The present paper reports the results of two experiments designed to demonstrate the utility of the concept of stimulus-response compatibility\textsuperscript{2} in the development of a theory of perceptual-motor behavior.

A task involves compatible S–R relations to the extent that the ensemble of stimulus and response combinations comprising the task results in a high rate of information transfer. Admittedly, degree of compatibility can be defined in terms of operations other than those used to secure a measure of information, for example, it could be specified in terms of measures of speed or accuracy. However, the present writers prefer the preceding definition because of the theoretical interpretation that they wish to give to compatibility effects. This interpretation makes use of the idea of a hypothetical process of information transformation or recoding in the course of a perceptual-motor activity, and assumes that the degree of compatibility is at a maximum when recoding processes are at a minimum. The concept of compatibility can be extended to cover relations between concurrent stimulus activities, such as take place during simultaneous listening and looking, as well as to relations between concurrent motor responses. However, the present paper will be limited to a consideration of stimulus-response compatibility effects in which the relevant information in the stimulus source is that generated by changes in its spatial characteristics, and the relevant aspect of a response is its direction of movement.

One of the earliest studies of the behavioral effects of changes in the spatial correspondence of S–R relations is the well-known experiment of Stratton (12) on vision without inversion of the retinal image. Recently the majority of studies of such effects have dealt with the spatial relations between machine controls and remote visual displays that are connected to them by mechanical or electrical means (1, 3, 5, 9, 13). These relations are important for human engineering. They have also become a matter of considerable interest for learning theory, in part because of Gagné, Foster, and Baker’s (2) proposal that a reversal of S–R relations is the only condition leading to negative transfer effects in perceptual-motor learning. However, studies in both of these areas have consistently dealt with only one aspect of the compatibility problem. Investigations in the human-engineering area usually have compared different ways of displaying information when the S responds with a single type
of control, or else have examined the effectiveness of several kinds of controls or control motions when used with a single display. Investigations of transfer effects have typically considered only reversals of the stimulus-response relations within one S–R ensemble, such as changes in the direction of motion of the control on a particular psychomotor test (8), without specifying the location of the original task along a dimension of S–R compatibility. Only two studies (7, 11) have systematically investigated the effects of varying both stimulus sets and response sets. Such an approach is indicated by the concept of compatibility.

Information theory provides a convenient formulation of compatibility in terms of information coding. Shannon (10) and others have pointed out, in the case of physical communication channels, that information can be transmitted at a rate approaching channel capacity only if messages are optimally encoded for the particular channel being used. In such systems optimum encoding requires that messages be expressed in a suitable form and that their probability constraints be matched to the physical constraints of the channel. By analogy, it seems reasonable to hypothesize that man’s performance of a perceptual-motor task should be most efficient when the task necessitates a minimum amount of information transformation (encoding and/or decoding), in other words, when the information generated by successive stimulus events is appropriate to the set of responses that must be made in the task, or conversely, when the set of responses is appropriately matched to the stimulus source.

A few of the writers who have discussed the application of information theory to psychology have considered coding problems, such as the possibility of recoding as an aid in remembering. However, interest has been limited to stimulus coding (6).

Since the coding problem is central to the topic of compatibility, it will be worth while to consider briefly the available ways of choosing the sets of stimuli and the sets of responses employed in a visual-motor task. We shall consider only the case of a finite series of discrete code symbols. A stimulus code must utilize one or more stimulus dimensions such as the intensity, the wave length, the duration, or the extent of visible light. A unidimensional code is one in which the required number of unique code symbols is provided by the selection of points along a single stimulus dimension. A multidimensional code is one in which points are selected along two or more stimulus dimensions and each code symbol is defined by specifying a unique combination of several stimulus characteristics. In either case, several code symbols can be combined to form a larger multisymbol series.

A response code can be devised by similar procedures. A unidimensional response code utilizes a single effector member and a specified number of points along one of the physical dimensions of a particular response continuum, such as the direction, the force, or the duration of a movement. A multidimensional response code utilizes the complex response of a single effector member, such as the motion of the arm in positioning of a three-dimensional control, or the responses of several different effectors acting concurrently, such as the simultaneous responses of hands and feet in moving a set of conventional aircraft controls. Both the set of stimuli and the set of responses required in a particular task can thus be specified as

---

3 If an incompatible S–R relation exists, then a considerable amount of recoding of information (such as the looking for meaningful associations, etc.) may be highly desirable.
points lying in the multidimensional space comprising the set of all possible discriminable stimuli and motor responses.

Garner and Hake (4) have pointed out that in order to attain maximum average information transmission per symbol, stimulus points should be selected with reference to their position along a scale of equal discriminability. A similar principle undoubtedly holds for the selection of a set of responses. Discriminability, defined without reference to response rate, establishes an upper limit to the maximum average information that can be generated by the selection of particular stimulus or response points from a set of points lying along a specified dimension. The problem of S–R compatibility remains, however, even when sets of stimuli and responses are optimally chosen with respect to discriminability. It concerns the question of how the correspondence of the sets of points in the stimulus and the response symbol spaces affects information transfer.

**EXPERIMENT 1**

The first experiment was planned to test the hypothesis that, in tasks requiring directional motor responses to spatial stimulus patterns, effective performance depends to a large extent upon the unique characteristics of S–R ensembles rather than on specific aspects of particular stimulus or response sets. Three sets of stimuli (S<sub>a</sub>, S<sub>b</sub>, S<sub>c</sub>) each comprising eight easily discriminable light patterns, and three sets of eight responses (R<sub>a</sub>, R<sub>b</sub>, R<sub>c</sub>) were studied in the nine combinations formed by combining each stimulus set with each response set. Within each of these nine S–R ensembles, the pairings of stimuli and responses were those that the Es judged would be expected by most Ss, i.e., would agree with population stereotypes. These sets of stimuli and responses were chosen as abstractions of commonly used spatial patterns of stimuli and responses, i.e., as familiar ways of representing points on a two-dimensional surface.

A further consideration in selecting the sets of stimuli and responses was that three of the nine S–R ensembles (S<sub>a</sub>–R<sub>a</sub>, S<sub>b</sub>–R<sub>b</sub>, and S<sub>c</sub>–R<sub>c</sub>) should represent "corresponding" permutations and combinations. Correspondence, in this case, was judged by the Es on the basis of the direct physical similarity of the two patterns. The experiment therefore provides an incidental test of how accurately S–R compatibility can be predicted from a consideration of the correspondence, with respect to the number of coding dimensions employed, between the two sets of points employed in stimulus and response coding.

**Method**

**Apparatus.**—The response required of Ss was to move a stylus quickly in the direction indicated by a stimulus light or lights. The three stimulus and the three response panels used in the study are shown in Fig. 1 and a drawing of the S's position, with the S<sub>a</sub>–R<sub>c</sub> combination in place, is shown in Fig. 2.

Stimulus Set A and Response Set A each consisted of eight permutations of direction from a central reference point. The stimulus panel contained eight lights forming the outline of a circle. The response panel contained eight pathways radiating from a central point like the spokes of a wheel. Each light and each pathway were separated from their neighbors by an angle of 45°. These stimulus and response patterns are characteristic of those provided by a pictorial azimuth or bearing display and an aircraft-type control stick.

Stimulus Panel B consisted of four lights separated by 90°. It provided four single-light positions, plus the four two-light combinations formed by adjacent pairs of lights. Response Panel B consisted of four pathways originating at a central point and separated by 90° intervals. Each path, as seen in Fig. 1, branched in a T and per-
mitted Ss to move from the center point to one of eight terminal positions. The four corner points of the response panel could be reached by two alternative pathways; for example, the upper-right corner could be reached either by a right-up sequence or an up-right sequence. The Ss were told that these were equivalent responses. Response Panel B, therefore, involved the choice of a single directional movement, or of a sequence of two successive movements, and permitted Ss to terminate their responses in one of eight end states.

Stimulus Panel C contained a pair of horizontally separated lights and a pair of vertically separated lights. The set of eight stimulus conditions that it provided included the four possibilities that a single light would come on, plus the four two-light combinations that could be produced by the simultaneous presentation of one of each of the two pairs. Response Panel C permitted a left or right response of the left hand, an up or down response of the right hand, plus the four possible combined movements of the two hands. Stimulus Panel C corresponds in general to two separate single-scale instruments and Response Panel C provides a set of responses similar to those present whenever two separate hand controls are used.

At the center of each response panel was a %\text{in.} diameter metal disc, surrounded by a thin nonconductive ring. Reaction time was measured as the time taken by S to move the stylus off this metal button.

The stimulus panels were mounted at a 60° angle to the horizontal. Response panels were 30° to the horizontal. The Ss worked from a seated position. The stimulus panels were 15° downward and 28 in. from their eyes, and the response panels were at a convenient location in front of them. The Ss watched the stimulus panels and seldom looked at their hands or the response panels.

The E's station contained separate selector switches for the eight stimulus combinations and a single noiseless activation switch, which turned on the selected light(s) and simultaneously started a 1/100-sec. timer. The apparatus was designed so that different stimulus and response panels could be substituted at S's station without any change at E's station. The E sat where he could observe S's movements and could record errors as well as reaction time.

Fig. 2. The S's station showing the correct two-handed response on Panel C to the lower right-hand light of Stimulus Set A

![Diagram of the three stimulus panels (S_A, S_B, and S_C) and the three response panels (R_A, R_B, and R_C) used in Exp. I. The Ss held a metal stylus on the circular button in the center of the response panel.](image)
Subjects.—The Ss consisted of 72 airmen at Lockbourne Air Force Base, selected on the basis of two-choice reaction-time measures. Their ages ranged from 18 to 29 years. All were right handed. They were excused from drill during the time spent as Ss and appeared to be well motivated.

The 72 Ss were selected from a larger group of 153 men so as to form eight equal-sized strata which had homogeneous within-group average two-choice reaction-time scores. Each of the nine experimental groups was formed by drawing one S at random from each stratum. Each group of eight Ss was then tested under one of the nine S-R combinations.

Procedure.—The instructions for each of the groups were similar. Those for the S_a-R_a group were as follows:

"Here is a stimulus panel of eight lights and a response panel in which you can move this stylus to one of eight places. Hold the stylus in your right hand. When I say ‘center’ place it on this center disc. I shall then say ‘ready’ and a few seconds later one of the lights will come on. If this light (point) should come on, move the stylus straight up. If this light should come on, move the stylus quickly to this position (indicate upper-right corner). If you start in the wrong direction, correct your movement as soon as possible. Do not try to guess which light will come on as they will be presented in a random order. Work for both speed and accuracy since both reaction time and errors will be recorded."

Each S was given 20 practice trials on his particular S-R combination, followed by 40 test trials. The order of stimuli was randomized, with the restriction of equal frequency for all stimuli at the end of each series, and a further restriction against runs longer than two.

Results

The experiment provides three measures of the effectiveness of each of the S-R ensembles: (a) reaction time, (b) percentage of responses that were errors, and (c) average information lost per stimulus. The means for all three criterion measures are indicated in Fig. 3.
Reaction time.—In scoring the two-handed responses that were made on Panel C, reaction time was taken as the average of the times for the two hands. This procedure is justified by the finding that in two-handed responses the times for the right and left hands agree very closely. The average correlation of the two measures was .96 for the three groups that used Response Panel C, and the difference in the mean reaction time for the two hands was only .004 sec. However, for all one-handed responses on Panel C, the mean difference in reaction time between the two hands was .11 sec. in favor of the right hand, and 23 out of 24 Ss using Panel C were faster when using the right hand. In summary, when Ss had to move one hand alone, the right hand, which moved away from or toward the body, was significantly faster (by about .108 sec.) than the left hand, which moved to the right or the left, but when both hands had to be moved together, they had similar reaction times. The time for two-handed responses was approximately the same as the mean for one-hand responses by the left hand.

The means shown in Fig. 3 are for all stimuli in a set combined. The times for movements that were made in the wrong direction (errors) are combined with the data for correct responses, since the mean reaction times for erroneous and for correct responses did not differ significantly.

The mean reaction-time data for the different Ss were tested for homogeneity of variance by Bartlett’s test and no significant departure from homogeneity was found, even though variance tended to increase somewhat as the mean reaction time increased (the range of SD’s was from .036 sec. for the S_a–R_a ensemble to .097 sec. for the S_c–R_a combination). A double-classification analysis of variance for matched groups was then carried out (see Table 1). The most important finding is the highly significant interaction effect. The variance that can be attributed to interaction is very much larger than the variance attributable to the primary effects of either stimulus or response sets alone.

The differences in the means for the primary effects are interpreted as significantly different from chance, as indicated by the F ratios shown in Table I for which the residual term is used as the estimate of error. However, no reliable generalizations about these arbitrarily selected stimulus or response codes can be made to situations in which comparisons are made among different sets of stimuli or different sets of responses.

For every stimulus set there was a different best-response set, and for every response set there was a different best-stimulus set. For example, Response Set A led to the shortest mean reaction time in combination with Stimulus Set A, but to the longest reaction time in combination with Stimulus Set C. The difference of almost .4 sec. is 21 times as large as the estimate of the standard error of the difference. The three “best” combinations are those that were predicted by Es on the basis of the correspondence of the spatial codes.

<table>
<thead>
<tr>
<th>TABLE 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis of Variance for the Reaction-Time Data</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matched Ss</td>
<td>7</td>
<td>42896</td>
<td></td>
</tr>
<tr>
<td>Stimulus sets</td>
<td>2</td>
<td>86772</td>
<td>68.43*</td>
</tr>
<tr>
<td>Response sets</td>
<td>2</td>
<td>36352</td>
<td>28.67*</td>
</tr>
<tr>
<td>S–R interaction</td>
<td>4</td>
<td>170492</td>
<td>134.46*</td>
</tr>
<tr>
<td>Residual</td>
<td>56</td>
<td>1268</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>71</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Significant beyond the 1% level of confidence when tested against the residual as an estimate of error.
direct test of the significance of the difference between performance with corresponding and noncorresponding S–R ensembles is provided by a single-classification analysis of variance, which is equivalent to a conventional $t$ test between the means for matched Ss under the two conditions. Such an analysis was carried out and the difference was found to be highly significant ($p < .001$). The writers feel justified in predicting that if other sets of stimuli and sets of responses should be selected with due consideration to the correspondence of the stimulus and response codes, then it is highly probable that interaction effects would again account for a large portion of the variance and spatially corresponding codes would again give superior performance.

The reaction times for the $S_a-R_a$ and the $S_b-R_b$ combinations did not differ significantly, but both were superior to that for the $S_c-R_c$ combination. It might be hypothesized that a bidimensional stimulus and response coding scheme (two pairs of lights and a two-handed response) is inefficient. It seems appropriate, however, to suggest an alternative hypothesis, that the inferiority may have been due to failure on the part of Es to observe some principle of response-response compatibility in selecting the eight alternative movements constituting Response Set C. It should be mentioned that forward and back responses made with the right hand only averaged .41 sec. for the $S_c-R_c$ combination, which is comparable to the mean reaction times for the $S_a-R_a$ and $S_b-R_b$ combinations, where the right hand was also used.

Errors.—Approximately 10% of all responses were in error, an error being defined as an initial movement in the wrong direction. These error data agree in general with the time data in respect to the rankings assigned to the different S–R ensembles. The three "best" combinations within each row and column of Fig. 3 are the same for the two criteria. The difference between the best and the worst combinations appears to be relatively larger than was the case with the time scores. It is interesting to note that Response Set B, which in combination with its corresponding stimulus set ($S_b$) led to the fewest errors, resulted in the most errors when used in combination with another stimulus set ($S_c$).

The data for responses to separate stimuli revealed two important relations. The first is that those S–R ensembles having lowest mean error scores also tend to have the most uniform time scores from one stimulus to another. The second is that wherever there is marked variability within the responses to a set of stimuli, time and error scores tend to vary together, i.e., are positively correlated. For the four S–R ensembles with the greatest number of errors the rank correlations between time and error scores for the eight stimuli vary from .65 to .93.

Information lost.—The average information transmitted per stimulus was computed by Method I of Garner and Hake (4). The data for information lost, shown in Fig. 3, are the differences between the theoretical information in each stimulus event (3 bits) and the average information transmitted per stimulation.

The results of the information analysis agree closely with the total error frequencies and, in fact, add little to the grosser error analysis. The ranks of the nine groups are identical on the two criteria except for a reversal of the two last (worst) groups. This agreement is not surprising since the same type of error distribution, a piling up of errors in certain cells of the
matrix of transition probabilities, was found for all nine S–R combinations. It is not possible to compute a meaningful estimate of average rate of information transfer in this experiment because of the relatively long time delay between successive responses. However, short response times were associated with small loss of information both within the eight conditions of each S–R ensemble and between the means for the nine different S–R ensembles. Thus the results of the information or error analysis, considered in relation to the reaction-time analysis, provide empirical support for defining compatibility in terms of the average rate of information transmission.

Experimental II

An important question concerning the results of the preceding experiment is whether the differences between various S–R combinations are transitory or permanent. An extended learning study was carried out in an effort to answer this question.

Method

The most desirable way to have conducted this study would have been to practice all of the 72 Ss used in Exp. I over an extended period. However, this was not practical. Instead, a single new group was studied for 32 training sessions covering approximately 2½ months. The Ss practiced making a single set of eight responses (Response Set A) to each of the three different sets of eight stimuli employed in the previous study. As an alternative Ss could have been asked to make three sets of responses to a single set of stimuli, but this was not done because less initial habit interference is to be expected if the same responses are learned to three sets of stimuli, than if three sets of responses are learned to a single set of stimuli. Response Set A was chosen because there is no ambiguity in the scoring of these responses.

Six male students at Ohio State University were started on the training series but one dropped out after 20 sessions and his data are not reported (although they are comparable to the data that are reported). In order to maintain motivation Ss were scheduled in pairs (except for the last 12 sessions of the odd S). One S was given a series of 16 trials on one of the three sets of S–R combinations in an unbroken sequence, while the other S observed. The Ss then exchanged places. This procedure was continued until each S had been tested on each of the three stimulus sets. The sequence of work and rest and the order of trials on the three conditions were balanced. Each stimulus appeared twice in a random order within each run of 16 trials. Each S made a total of 48 responses per session.

Initial instructions and other procedures were the same as in the preceding experiment. At periodic intervals Ss were cautioned to try not to make errors. In addition to reaction time and errors, E recorded movement time, the time to traverse the selected pathway of Panel A. After each response Ss were told their reaction time and if they had made an error, this was pointed out.

On Sessions 27 to 30 inclusive, a secondary task, mental arithmetic, was carried on by Ss concurrently with the perceptual-motor task. The E read aloud a series of numbers at a predetermined rate, and S gave the successive differences between the last two numbers read by E. This secondary task was introduced to test the hypothesis that the least compatible S–R combination would show the most deterioration under conditions of additional load or stress. Standard conditions were resumed on Session 31. The experiment was terminated at the end of Session 32.

Results

The mean reaction-time and movement-time data for the five Ss are shown graphically in Fig. 4. Throughout all of the standard sessions performance was consistently best when Ss responded to Stimulus Set A. Stimulus Set B gave almost equally good results. Stimulus Set C was much the worst of the three. At no time during the 32 days did any of the Ss consistently respond as quickly to Stimulus Set C as they did to the other two sets of eight stimuli. The mean times for the three stimulus sets on Days 17 through 26 inclusive were as follows: $S_a .272$ sec.; $S_b .286$ sec.; $S_c .355$ sec.
Movement time in responding to Stimulus Set C was slightly but consistently slower than to Sets A and B. Averaged over Trials 17 through 26 these times were as follows: $S_a$, .059 sec.; $S_b$, .061 sec.; $S_c$, .067 sec.

The error data, averaged over all five Ss for successive blocks of time, are given in Table 2. It can be seen that the Ss never succeeded in eliminating all errors. In fact, errors occurred more frequently as training progressed, indicating perhaps that S emphasized speed at the expense of accuracy. Errors were made most frequently to Stimulus Set C throughout the training period. Thus both time and error data indicate consistently poorer performance on the $S_c-R_a$ condition.

On Days 27–30, when the load on Ss was increased by the addition of a secondary task, reaction times were substantially slower for all three S–R combinations. All five Ss continued to give the slowest reaction to Stimulus C; however, the relative differences between groups were much smaller than they had been previously. The frequency of arithmetic errors was approximately the same for all three groups, but there was a large differ-
ence between groups in the frequency of movement errors. The introduction of an additional task late in learning in this case apparently served to reduce the degree of readiness for the motor task, minimizing reaction-time differences, but increasing differences with respect to errors in the motor task. However, the data are not conclusive on these points and the test of the hypothesis regarding the effects of increased load is considered inconclusive.

Discussion

The results of the two experiments demonstrate clearly the importance of stimulus and response coding for the maximum rate of information transfer in a perceptual-motor task. It also is clear that some S-R compatibility effects are relatively unaffected by extended practice.

An interpretation of the basis for the relative permanence of these effects, which seems appropriate to the concept of compatibility, is one stated in terms of the capacity to learn to deal with sets of probabilities (probability learning). The response that a person makes to a particular stimulus event can be considered to be a function of two sets of probabilities; (a) the probabilities (uncertainties) appropriate to the situational constraints established by E's instructions, by the reinforcements experienced in the experimental situation, and by other aspects of the immediate situation; and (b) the more general and more stable expectancies or habits based on S's experiences in many other situations. It is suggested that extended training will nearly always lead to changes in the former but often will have relatively little effect on the latter.

This view holds that S's behavior, notwithstanding long experience in a particular situation, is never entirely relevant to the specific constraints of that particular situation. Instead, S responds as if additional possibilities were present.

This interpretation supports the view that stimulus and response sets are optimally matched when the resulting ensemble agrees closely with the basic habits or expectancies of individuals, i.e., with individual and with population stereotypes. It must be remembered, however, that the expectancies referred to are those which hold for the particular situation under study, and that population stereotypes should be determined with due regard to the total situation. For example, in the present experiment the Es determined, by preliminary trial, that Ss "preferred" to move toward rather than away from a stimulus light. However, it is known that in many stimulus tasks where the S controls the stimulus, the opposite motion relation is the expected one. When Ss have to learn to deal with the probabilities inherent in a particular situation, the correspondence of these specific expectancies to the more general expectancies of the individual with respect to that kind of situation is an important aspect of the learning task.

Further support for considering S-R compatibility to be a function of stimulus and response matching comes from the positive correlation between the two criteria, time and errors, used in evaluating performance. This positive correlation would be expected if additional information transformations, or re-encoding steps, were added to a communication system, since each transformation would be likely to add an additional time delay and to in-
crease the total probability of errors. In the human it is hypothesized that the fastest responses may be the most accurate because they involve S–R ensembles in which the transfer of information from stimulus to response is most direct, i.e., involves the minimum number of recoding steps. The concept of intervening information transformations does not attempt to explain how such recoding occurs within the nervous system, but it is in agreement with the subjective reports of Ss who maintained that it was difficult, in the case of the less compatible S–R ensembles used in the present study, to be prepared to respond to all of the eight stimulus possibilities.

SUMMARY

Experiment I was planned to test the hypothesis that information transfer in a perceptual-motor task is in large measure a function of the matching of sets of stimuli and sets of responses. Nine S–R ensembles, involving variations of the spatial patterns of stimuli and responses, were studied in an eight-choice situation, using groups of matched Ss.

The results, analyzed in terms of reaction time, errors, and information lost, support the hypothesis. They indicate that it is not permissible to conclude that any particular set of stimuli, or set of responses, will provide a high rate of information transfer; it is the ensemble of S–R combinations that must be considered.

Experiment II was planned to test the permanence of three selected S–R compatibility effects. Five Ss were trained for 32 days to make a particular set of responses to each of three sets of stimuli. Differences in reaction time, movement time, and frequency of errors in responding to the three sets of stimuli persisted over the 32 days.

The results are interpreted in terms of probability learning and the necessity for (hypothetical) information transformation or re-encoding steps. It appears that it is very difficult for Ss to learn to deal effectively with the information (uncertainties) characteristic of a specific situation, if these uncertainties are different from the more general set of probabilities which have been learned in similar life situations.

REFERENCES


(Received for early publication May 22, 1953)