A comparison of semantic, acoustic and visual criteria for matching of word pairs

GILLIAN COHEN[1]
INSTITUTE OF EXPERIMENTAL PSYCHOLOGY, OXFORD, ENGLAND

Ss judged the same pairs of words "same" or "different" under semantic, acoustic and visual criteria. RTs were compared for each criterion, and the effects of different kinds of confusability, such as acoustic similarity in the semantic matching task, or semantic similarity in the acoustic matching task, were also studied.

The purpose of this experiment was to compare matching of visually presented word pairs under three different criteria, semantic, acoustic and visual. The S was required to judge the members of the pair "same," or "different," in respect of (a) meaning, (b) sound, and (c) physical identity or spelling. The experiment was designed to reveal any systematic differences which might exist between the three levels of processing, and also any sequential dependencies between them. A knowledge of the relationships between the different levels of processing should assist our understanding of the operations involved in word recognition and in reading, and go some way towards answering the following questions:

(1) Is word recognition performed by hierarchical analysis proceeding sequentially from visual to acoustic to semantic?

(2) Can any of the three stages be omitted, and under what conditions is this possible?

(3) Does the evidence support a parallel or a serial processing model?

The design of the experiment was suggested by that of Posner and Mitchell (1967). In their experiment the S matched pairs of letters at three different levels, physical identity (AA), name identity (Aa), and rule identity (AE–both vowels). In the present experiment the same method is extended to classification of words, with very similar results.

METHOD

A set of 20 word pairs was compiled. Each pair was typed in capitals side by side on a 4 x 6 card, and subtended a visual angle of approximately 5 deg. The cards were exposed singly in a Cambridge tachistoscope. Reaction times were measured by a Decatron timer from the onset of exposure until the S terminated the display by responding. Ss were instructed to classify each pair as "same" or "different" by pressing the appropriately labelled switch. Half of the right-handed Ss had the right switch for "same" and the left switch for "different," and half had the opposite arrangement. The left-handed Ss were similarly divided between the two arrangements. Reaction times, which are accurate to within 10 msec, and errors were recorded.

Thirteen Ss completed the experiment. They were graduates and undergraduates of both sexes. Each S had one complete block of 20 practice trials with a special set of practice cards. The experimental set was then presented to each S three times, once under each of the three criteria. The order of the criteria, and the order of the pairs within the set were counterbalanced over Ss. The complete session lasted about 45 min. The S rested for 3 min between each change of criterion. Ss were paid for their services. They were instructed to respond as fast as possible but to avoid errors, but there was no pay-off for success.

The set of 20 word pairs was composed of the following types:

Type I (identical)–five of the pairs were identical as in sew-sew.

Type NC (not confusable)–five of the pairs were not confusable under any of the three criteria, and shared no letter as in sew-run.

Type S (semantically same) or SC (semantically confusable)–five of the pairs were "same" by the semantic criterion, but "different" by the other two criteria, although semantically confusable, as in sew-stitch.

Type S (acoustically same) or AC (acoustically confusable)–five of the pairs were "same" by the acoustic criterion, but "different" by the other two criteria, although acoustically confusable, as in sew-so.

Thus in the semantic matching task two types of pairs required the response "same," the identical pairs (e.g., sew-sew) and the semantically equivalent pairs (e.g., sew-stitch); and two types of pairs required the response "different," the NC pair (e.g., sew-run) and the AC pair (e.g., sew-so). In the acoustic matching task the I pairs (sew-sew, etc.) and the S pairs (sew-so) required the response "same" and the NC (sew-run) and SC (sew-stitch) pairs were "different." In the visual matching task only the identical pairs were "same"; all the other pairs were "different."

Thorndike-Lorge frequencies were balanced over each of these five types of pairs.

RESULTS

The mean RTs for each type of word pair under each criterion are tabulated below. Error scores were excluded.

For the I and NC types the stimulus-response combination remains constant under all three criteria, and differences in response times can fairly be attributed to differences in processing time. Accordingly an analysis of variance was performed on the I and NC RTs under the three criteria. The difference between the criteria yields $F = 31.26$, $df = 2/24$, $p < .001$). The difference between I RTs and NC RTs is also significant, $F = 22.8$, $df = 1/12$, $p < .001$. Three analyses of variance were performed on the RTs under each criterion. The difference between the pair types proved significant in each case, as follows:

Semantic matching–$F = 14.9$, $df = 3/36$, $p < .001$
Acoustic matching–$F = 6.805$, $df = 3/36$, $p < .001$
Visual matching–$F = 6.76$, $df = 3/36$, $p < .001$

The Newman-Keuls test was used to compare each of the pair types with all the others within each criterion. Under the semantic criterion I differs significantly from S, from NC and from AC, but these three do not differ from each other. In the acoustic task the result is similar; I differs from S, from NC and from SC which do not differ from each other. In the visual matching task AC differs from SC and from NC. The level of significance is $p < .01$.

The Newman-Keuls method was also used to further explore the differences between I and NC scores across all three criteria. I differs significantly from NC in the semantic and acoustic tasks, but not in the visual. Acoustic I and Visual I both differ from Semantic I, but not from each other. Semantic NC, Acoustic NC and Visual NC all differ significantly from each other. Again the level of significance is $p < .01$.

Error rate was 4.4%. The error distribution is set out in Table 2.

There was no effect of the order in which an S was presented with the three criteria. To test this the means of those Ss performing the semantic task first were compared with those who performed it last, and similarly for each criterion. No significant

Table 1
Mean RTs (in milliseconds) for the Pair Types under Each of the Three Criteria

Response	Type	Semantic	Acoustic	Visual
Same	I	839	716	678
	S	1001	839	—
Different	NC	1089	842	612
	AC	1099	—	725
	SC	—	819	612

Table 2
Error Distribution

	Semantic	Acoustic	Visual	Totals
I	4	0	3	7
S	5	2	–	7
NC	2	1	1	4
AC	5	–	6	11
SC	–	3	0	3
Totals	16	6	10	32

differences appeared. Individual S's scores were examined to see if fast matching on a particular criterion correlated with susceptibility to interference from that criterion, i.e., if a fast acoustic matcher had more trouble with AC pairs in semantic matching. Susceptibility to interference was also examined to see if it was general (i.e., both semantic and acoustic) or specific. No systematic trends of this kind could be discerned. Some Ss tended to be fast for all criteria and over all types of pairs, while others were slow throughout.

DISCUSSION

Points arising from these results are as follows:

(1) *The difference between I RTs under different criteria.* The matching of identical word pairs (e.g., sew-sew) takes longer under the acoustic than under the visual criterion, although this difference just fails to reach a significant level. Under the semantic criterion RTs to such pairs are significantly longer. Posner, who obtained a similar result, attributes it to a low level of practice, and found that it tended to disappear with more practice. Two interpretations of this finding are possible:

(a) The S does not make all three kinds of match on the basis of visual identity alone, as he logically could, but proceeds to analyze sound or meaning also, and matches along the relevant criterion. But with this interpretation, how can the difference between I and S RTs which occurs both in Posner's study, and in the present experiment be explained? It appears that physical identity or physical similarity facilitates judgments of "same" by other non-physical criteria. In an earlier study Posner (1964) found classification time to be a linear function of the similarity of the members of the pair. In his 1967 experiment he found it was fastest to judge an identical pair (AA) "same" by the name identity criterion; next fastest to classify a highly similar pair (Cc), and slowest to classify a physically dissimilar pair (Bb) as "same." It is reasonable to conclude that physical similarity cuts down processing time, and the saving is most likely to occur in the identification of the second member of the pair.

(b) The S analyzes and compares the members of the pair visually and decides they are visually identical. Extra time is then consumed in deciding that visual identity entails the pair being "same" by the semantic or acoustic criteria (or by name or rule criteria). This strategy would give an I time shorter than S time, since S pairs would require analysis on more criteria. (b) is also easier to reconcile with the practice effect reducing the differences between I RTs under different criteria, because the entailment relation need not be worked out afresh each time. Indeed a (b) interpretation would be strongly supported if further experiments showed that the I/S difference remained stable with practice, while the I visual/I semantic/I acoustic differences were reduced.

(2) *The difference between "same" and "different" responses.* The I RTs are shorter than the NC RTs in semantic and acoustic matching, but the difference is not significant in the visual matching task. A possible explanation can be found for this reversal, since sew-run, an NC pair, can be judged different by the visual criterion from a scrutiny of only the first letters, while sew-sew, an I pair, requires scrutiny of all the component letters. The longer scores in the AC category of the visual matching task fit in with this explanation, since AC pairs are also extremely similar visually (e.g., week-weak) and would need thorough inspection. The results suggest a self-terminating search in the "different" visual pairs. However, it is puzzling to find this does not hold in the acoustic matching task where road-street, an NC pair, could be judged different faster than rode-road could be judged "same" if a self-terminating search were employed. This inconsistency may be due to the unstable expectancies to which the vagaries of English spelling give rise; or it may be that some other factor such as codability is affecting "same"/"different" latencies here, as Bindra et al have recently suggested (1968), but this problem is outside the scope of this paper.

(3) *The difference between NC and AC or SC pairs.* It was expected that it might take longer to judge the AC pairs semantically different than to judge the NC pairs semantically different; and similarly NC and SC pairs might differ significantly in the acoustic matching task. No such differences were found, except in the visual matching task where the difference between NC and AC is attributable to the visual similarity of the members of the AC pair. The absence of such differences might indicate, for instance, that acoustic analysis is bypassed when meaning is the relevant criterion. Such a conclusion is plausible since this is what appears to happen during fast, skilled reading. Alternatively, the S may *derive* the sound of the words, but not compare them along this irrelevant criterion. Or again it may be that acoustic characteristics of the words are both derived and compared, but that similarity on an irrelevant criterion does not delay or interfere with a judgment of difference on the relevant one. Posner also found that it did not take longer to judge the pair Aa, which has a common name, physically different than to judge AB physically different, i.e., there was no interference from the irrelevant criterion.

(4) *The difference between NC RTs under different criteria.* All Ss show longer RTs for NC acoustic than for NC visual, and longest of all for NC semantic. In Posner's experiment the time to judge AB "different" also varied with the criterion. This finding is not necessarily evidence of a sequential hierarchical relationship, and the absence of interference from the irrelevant characteristics of the stimuli makes such a mechanism less likely. The longer times may well reflect the greater complexity of the processing or of the comparison required at that particular stage, rather than the presence of more stages. Although of course, both semantic and acoustic analysis must logically depend on prior visual identification, semantic matching might well omit the acoustic stage, and acoustic matching need not entail processing meanings. The results are consonant with either of the following interpretations:

(a) Parallel processing of the other characteristics following visual identification, and terminating as soon as a match is obtained on the relevant one. Different characteristics must then take different lengths of time either to analyze or to compare.

(b) Serial processing which can omit irrelevant analyses or ignore the results of irrelevant comparisons.

REFERENCES

BINDRA, D., DONDERI, D. C., & NISHISATO, S. Decision latencies of "same" and "different" judgments. *Percept. & Psychophys.*, 1968, Vol. 3, No. 2B, 121-126.

POSNER, M. I. Information reduction in the analysis of sequential tasks. *Psychol. Rev.*, 1964, 71, 491-504.

POSNER, M. I., & MITCHELL, R. F. Chronometric analysis of classification. *Psychol. Rev.*, 1967, 74, 392-409.

NOTE

1. Address: Institute of Experimental Psychology, University of Oxford, Oxford, England.

(Accepted for publication June 10, 1968.)