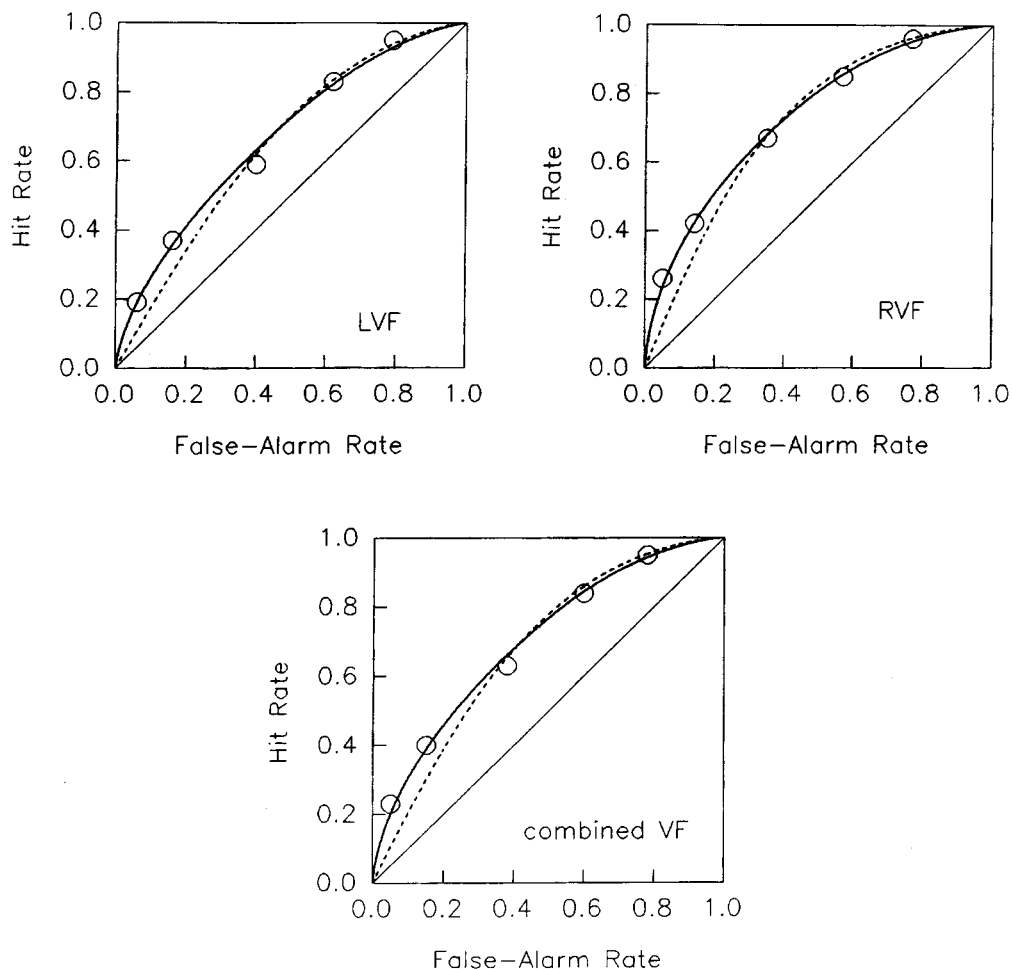
**Pooled data.** Figure 2 shows the ROCs for the pooled ratings of the 7 observers. Separate ROCs are shown for the LVF, the RVF, and the combined visual fields. For these pooled ratings,  $p(A)$  was 0.66 for the LVF and 0.73 for the RVF. This nonparametric measure, like the values of  $d'$  (see Table 4), denotes better performance in the RVF. Visual inspection of the curves in Figure 2 shows a better fit for the optimal model than for the differenc-

ing model for each visual field and for the combined data.

On the basis of Chiarello's (1988) suggestion that visual-field asymmetries may vary with word length, we performed an ANOVA on the seven obtained values of  $p(A)$  for each word length within each visual field. On average, accuracy for words in the RVF [ $p(A) = 0.71$ ] was significantly greater than it was for words in the LVF [ $p(A) = 0.65$ ,  $F(1,6) = 15.17$ ,  $p < .01$ ]. The values of  $p(A)$  decreased with increasing word length, especially in the LVF; however, the effect of word length was not significant.

### Discussion

The results of this experiment confirmed those of Experiment 1, in that (1) observers adopted the optimal decision strategy when judging whether two words belong to the same semantic category; and (2) they achieved greater accuracy for words in the RVF than they did for those in the LVF. In Experiment 1, one word was presented in central fixation and the other in the LVF or RVF, whereas in Experiment 2, both words were pre-



**Figure 2.** Best-fitting jackknifed ROCs for 7 observers in Experiment 2, for the LVF, the RVF, and the combined visual fields. The theoretical curves are for the differencing model (broken lines) and the optimal model (solid lines).



sented in either the LVF or the RVF. Despite this difference, the basic pattern of results was the same in both experiments; evidently, the presentation in central fixation of one of the words was not necessary for the adoption of the optimal strategy.

Of the 7 observers who completed the three sessions, the data of 2 were not significantly different from the prediction of either model, and did not allow the models to be distinguished. Of these 2 observers, one (K.W.) found the task very difficult, and as this observer's ROC traced a path near the major diagonal, the data do not discriminate between the two models. In this limiting case, both models predict the same result. The other observer (J.W.) used the rating categories in such a way that the points on the resultant ROC were clustered. As a result, the path of the ROC was not clearly defined for this observer and, again, it was not possible to discriminate between the two models.

The results of the ROC analysis of the three-, four-, and five-letter words did not support the hypothesis that three-letter words are responded to more accurately when presented in the LVF than they are when presented in the RVF. As word length increased, however, there was possibly a greater decrement in performance in the LVF than in the RVF. If genuine, this result could identify a right-hemisphere inefficiency in encoding words in a letter-by-letter fashion.

## GENERAL DISCUSSION

### The Decision Strategy Adopted

The results of the two experiments showed that observers adopted an optimal decision strategy when making *same-different* judgments about the meaning of words. The proper model for the ROC analysis of semantic *same-different* tasks is, therefore, the optimal model. This result contrasts with the findings of Hautus et al. (in press), Irwin et al. (1993), and Vogels and Orban (1986), who found that observers adopted a differencing strategy when judging whether two tones, two tastes, or two lines were the same or different. It may be that for simple stimuli, observers are likely to adopt a differencing strategy because, for such stimuli, it is possible to make only relative judgments—of one stimulus relative to the other—and not absolute judgments of each stimulus on its own.

The capacity to make absolute judgments may, for its part, be understood in terms of a more fundamental principle derived from psychophysical experiments. These experiments (e.g., Hautus et al., 1994; Pollack, 1952) have shown that observers are unable to make absolute judgments of the magnitude of simple sensory dimensions, and in this sense, therefore, psychophysical judgments of this kind are relative judgments. The ability to make absolute judgments is probably confined to judgments about more complex stimuli. The fact that the observers in the present experiments were able to adopt the optimal strategy in making their decisions must mean that the semantic dimension of naturalness or artificiality is suffi-

ciently complex to permit the absolute judgments that the optimal strategy requires. In contrast, other studies have found that when events lie on a sensory dimension that precludes absolute judgments of their magnitude, observers have adopted only the differencing strategy when deciding whether those events are the same or different (Hautus et al., 1994; Irwin et al., 1993).

The brevity with which the words were exposed in our experiments meant that they were difficult to perceive, and it was partly for this reason, no doubt, that the observers did not achieve near-perfect accuracy in their *same-different* judgments. Presumably, also, some words conveyed their meanings more decisively than others. The net effect of such factors meant that the distinction between those words that referred to natural objects and those that referred to artificial things was to some degree ambiguous—in fact, on the evidence of Experiment 1, the underlying distributions of the two classes of words were separated by exactly  $d' = 1$  (to two decimal places; see the last entry of Table 2). Regardless of the source of the uncertainty about the category to which a word belonged, the *decision* had to be based on the semantic properties and not on the physical properties of the word. The *evidence* for that decision was uncertain, but the *process* by which the words were encoded as artificial or natural was not addressed by our experiments and their associated theory.

We made the semantic decision difficult in order to allow the nature of the decision strategy to be inferred from the shape of the ROC. In fact, we may have overestimated the desirable degree of difficulty in Experiment 1, and for this reason, we increased the duration of exposure in Experiment 2. This increased the level of accuracy, so that we did not obtain ROCs that lay along or near the major diagonal of the ROC square. In other words, we used the duration of exposure as a method for controlling level of performance, rather than as a systematic variable of our study. This procedure may have affected the way the words were processed in each hemisphere. Chiarello (1988) reported that masking words with patterns favored their processing in the left hemisphere. On the other hand, Sergent (1987) reported that degrading the fidelity of pictures favored their processing by the right hemisphere. We do not know whether either of these phenomena obtained in our experiment.

It should be emphasized that the symmetrical ROCs of the optimal strategy displayed in Figures 1 and 2 are not those of the familiar normal-normal equal-variance model for a single observation interval, and, similarly, that the asymmetrical ROCs of the differencing strategy are not those of the unequal-variance normal model; for one thing, the unequal-variance normal model has two free parameters, whereas the differencing model has only one. The ROCs we have fitted are, however, *derived* from the normal-normal equal-variance model (see Appendix, and Macmillan & Creelman, 1991, Chapter 6), and the fits therefore provide an estimate of the model's parameter,  $d'$ . For this reason, our estimate of  $d'$  has exactly the same significance in this task as it does in other

detection tasks. Tables 1, 2, 3, and 4 show that the estimated value of  $d'$  (whether for groups or individual observers) is larger for the differencing strategy than for the optimal strategy. This is because the differencing strategy is suboptimal, and a larger signal-to-noise ratio (i.e., a larger  $d'$ ) is therefore required to achieve a given level of performance when this strategy is used than when the optimal strategy is used. However, since the differencing strategy does not provide a good fit to much of our data, the more appropriate estimates of  $d'$  are obtained by fitting the optimal strategy.

Our analysis draws on a theory that is usually applied to psychophysical judgments about sensory dimensions, but we have attempted to extend it to cognitive judgments of word meaning. Judgments of sameness and difference can be viewed as judgments about the category to which a stimulus belongs; thus in our experiments, observers judged whether the objects that words referred to should be categorized as natural or as artificial. The affinities between the processes involved in these two kinds of judgments, one about sensory dimensions and the other about abstract categories, have been tabulated by Medin and Barsalou (1987). Although the two fields of inquiry have largely developed independently, Medin and Barsalou argue that the theoretical and empirical similarities of the research in the two areas are sufficiently deep to suggest that both kinds of judgment reflect common processes—a view shared by Harnad (1987). To illustrate these affinities, we note that the model for semantic classifications proposed by Smith, Shoben, and Rips (1974) includes the idea of a criterion boundary, not unlike that of the optimal model of detection theory. In their model, however, there are two criteria, and observations falling between the criteria are subjected to further processing, and thus to delayed responding. Their dual-stage model is intended to account for reaction time in semantic categorization, whereas the detection-theoretic model we have presented is silent on that matter.

### Visual-Field Asymmetries

This ROC analysis of *same-different* judgments revealed a consistent performance advantage for words presented in the RVF. An RVF advantage for word recognition is one of the most robust findings of visual-field asymmetries (Bryden, 1982), and may be attributable either to left-to-right scanning or to a left-hemisphere superiority for word recognition. Since the RVF advantage that we found was independent of word length, our results do not help resolve differences among competing theories about how word length affects the way words are processed by each hemisphere (see Chiarello, 1988; Coltheart, 1980).

### Conclusions

We have emphasized how the *same-different* experiment is subject to complexities that arise from the different decision strategies that observers can adopt in deciding whether words have the same semantic properties

or not. We have tried to show how an appropriate ROC analysis may uncover those decision strategies and thereby contribute to the understanding of cognitive functioning. In particular, we have shown that for *same-different* judgments of semantic category, observers adopt an optimal, rather than a differencing, decision strategy. Not only does ROC analysis resolve this ambiguity in *same-different* judgments, it provides unbiased indices of the accuracy of these judgments. We have illustrated the potential usefulness of this approach by extracting those indices from *same-different* ROCs to re-examine hemispheric asymmetries in the processing of semantic information. We found a distinct advantage for the RVF in this task.

### REFERENCES

- ALGOM, D. (1992). In D. Algom (Ed.), *Psychophysical approaches to cognition: Advances in psychology* (Vol. 92, pp. 1-12). Amsterdam: Elsevier.
- BAGNARA, S., BOLES, D. B., SIMION, F., & UMITLÀ, C. (1983). Symmetry and similarity effects in the comparison of visual patterns. *Perception & Psychophysics*, **34**, 578-584.
- BARRON, R. W., & HENDERSON, L. (1977). The effects of lexical and semantic information on same-different visual comparison of words. *Memory & Cognition*, **5**, 566-579.
- BARRON, R. W., & PITTINGER, J. B. (1974). The effect of orthographic structure and lexical meaning on "same"- "different" judgments. *Quarterly Journal of Experimental Psychology*, **26**, 566-581.
- BORLAND (1990). *Turbo Pascal 6.0 library reference*. Scotts Valley, CA: Borland International.
- BRADSHAW, J. L. (1989). *Hemispheric specialization and psychological function*. Chichester, U.K.: Wiley.
- BRYDEN, M. P. (1982). *Laterality: Functional asymmetry in the intact brain*. New York: Academic Press.
- BRYDEN, M. P. (1988). Cerebral specialization: Clinical and experimental assessment. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 1, pp. 143-159). Amsterdam: Elsevier.
- CHIARELLO, C. (1988). Lateralization of lexical processes in the normal brain: Review of visual half-field research. In H. A. Whitaker (Ed.), *Contemporary reviews in neuropsychology* (pp. 36-76). New York: Springer-Verlag.
- COLTHEART, M. (1980). Deep dyslexia: A right hemisphere hypothesis. In M. Coltheart, K. Patterson, & J. Marshall (Eds.), *Deep dyslexia* (pp. 326-380). London: Routledge & Kegan Paul.
- DORFMAN, D. D., & BERBAUM, K. S. (1986). RSCORE-J: Pooled rating-method data: A computer program for analyzing pooled ROC curves. *Behavior Research Methods, Instruments, & Computers*, **18**, 452-462.
- EGETH, H., & EPSTEIN, J. (1972). Differential specialization of the cerebral hemispheres for the perception of sameness and difference. *Perception & Psychophysics*, **12**, 218-220.
- FARELL, B. (1985). "Same"- "different" judgments: A review of current controversies in perceptual comparisons. *Psychological Bulletin*, **98**, 419-456.
- FRANCIS, M. A. (1993). *Facilitation and inhibition between attended and unattended words: Addressing the issue of early versus late selection*. Unpublished doctoral dissertation, University of Auckland, Auckland.
- GIBSON, A. R., DIMOND, S. J., & GAZZANIGA, M. S. (1972). Left field superiority for word matching. *Neuropsychologia*, **10**, 463-466.
- GREEN, D. M., & SWETS, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- GROSS, M. M. (1972). Hemispheric specialization for processing of visually presented verbal and spatial stimuli. *Perception & Psychophysics*, **12**, 357-363.
- HARNAD, S. (1987). Category induction and representation. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 535-565). Cambridge: Cambridge University Press.

- HAUTUS, M. J., IRWIN, R. J., & SUTHERLAND, S. (1994). Relativity of judgements about sound amplitude and the asymmetry of the same-different ROC. *Quarterly Journal of Experimental Psychology*, *47A*, 1035-1045.
- HERON, W. (1957). Perception as a function of retinal locus and attention. *American Journal of Psychology*, *70*, 38-48.
- IRWIN, R. J., & HAUTUS, M. J. (1993). *Decision strategies in same-different judgments revealed by their receiver operating characteristics and psychometric functions*. Unpublished manuscript.
- IRWIN, R. J., STILLMAN, J. A., HAUTUS, M. J., & HUDDLESTON, L. M. (1993). Measurement of taste discrimination with the same-different task: A detection-theory analysis. *Journal of Sensory Studies*, *8*, 229-239.
- JOHNSON, K. O. (1980). Sensory discrimination: Decision processes. *Journal of Neurophysiology*, *43*, 1771-1792.
- KINSBOURNE, M. (1970). The cerebral basis of lateral asymmetries in attention. *Acta Psychologica*, *33*, 193-201.
- MACMILLAN, N. A., & CREELMAN, C. D. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.
- MACMILLAN, N. A., KAPLAN, H. L., & CREELMAN, C. D. (1977). The psychophysics of categorical perception. *Psychological Review*, *84*, 452-471.
- MCNICOL, D. M. (1972). *A primer of signal detection theory*. London: Allen and Unwin.
- MEDIN, D. L., & BARSALOU, L. W. (1987). Categorization processes and categorical perception. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 455-490). Cambridge: Cambridge University Press.
- NOREEN, D. L. (1981). Optimal decision rules for some common psychophysical paradigms. In S. Grossberg (Ed.), *Mathematical psychology and psychophysiology* (SIAMS-AMS Proceedings, Vol. 13, pp. 237-279). Providence, RI: American Mathematical Society.
- POLLACK, I. (1952). Information of elementary auditory displays. *Journal of the Acoustical Society of America*, *24*, 745-750.
- SERGEANT, J. (1987). Information processing and laterality effects for object and face perception. In G. W. Humphreys & M. J. Riddoch (Eds.), *Visual object processing: A cognitive neuropsychological approach* (pp. 145-173). Hove, U.K.: Erlbaum.
- SMITH, E. E., SHOEN, E. J., & RIPS, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, *81*, 214-241.
- SORKIN, R. D. (1962). Extension of the theory of signal detectability to matching procedures in psychoacoustics. *Journal of the Acoustical Society of America*, *34*, 1745-1751.
- SWETS, J. A., & PICKETT, R. M. (1982). *Evaluation of diagnostic systems: Methods from signal detection theory*. New York: Academic Press.
- THORNDIKE, E. L., & LORGE, I. (1944). *The teacher's handbook of 30,000 words*. New York: Columbia University, Teachers College Press.
- URCUOLI, P. J., KLEIN, R. M., & DAY, J. (1981). Hemispheric differences in semantic processing: Category matching is not the same as category membership. *Perception & Psychophysics*, *29*, 343-351.
- VOGELS, R., & ORBAN, G. A. (1986). Decision processes in visual discrimination of line orientation. *Journal of Experimental Psychology: Human Perception & Performance*, *12*, 115-132.

#### NOTES

1. Reaction times to the central word were measured in order to determine whether the semantic category of the side word facilitated or inhibited responses to the central word. An ANOVA of the reaction times to the central word showed that reaction time was unrelated to the visual field in which the side word appeared, to the length of the word, or to whether the side word belonged to the same category as the central word or to a different category. None of the interactions between these factors were significant, either. This work is reported in more detail elsewhere (Francis, 1993).

2. See Note 1.

#### APPENDIX

The shape of the ROC for the differencing strategy has been known to be asymmetrical since Sorkin's (1962) analysis of this design. The parametric equations for this ROC have been presented by Macmillan, Kaplan, and Creelman (1977), and in slightly different form by Irwin, Stillman, Hautus, and Huddleston (1993). The nature of the asymmetry depends on the definition of hits and false alarms. For the definition adopted here (i.e., that a hit consists of calling two words the same when they are in fact the same, and a false alarm consists of calling two words the same when they are in fact different), the asymmetry takes the form shown by the broken curves in Figures 1 and 2. When hits and false alarms are defined differently, as by Macmillan and Creelman (1991), the ROC is similar to those in Figures 1 and 2, but is reflected about the negative diagonal.

The shape of the ROC for the optimal strategy is less well known. In this appendix, we demonstrate its symmetry by a geometric argument. In Figure A1, which shows the decision space for the same-different task in a diagram similar to one originally presented by Noreen (1981),  $X_1$  is a random variable associated with the observation of one word on a trial, and  $X_2$  is a random variable associated with the other word on a trial. There are four possible kinds of trial in our experiments: (1) <NN> trials, in which both words can refer to natural things; (2) <AA> trials, in which both words can refer to artificial things; (3) <NA> trials, in which the two words can refer to things of different semantic categories; and (4) <AN> trials, in which the two words can refer to things of different semantic categories, but in locations that are the reverse of those in (3).

The observations associated with each class of word are assumed to be normally distributed, with means separated by  $d'$ . Because there are two words in a trial, the outcome of all possible trials of that type is represented by a bivariate normal distribution rather than by the more familiar distributions of the single-interval experiment. In Figure A1, the means of the four bivariate distributions lie at the center of the circles in each quadrant of the space. The circles depict the locus of equal-

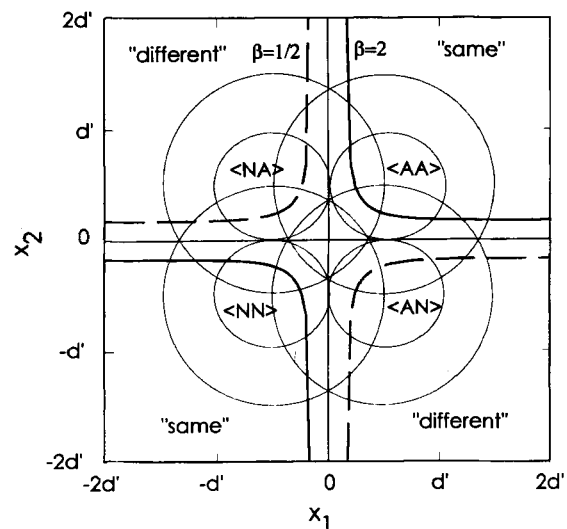


Figure A1. Decision space for the same-different task. (See text for explanation of terms.)

probability densities at the first two *SDs* from the mean of each bivariate distribution.

A decision based on the optimal strategy is optimal in the sense that the criterion is monotonic with likelihood ratio,  $L(x)$ . In the two-dimensional decision space of the *same-different* task, a given likelihood ratio is represented by a line rather than a point (as in the single-interval experiment). For the *same-different* experiment,  $L(x)$  can be expressed in terms of  $d'$  by making use of the relation  $L(x) = e^{d'x}$  (Green & Swets, 1966, Equation 3.3b). For the decision space in Figure A1, Noreen (1981) has shown that  $L(x)$  is given by:

$$L(x) = \frac{e^{d'(x_1+x_2)} + 1}{e^{d'x_1} + e^{d'x_2}}. \quad (\text{A1})$$

The loci of three examples of constant  $L(x)$  are shown in Figure A1: The axes show the special case when  $L(x) = 1$ ; the solid curves in the upper-right and lower-left quadrants are for the case when  $L(x) = 2$ ; and the broken curves in the other quadrants are for the case when  $L(x) = 1/2$ .

Consider the case when  $L(x) = 2$ . As Figure A1 illustrates, the outcome of any trial that lies beyond this contour in the upper-right quadrant will be called "same"; these points will mostly derive from  $\langle \text{AA} \rangle$  trials, but a few will derive from  $\langle \text{NN} \rangle$  trials; in either case, the result will be a hit. On the other hand, when the outcome of an  $\langle \text{NA} \rangle$  or  $\langle \text{AN} \rangle$  trial lies beyond this boundary, the result will be a false alarm. Likewise, trial outcomes that fall below the contour in the lower-left quadrant will be called "same." These will mostly stem from  $\langle \text{NN} \rangle$  trials, in which case they will be hits. False alarms will again derive from  $\langle \text{NA} \rangle$  or  $\langle \text{AN} \rangle$  trials that fall in this space. Other parts of

the decision space represent cases in which the decision will be "different." When that decision follows an  $\langle \text{NA} \rangle$  or  $\langle \text{AN} \rangle$  trial, it will result in a correct rejection, and when it follows an  $\langle \text{AA} \rangle$  or  $\langle \text{NN} \rangle$  trial, it will result in a miss.

It can now be seen why the ROC for the optimal strategy is symmetric about the negative diagonal of the ROC square. The demonstration rests on certain symmetries between the areas of the decision space bounded by a given likelihood ratio (which we will call  $\beta$ ), and the areas bounded by  $1/\beta$ , the reciprocal of that likelihood ratio. The hit rate and the false-alarm rate are determined from the bivariate normal distributions that lie within these areas. Figure A1 shows that the false-alarm rate for  $\beta = 2$  is identical to the miss rate for  $\beta = 1/2$ , and this will be true for any pair of likelihood ratios that stand in a reciprocal relation. An identity also holds for the correct-rejection rate and the hit rate (the complements of the false-alarm and miss rates), whereby the correct-rejection rate for  $\beta = 2$  is identical to the hit rate for  $\beta = 1/2$ . These pairs of coordinates specify points in the ROC square that have reflectional symmetry about the negative diagonal. This symmetry holds for all reciprocally related likelihood ratios, and so the complete ROC is symmetrical about the negative diagonal.

Although this geometric reasoning shows that the *same-different* ROC for the optimal strategy is symmetric about the negative diagonal of the ROC square, it does not define its exact location or shape. That definition requires solving the parametric equations for the hit rate and the false-alarm rate as a function of  $\beta$ . These equations have no analytic solution (Irwin & Hautus, 1993), but they can be solved numerically for any value of  $d'$ . Examples are shown by the solid lines in Figures 1 and 2 for the best-fitting values of  $d'$  for our data.

(Manuscript received December 7, 1993;  
revision accepted for publication June 20, 1994.)