4 Studies of Similarity

Amos Tversky
Itamar Gati
Hebrew University, Jerusalem

Any event in the history of the organism is, in a sense, unique. Consequently, recognition, learning, and judgment presuppose an ability to categorize stimuli and classify situations by similarity. As Quine (1969) puts it: "There is nothing more basic to thought and language than our sense of similarity; our sorting of things into kinds [p. 116]." Indeed, the notion of similarity – that appears under such different names as proximity, resemblance, communality, representativeness, and psychological distance – is fundamental to theories of perception, learning, and judgment. This chapter outlines a new theoretical analysis of similarity and investigates some of its empirical consequences.

The theoretical analysis of similarity relations has been dominated by geometric models. Such models represent each object as a point in some coordinate space so that the metric distances between the points reflect the observed similarities between the respective objects. In general, the space is assumed to be Euclidean, and the purpose of the analysis is to embed the objects in a space of minimum dimensionality on the basis of the observed similarities, see Shepard (1974).

In a recent paper (Tversky, 1977), the first author challenged the dimensional-metric assumptions that underlie the geometric approach to similarity and developed an alternative feature-theoretical approach to the analysis of similarity relations. In this approach, each object a is characterized by a set of features, denoted A, and the observed similarity of a to b, denoted $s(a, b)$, is expressed as a function of their common and distinctive features (see Fig. 4.1). That is, the observed similarity $s(a, b)$ is expressed as a function of three arguments: $A \cap B$, the features shared by a and b; $A - B$, the features of a that are not shared by b; $B - A$, the features of b that are not shared by a. Thus the similarity between

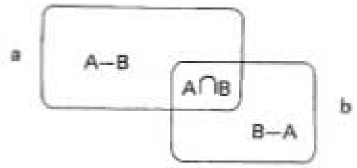

FIG. 4.1. A graphical illustration of the relation between two feature sets.

objects is expressed as a feature-matching function (i.e., a function that measures the degree to which two sets of features match each other) rather than as the metric distance between points in a coordinate space.

The theory is based on a set of qualitative assumptions about the observed similarity ordering. They yield an interval similarity scale S, which preserves the observed similarity order [i.e., $S(a, b) > S(c, d)$ iff $s(a, b) > s(c, d)$], and a scale f, defined on the relevant feature space such that

$$S(a, b) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A) \quad \text{where } \theta, \alpha, \beta \geqslant 0. \quad (1)$$

According to this form, called the *contrast model*, the similarity of a to b is described as a linear combination (or a contrast) of the measures of their common and distinctive features. Naturally, similarity increases with the measure of the common features and decreases with the measure of the distinctive features.

The contrast model does not define a unique index of similarity but rather a family of similarity indices defined by the values of the parameters θ, α, and β. For Example, if $\theta = 1$, and $\alpha = \beta = 0$, then $S(a, b) = f(A \cap B)$; that is, similarity equals the measure of the common features. On the other hand, if $\theta = 0$, and $\alpha = \beta = 1$, then $-S(a, b) = f(A - B) + f(B - A)$; that is, the dissimilarity of a to b equals the measure of the symmetric difference of the respective feature sets, see Restle (1961). Note that in the former case ($\theta = 1, \alpha = \beta = 0$), the similarity between objects is determined only by their common features, whereas in the latter case ($\theta = 0, \alpha = \beta = 1$), it is determined by their distinctive features only. The contrast model expresses similarity between objects as the weighted difference of the measures of their common and distinctive features, thereby allowing for a variety of similarity relations over the same set of objects.

The contrast model is formulated in terms of the parameters (θ, α, β) that characterize the task, and the scale f, which reflects the salience or prominence of the various features. Thus f measures the contribution of any particular (common or distinctive) feature to the similarity between objects. The scale value $f(A)$ associated with stimulus a is regarded, therefore, as a measure of the overall salience of that stimulus. The factors that contribute to the salience of a stimulus include: intensity, frequency, familiarity, good form, and informational content. The manner in which the scale f and the parameters (θ, α, β) depend on the context and the task are discussed in the following sections.

This chapter employs the contrast model to analyze the following three problems: the relation between judgments of similarity and difference; the nature of

asymmetric similarities; and the effects of context on similarity. All three problems concern changes in similarity induced, respectively, by the formulation of the *task* (as judgment of similarity or as judgment of difference), the *direction* of comparison, and the effective *context* (i.e., the set of objects under consideration).

To account for the effects of these manipulations within the present theoretical framework, we introduce several hypotheses that relate focus of attention to the experimental task. In particular, it is assumed that people attend more to common features in judgments of similarity than in judgments of difference, that people attend more to the subject than to the referent of the comparison, and that people attend primarily to features that have classificatory significance.

These hypotheses are formulated in terms of the contrast model and are tested in several experimental studies of similarity. For a more comprehensive treatment of the contrast model and a review of relevant data (including the present studies), see Tversky (1977).

SIMILARITY VERSUS DIFFERENCE

What is the relation between judgments of similarity and judgements of difference? Some authors emphasized that the two judgments are conceptually independent; others have treated them as perfectly correlated. The data appear to support the latter view. For example, Hosman and Kuennapas (1972) obtained independent judgments of similarity and difference for all pairs of lower-case letters on a scale from 0 to 100. The product-moment correlation between the judgments was −.98, and the slope of the regression line was −.91. We also collected judgments of similarity and difference for 21 pairs of countries using a 20-point rating scale. The product moment correlation between the ratings was again −.98. The near-perfect negative correlation between similarity and difference, however, does not always hold.

In applying the contrast model to judgments of similarity and of difference, it is reasonable to assume that enlarging the measure of the common features increases similarity and decreases difference, whereas enlarging the measure of the distinctive features decreases similarity and increases difference. More formally, let $s(a, b)$ and $d(a, b)$ denote ordinal measures of similarity and difference, respectively. Thus $s(a, b)$ is expected to increase with $f(A \cap B)$ and to decrease with $f(A - B)$ and with $f(B - A)$, whereas $d(a, b)$ is expected to decrease with $f(A \cap B)$ and to increase with $f(A - B)$ and with $f(B - A)$.

The relative weight assigned to the common and the distinctive features may differ in the two judgments because of a change in focus. In the assessment of similarity between stimuli, the subject may attend more to their common features, whereas in the assessment of difference between stimuli, the subject may attend more to their distinctive features. Stated differently, the instruction to consider similarity may lead the subject to focus primarily on the features that contribute to the similarity of the stimuli, whereas the instruction to consider difference may lead the subject to focus primarily on the features that contribute to the difference between the stimuli. Consequently, the relative weight of the common features is expected to be greater in the assessment of similarity than in the assessment of difference.

To investigate the consequences of this focusing hypothesis, suppose that both similarity and difference measures satisfy the contrast model with opposite signs but with different weights. Furthermore, suppose for simplicity that both measures are symmetric. Hence, under the contrast model, there exist non-negative constants θ and λ such that

$$s(a, b) > s(c, e) \quad \text{iff} \quad \theta f(A \cap B) - f(A - B) - f(B - A) > \theta f(C \cap E) - f(C - E) - f(E - C), \qquad (2)$$

and

$$d(a, b) > d(c, e) \quad \text{iff} \quad f(A - B) + f(B - A) - \lambda f(A \cap B) > f(C - E) + f(E - C) - \lambda f(C \cap E) \qquad (3)$$

The weights associated with the distinctive features can be set equal to 1 in the symmetric case with no loss of generality. Hence, θ and λ reflect the *relative* weight of the common features in the assessment of similarity and difference, respectively.

Note that if θ is very large, then the similarity ordering is essentially determined by the common features. On the other hand, if λ is very small, then the difference ordering is determined primarily by the distinctive features. Consequently, both $s(a, b) > s(c, e)$ and $d(a, b) > d(c, e)$ may be obtained whenever

$$f(A \cap B) > f(C \cap E) \quad \text{and} \quad f(A - B) + f(B - A) > f(C - E) + f(E - C). \qquad (4)$$

That is, if the common features are weighed more heavily in judgments of similarity than in judgments of difference, then a pair of objects with many common and many distinctive features may be perceived as both more similar and more different than another pair of objects with fewer common and fewer distinctive features.

Study 1: Similarity Versus Difference

All subjects that took part in the experiments reported in this chapter were undergraduate students majoring in the social sciences from the Hebrew University in Jerusalem and the Ben-Gurion University in Beer-Sheba. They participated in the studies as part of the requirements for a psychology course. The material was presented in booklets and administered in the classroom. The instructions were

printed in the booklet and also read aloud by the experimenter. The different forms of each booklet were assigned randomly to different subjects.

Twenty sets of four countries were constructed. Each set included two pairs of countries: a prominent pair and a nonprominent pair. The prominent pairs consisted of countries that were well known to the subjects (e.g., U.S.A.–U.S.S.R.). The nonprominent pairs consisted of countries that were known to our subjects but not as well as the prominent pairs (e.g., Paraguay–Ecuador). This assumption was verified in a pilot study in which 50 subjects were presented with all 20 quadruples of countries and asked to indicate which of the two pairs include countries that are more prominent, or better known. For each quadruple, over 85% of the subjects ordered the pairs in accord with our a priori ordering. All 20 sets of countries are displayed in Table 4.1.

Two groups of 30 subjects each participated in the main study. All subjects were presented with the same 20 sets in the same order. The pairs within each set were arranged so that the prominent pairs appeared an equal number of times on the left and on the right. One group of subjects – the similarity group – selected between the two pairs of each set the pair of countries that are more

TABLE 4.1
Percentage of Subjects That Selected the Prominent Pair in the Similarity Group (Π_s) and in the Difference Group (Π_d)

	Prominent Pairs	*Nonprominent Pairs*	Π_s	Π_d	$\Pi_s + \Pi_d$
1	W. Germany–E. Germany	Ceylon–Nepal	66.7	70.0	136.7
2	Lebanon–Jordan	Upper Volta–Tanzania	69.0	43.3	112.3
3	Canada–U.S.A.	Bulgaria–Albania	80.0	16.7	96.7
4	Belgium–Holland	Peru–Costa Rica	78.6	21.4	100.0
5	Switzerland–Denmark	Pakistan–Mongolia	55.2	28.6	83.8
6	Syria–Iraq	Liberia–Kenya	63.3	28.6	91.9
7	U.S.S.R.–U.S.A.	Paraguay–Ecuador	20.0	100.0	120.0
8	Sweden–Norway	Thailand–Burma	69.0	40.7	109.7
9	Turkey–Greece	Bolivia–Honduras	51.7	86.7	138.4
10	Austria–Switzerland	Zaire–Madagascar	79.3	24.1	103.4
11	Italy–France	Bahrain–Yemen	44.8	70.0	114.8
12	China–Japan	Guatemala–Costa Rica	40.0	93.1	133.1
13	S. Korea–N. Korea	Nigeria–Zaire	63.3	60.0	123.3
14	Uganda–Libya	Paraguay–Ecuador	23.3	65.5	88.8
15	Australia–S. Africa	Iceland–New Zealand	57.1	60.0	117.1
16	Poland–Czechoslovakia	Colombia–Honduras	82.8	37.0	119.8
17	Portugal–Spain	Tunis–Morocco	55.2	73.3	128.5
18	Vatican–Luxembourg	Andorra–San Marino	50.0	85.7	135.7
19	England–Ireland	Pakistan–Mongolia	80.0	58.6	138.6
20	Norway–Denmark	Indonesia–Philippines	51.7	25.0	76.7
	Average		59.1	54.4	113.5

similar. The second group of subjects – the difference group – selected between the two pairs in each set the pair of countries that are more different.

Let Π_s and Π_d denote, respectively, the percentage of subjects who selected the prominent pair in the similarity task and in the difference task. (Throughout this chapter, percentages were computed relative to the number of subjects who responded to each problem, which was occasionally smaller than the total number of subjects.) These values are presented in Table 4.1 for all sets. If similarity and difference are complementary (i.e., $\theta = \lambda$), then the sum $\Pi_s + \Pi_d$ should equal 100 for all pairs. On the other hand, if $\theta > \lambda$, then this sum should exceed 100. The average value of $\Pi_s + \Pi_d$ across all subjects and sets is 113.5, which is significantly greater than 100 ($t = 3.27$, $df = 59$, $p < .01$). Moreover, Table 4.1 shows that, on the average, the prominent pairs were selected more frequently than the nonprominent pairs both under similarity instructions (59.1%) and under difference instructions (54.4%), contrary to complementarity. These results demonstrate that the relative weight of the common and the distinctive features vary with the nature of the task and support the focusing hypothesis that people attend more to the common features in judgments of similarity than in judgments of difference.

2 DIRECTIONALITY AND ASYMMETRY

Symmetry has been regarded as an essential property of similarity relations. This
454
view underlies the geometric approach to the analysis of similarity, in which dissimilarity between objects is represented as a metric distance function. Although many types of proximity data, such as word associations or confusion probabilities, are often nonsymmetric, these asymmetries have been attributed to response biases. In this section, we demonstrate the presence of systematic asymmetries in direct judgments of similarity and argue that similarity should not be viewed as a symmetric relation. The observed asymmetries are explained in the contrast model by the relative salience of the stimuli and the directionality of the comparison.

Similarity judgments can be regarded as extensions of similarity statements (i.e., statements of the form "*a* is like *b*"). Such a statement is directional; it has a subject, *a*, and a referent, *b*, and it is not equivalent in general to the converse similarity statement "*b* is like *a*." In fact, the choice of a subject and a referent depends, in part at least, on the relative salience of the objects. We tend to select the more salient stimulus, or the prototype, as a referent and the less salient stimulus, or the variant, as a subject. Thus we say "the portrait resembles the person" rather than "the person resembles the portrait." We say "the son resembles the father" rather than "the father resembles the son," and we say "North Korea is like Red China" rather than "Red China is like North Korea."

As is demonstrated later, this asymmetry in the *choice* of similarity statements is associated with asymmetry in *judgments* of similarity. Thus the judged similarity

of North Korea to Red China exceeds the judged similarity of Red China to North Korea. In general, the direction of asymmetry is determined by the relative salience of the stimuli: The variant is more similar to the prototype than vice versa.

If $s(a, b)$ is interpreted as the degree to which *a* is similar to *b*, then *a* is the subject of the comparison and *b* is the referent. In such a task, one naturally focuses on the subject of the comparison. Hence, the features of the subject are weighted more heavily than the features of the referent (i.e., $\alpha > \beta$). Thus similarity is reduced more by the distinctive features of the subject than by the distinctive features of the referent. For example, a toy train is quite similar to a real train, because most features of the toy train are included in the real train. On the other hand, a real train is not as similar to a toy train, because many of the features of a real train are not included in the toy train.

It follows readily from the contrast model, with $\alpha > \beta$, that

$$\begin{aligned} s(a, b) &> s(b, a) \quad \text{iff} \\ \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A) &> \theta f(A \cap B) - \alpha f(B - A) - \beta f(A - B) \\ \text{iff } f(B - A) &> f(A - B). \end{aligned} \tag{5}$$

Thus $s(a, b) > s(b, a)$ whenever the distinctive features of *b* are more salient than the distinctive features of *a*, or whenever *b* is more prominent than *a*. Hence, the conjunction of the contrast model and the focusing hypothesis ($\alpha > \beta$) implies that the direction of asymmetry is determined by the relative salience of the stimuli so that the less salient stimulus is more similar to the salient stimulus than vice versa.

In the contrast model, $s(a, b) = s(b, a)$ if either $f(A - B) = f(B - A)$ or $\alpha = \beta$. That is, symmetry holds whenever the objects are equally salient, or whenever the comparison is nondirectional. To interpret the latter condition, compare the following two forms:

1. Assess the degree to which *a* and *b* are similar to each other.
2. Assess the degree to which *a* is similar to *b*.

In (1), the task is formulated in a nondirectional fashion, and there is no reason to emphasize one argument more than the other. Hence, it is expected that $\alpha = \beta$ and $s(a, b) = s(b, a)$. In (2), on the other hand, the task is directional, and hence the subject is likely to be the focus of attention rather than the referent. In this case, asymmetry is expected, provided the two stimuli are not equally salient. The directionality of the task and the differential salience of the stimuli, therefore, are necessary and sufficient for asymmetry.

In the following two studies, the directional asymmetry prediction, derived from the contrast model, is tested using semantic (i.e., countries) and perceptual (i.e., figures) stimuli. Both studies employ essentially the same design. Pairs of stimuli that differ in salience are used to test for the presence of asymmetry in the choice of similarity statements and in direct assessments of similarity.

Study 2: Similarity of Countries

In order to test the asymmetry prediction, we constructed 21 pairs of countries so that one element of the pair is considerably more prominent than the other (e.g., U.S.A.–Mexico, Belgium–Luxembourg). To validate this assumption, we presented all pairs to a group of 68 subjects and asked them to indicate in each pair the country they regard as more prominent. In all cases except one, more than two-thirds of the subjects agreed with our initial judgment. All 21 pairs of countries are displayed in Table 4.2, where the more prominent element of each pair is denoted by p and the less prominent by q.

Next, we tested the hypothesis that the more prominent element is generally chosen as the referent rather than as the subject of similarity statements. A group of 69 subjects was asked to choose which of the following two phrases they prefer to use: "p is similar to q," or "q is similar to p." The percentage of subjects that selected the latter form, in accord with our hypothesis, is displayed in Table 4.2 under the label II. It is evident from the table that in all cases the great majority of subjects selected the form in which the more prominent country serves as a referent.

455

TABLE 4.2
Average Similarities and Differences for 21 Pairs of Countries

	p	q	II	$s(p, q)$	$s(q, p)$	$d(p, q)$	$d(q, p)$
1	U.S.A.	Mexico	91.1	6.46	7.65	11.78	10.58
2	U.S.S.R.	Poland	98.6	15.12	15.18	6.37	7.30
3	China	Albania	94.1	8.69	9.16	14.56	12.16
4	U.S.A.	Israel	95.6	9.70	10.65	13.78	12.53
5	Japan	Philippines	94.2	12.37	11.95	7.74	5.50
6	U.S.A.	Canada	97.1	16.96	17.33	4.40	3.82
7	U.S.S.R.	Israel	91.1	3.41	3.69	18.41	17.25
8	England	Ireland	97.1	13.32	13.49	7.50	5.04
9	W. Germany	Austria	87.0	15.60	15.20	6.95	6.67
10	U.S.S.R.	France	82.4	5.21	5.03	15.70	15.00
11	Belgium	Luxembourg	95.6	15.54	16.14	4.80	3.93
12	U.S.A.	U.S.S.R.	65.7	5.84	6.20	16.65	16.11
13	China	N. Korea	95.6	13.13	14.22	8.20	7.48
14	India	Ceylon	97.1	13.91	13.88	5.51	7.32
15	U.S.A.	France	86.8	10.42	11.09	10.58	10.15
16	U.S.S.R.	Cuba	91.1	11.46	12.32	11.50	10.50
17	England	Jordan	98.5	4.97	6.52	15.81	14.95
18	France	Israel	86.8	7.48	7.34	12.20	11.88
19	U.S.A.	W. Germany	94.1	11.30	10.70	10.25	11.96
20	U.S.S.R.	Syria	98.5	6.61	8.51	12.92	11.60
21	France	Algeria	95.6	7.86	7.94	10.58	10.15

Sim
1–20
scal

To test the hypothesis that $s(q, p) > s(p, q)$, we instructed two groups of 77 subjects each to assess the similarity of each pair on a scale from 1 (no similarity) to 20 (maximal similarity). The two groups were presented with the same list of 21 pairs, and the only difference between the two groups was the order of the countries within each pair. For example, one group was asked to assess "the degree to which Red China is similar to North Korea," whereas the second group was asked to assess "the degree to which North Korea is similar to Red China." The lists were balanced so that the more prominent countries appeared about an equal number of times in the first and second position. The average ratings for each ordered pair, denoted $s(p, q)$ and $s(q, p)$ are displayed in Table 4.2. The average $s(q, p)$ was significantly higher than the average $s(p, q)$ across all subjects and pairs. A t-test for correlated samples yielded $t = 2.92$, $df = 20$, and $p < .01$. To obtain a statistical test based on individual data, we computed for each subject a directional asymmetry score, defined as the average similarity for comparisons with a prominent referent [i.e., $s(q, p)$ minus the average similarity for comparison with a prominent subject, i.e., $s(p, q)$]. The average difference (.42) was significantly positive: $t = 2.99$, $df = 153$, $p < .01$.

The foregoing study was repeated with judgments of difference instead of judgments of similarity. Two groups of 23 subjects each received the same list of 21 pairs, and the only difference between the groups, again, was the order of the countries within each pair. For example, one group was asked to assess "the degree to which the U.S.S.R. is different from Poland," whereas the second group was asked to assess "the degree to which Poland is different from the U.S.S.R." All subjects were asked to rate the difference on a scale from 1 (minimal difference) to 20 (maximal difference).

DIFF
1–20
scal

If judgments of difference follow the contrast model (with opposite signs) and the focusing hypothesis ($\alpha > \beta$) holds, then the prominent stimulus p is expected to differ from the less prominent stimulus q more than q differs from p [i.e., $d(p, q) > d(q, p)$]. The average judgments of difference for all ordered pairs are displayed in Table 4.2. The average $d(p, q)$ across all subjects and pairs was significantly higher than the average $d(q, p)$. A t-test for correlated samples yielded $t = 2.72$, $df = 20$, $p < .01$. Furthermore, the average difference between $d(p, q)$ and $d(q, p)$, computed as previously for each subject (.63), was significantly positive: $t = 2.24$, $df = 45$, $p < .05$. Hence, the predicted asymmetry was confirmed in direct judgments of both similarity and difference.

Study 3: Similarity of Figures

Two sets of eight pairs of geometric figures served as stimuli in the present study. In the first set, one figure in each pair, denoted p, had better form than the other, denoted q. In the second set, the two figures in each pair were roughly equivalent with respect to goodness of form, but one figure, denoted p, was richer or more complex than the other, denoted q. Examples of pairs of figures from each set are presented in Fig. 4.2.

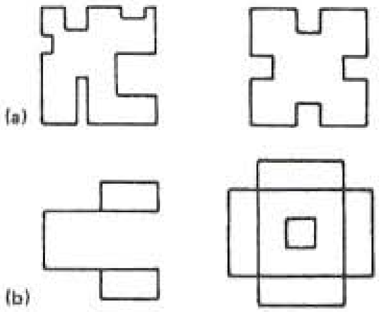

FIG. 4.2. Examples of pairs of figures used to test the prediction of asymmetry. (a) Example of a pair of figures (from Set 1) that differ in goodness of form. (b) Example of a pair of figures (from Set 2) that differ in complexity.

A: We hypothesized that both goodness of form and complexity contribute to the salience of geometric figures. Moreover, we expected a "good figure" to be more salient than a "bad figure," although the latter is generally more complex. For pairs of figures that do not vary much with respect to goodness of form, however, the more complex figure is expected to be more salient.

A group of 69 subjects received the entire list of 16 pairs of figures. The two elements of each pair were displayed side by side. For each pair, the subjects were asked to choose which of the following two statements they preferred to use: "the left figure is similar to the right figure," or "the right figure is similar to the left figure." The positions of the figures were randomized so that p and q appeared an equal number of times on the left and on the right. The proportion of subjects that selected the form "q is similar to p" exceeded 2/3 in all pairs except one. Evidently, the more salient figure (defined as previously) was generally chosen as the referent rather than as the standard.

To test for asymmetry in judgments of similarity, we presented two groups of 66 subjects each with the same 16 pairs of figures and asked the subjects to rate (on a 20-point scale) the degree to which the figure on the left is similar to the figure on the right. The two groups received identical booklets, except that the left and right positions of the figures in each pair were reversed. The data shows that the average $s(q, p)$ across all subjects and pairs was significantly higher than the average $s(p, q)$. A t-test for correlated samples yielded $t = 2.94$, $df = 15$, $p < .01$. Furthermore, in both sets the average difference between $s(q, p)$ and $s(p, q)$ computed as previously for each individual subject (.56) were significantly positive. In Set 1, $t = 2.96$, $df = 131$, $p < .01$, and in Set 2, $t = 2.79$, $df = 131$, $p < .01$.

The preceding two studies revealed the presence of systematic and significant asymmetries in judgments of similarity between countries and geometric figures. The results support the theoretical analysis based on the contrast model and the focusing hypothesis, according to which the features of the subject are weighted more heavily than the features of the referent. Essentially the same results were obtained by Rosch (1975) using a somewhat different design. In her studies, one

stimulus (the standard) was placed at the origin of a semicircular board, and the subject was instructed to place the second (variable) stimulus on the board so as "to represent his feeling of the distance between that stimulus and the one fixed at the origin." Rosch used three stimulus domains: color, line orientation, and number. In each domain, she paired prominent, or focal, stimuli with nonfocal stimuli. For example, a pure red was paired with an off-red, a vertical line was paired with a diagonal line, and a round number (e.g., 100) was paired with a nonround number (e.g., 103).

In all three domains, Rosch found that the measured distance between stimuli was smaller when the more prominent stimulus was fixed at the origin. That is, the similarity of the variant to the prototype was greater than the similarity of the prototype to the variant. Rosch also showed that when presented with sentence frames containing hedges such as "_____ is virtually _____," subjects generally placed the prototype in the second blank and the variant in the first. For example, subjects preferred the sentence "103 is virtually 100" to the sentence "100 is virtually 103."

In contrast to direct judgments of similarity, which have traditionally been viewed as symmetric, other measures of similarity such as confusion probability or association were known to be asymmetric. The observed asymmetries, however, were commonly attributed to a response bias. Without denying the important role of response biases, asymmetries in identification tasks occur even in situations to which a response bias interpretation does not apply (e.g., in studies where the subject indicates whether two presented stimuli are identical or not). Several experiments employing this paradigm obtained asymmetric confusion probabilities of the type predicted by the present analysis. For a discussion of these data and their implications, see Tversky (1977).

3 CONTEXT EFFECTS

The preceding two sections deal with the effects of the formulation of the task (as judgment of similarity or of difference) and of the direction of comparison (induced by the choice of subject and referent) on similarity. These manipulations were related to the parameters (θ, α, β) of the contrast model through the focusing hypothesis. The present section extends this hypothesis to describe the manner in which the measure of the feature space f varies with a change in context.

The scale f is generally not invariant with respect to changes in context or frame of reference. That is, the salience of features may vary widely depending on implicit or explicit instructions and on the object set under consideration. East Germany and West Germany, for example, may be viewed as highly similar from a geographical or cultural viewpoint and as quite dissimilar from a political viewpoint. Moreover, the two Germanys are likely to be viewed as more similar to each other in a context that includes many Asian and African countries than in a context that includes only European countries.

How does the salience of features vary with changes in the set of objects under consideration? We propose that the salience of features is determined, in part at least, by their diagnosticity (i.e., classificatory significance). A feature may acquire diagnostic value (and hence become more salient) in a particular context if it serves as a basis for classification in that particular context. The relations between similarity and diagnosticity are investigated in several studies that show how the similarity between a given pair of countries is varied by changing the context in which they are embedded.

Study 4. The Extension of Context

457

According to the preceding discussion, the diagnosticity of features is determined by the prevalence of the classifications that are based on them. Hence, features that are shared by all the objects under study are devoid of diagnostic value, because they cannot be used to classify these objects. However, when the context is extended by enlarging the object set, some features that had been shared by all objects in the original context may not be shared by all objects in the broader context. These features then acquire diagnostic value and increase the similarity of the objects that share them. Thus the similarity of a pair of objects in the original context is usually smaller than their similarity in the extended context.

To test this hypothesis, we constructed a list of pairs of countries with a common border and asked subjects to assess their similarity on a 20-point scale. Four sets of eight pairs were constructed. Set 1 contained eight pairs of American countries, Set 2 contained eight pairs of European countries, Set 3 contained four pairs from Set 1 and four pairs from Set 2, and Set 4 contained the remaining pairs from Sets 1 and 2. Each one of the four sets was presented to a different group of 30–36 subjects. The entire list of 16 pairs is displayed in Table 4.3.

Recall that the features "American" and "European" have no diagnostic value in Sets 1 and 2, although they both have diagnostic value in Sets 3 and 4. Consequently, the overall average similarity in the heterogeneous sets (3 and 4) is expected to be higher than the overall average similarity in the homogeneous sets (1 and 2). The average similarity for each pair of countries obtained in the homogeneous and the heterogeneous contexts, denoted s_o and s_e, respectively, are presented in Table 4.3. In the absence of context effects, the similarity for any pair of countries should be independent of the list in which it was presented. In contrast, the average difference between s_e and s_o (.57) is significantly positive: $t = 2.11$, $df = 15$, $p < .05$.

Similar results were obtained in an earlier study by Sjöberg (1972) who showed that the similarities between string instruments (banjo, violin, harp, electric guitar) were increased when a wind instrument (clarinet) was added to

TABLE 4.3
Average Similarities of Countries in Homogeneous (s_o) and Heterogeneous (s_e) Contexts

So HOMO HETERO

	Countries s_c	$s_o(a, b)$	$s_e(a, b)$
American countries	Panama–Costa Rica	12.30	13.29
	Argentina–Chile	13.17	14.36
	Canada–U.S.A.	16.10	15.86
	Paraguay–Bolivia	13.48	14.43
	Mexico–Guatemala	11.36	12.81
	Venezuela–Colombia	12.06	13.06
	Brazil–Uruguay	13.03	14.64
	Peru–Ecuador	13.52	14.61
European countries	England–Ireland	13.88	13.37
	Spain–Portugal	15.44	14.45
	Bulgaria–Greece	11.44	11.00
	Sweden–Norway	17.09	15.03
	France–W. Germany	10.88	11.81
	Yugoslavia–Austria	8.47	9.86
	Italy–Switzerland	10.03	11.14
	Belgium–Holland	15.39	17.06

this set. Hence, Sjöberg found that the similarity in the homogeneous pairs (i.e., pairs of string instruments) was increased when heterogeneous pairs (i.e., a string instrument and a wind instrument) were introduced into the list. Because the similarities in the homogeneous pairs, however, are greater than the similarities in the heterogeneous pairs, the above finding may be attributed, in part at least, to the common tendency of subjects to standardize the response scale (i.e., to produce the same average similarity for any set of comparisons).

Recall that in the present study all similarity assessments involve only homogeneous pairs (i.e., pairs of countries from the same continent sharing a common border). Unlike Sjöberg's (1972) study that extended the context by introducing heterogeneous pairs, our experiment extended the context by constructing heterogeneous lists composed of homogeneous pairs. Hence, the increase of similarity with the enlargement of context, observed in the present study, cannot be explained by the tendency to standardize the response scale.

Study 5. Similarity and Clustering

When faced with a set of stimuli, people often organize them in clusters to reduce information load and facilitate further processing. Clusters are typically selected in order to maximize the similarity of objects within the cluster and the dissimilarity of objects from different clusters. Clearly, the addition and/or deletion of

H: objects can alter the clustering of the remaining objects. We hypothesize that changes in clustering (induced by the replacement of objects) increase the diagnostic value of the features on which the new clusters are based and consequently the similarity of objects that share these features. Hence, we expect that changes in context which affect the clustering of objects will affect their similarity in the same manner.

The procedure employed to test this hypothesis (called the *diagnosticity hypothesis*) is best explained in terms of a concrete example, taken from the present study. Consider the two sets of four countries displayed in Fig. 4.3, which differ only in one of their elements (p or q).

The sets were constructed so that the natural clusterings of the countries are: p and c vs. a and b in Set 1; and b and q vs. c and a in Set 2. Indeed, these were the modal classifications of subjects who were asked to partition each quadruple into two pairs. In Set 1, 72% of the subjects partitioned the set into Moslem countries (Syria and Iran) vs. non-Moslem countries (England and Israel); whereas in Set 2, 84% of the subjects partitioned the set into European countries (England and France) vs. Middle-Eastern countries (Iran and Israel). Hence, the replacement of p by q changed the pairing of a: In Set 1, a was paired with b; whereas in Set 2, a was paired with c. The diagnosticity hypothesis implies that the change in clustering, induced by the substitution of the odd element (p or q), should produce a corresponding change in similarity. That is, the similarity of England to Israel should be greater in Set 1, where it is natural to group them together, than in Set 2 where it is not. Likewise, the similarity of Iran to Israel should be greater in Set 2, where they tend to be grouped together, than in Set 1 where they are not.

To investigate the relation between clustering and similarity, we constructed 20 pairs of sets of four countries of the form (a, b, c, p) and (a, b, c, q), whose elements are listed in Table 4.4. Two groups of 25 subjects each were presented with 20 sets of four countries and asked to partition each quadruple into two pairs. Each group received one of the two matched quadruples, displayed in a row in random order.

		a ISRAEL	
Set 1	b ENGLAND 37.5%	$\boxed{p}$ SYRIA 25%	c IRAN 37.5%
		a ISRAEL	
Set 2	b ENGLAND 24.2%	$\boxed{q}$ FRANCE 30.3%	c IRAN 45.5%

FIG. 4.3. An example of two matched sets of countries used to test the diagnosticity hypothesis. The percentage of subjects that ranked each country below (as most similar to the target) is presented under the country.

458

a, b, c, p
a, b, c, q

TABLE 4.4
Classification and Similarity Data for the Test of the Diagnosticity Hypothesis

% to a　　% to a　　%

	a	b	c	q	p	$b(p) - b(q)$	$c(q) - c(p)$	$D(p, q)$
1	U.S.S.R.	Poland	China	Hungary	India	6.1	24.2	66.7
2	England	Iceland	Belgium	Madagascar	Switzerland	10.4	−7.5	68.8
3	Bulgaria	Czechoslovakia	Yugoslavia	Poland	Greece	13.7	19.2	56.6
4	U.S.A.	Brazil	Japan	Argentina	China	11.2	30.2	78.3
5	Cyprus	Greece	Crete	Turkey	Malta	9.1	−6.1	63.2
6	Sweden	Finland	Holland	Iceland	Switzerland	6.5	6.9	44.1
7	Israel	England	Iran	France	Syria	37.5−24.2= 13.3	45.5−37.5= 8.0	87.5
8	Austria	Sweden	Hungary	Norway	Poland	3.0	15.2	60.0
9	Iran	Turkey	Kuwait	Pakistan	Iraq	−6.1	0.0	58.9
10	Japan	China	W. Germany	N. Korea	U.S.A.	24.2	6.1	66.9
11	Uganda	Libya	Zaire	Algeria	Angola	23.0	−1.0	48.8
12	England	France	Australia	Italy	New Zealand	36.4	15.2	73.3
13	Venezuela	Colombia	Iran	Brazil	Kuwait	0.3	31.5	60.7
14	Yugoslavia	Hungary	Greece	Poland	Turkey	9.1	9.1	76.8
15	Libya	Algeria	Syria	Tunis	Jordan	3.0	24.2	73.2
16	China	U.S.S.R.	India	U.S.A.	Indonesia	30.3	−3.0	42.2
17	France	W. Germany	Italy	England	Spain	−12.1	30.3	74.6
18	Cuba	Haiti	N. Korea	Jamaica	Albania	−9.1	0.0	35.9
19	Luxembourg	Belgium	Monaco	Holland	San Marino	30.3	6.1	52.2
20	Yugoslavia	Czechoslovakia	Austria	Poland	France	3.0	24.2	39.6

$\bar{X}$ =　　11%　　61.4%

Let $a_p(b, c)$ denote the percentage of subjects that paired a with b rather than with c when the odd element was p, etc. the difference $D(p, q) = a_p(b, c) - a_q(b, c)$, therefore, measures the effect of replacing q by p on the tendency to classify a with b rather than with c. The values of $D(p, q)$ for each one of the pairs is presented in the last column of Table 4.4. The results show that, in all cases, the replacement of q by p changed the pairing of a in the expected direction; the average difference is 61.4%.

Next, we presented two groups of 33 subjects each with 20 sets of four countries in the format displayed in Fig. 4.3. The subjects were asked to rank, in each quadruple, the three countries below (called the *choice set*) in terms of their similarity to the country on the top (called the *target*). Each group received exactly one quadruple from each pair. If the similarity of b to a, say, is independent of the choice set, then the proportion of subjects who ranked b rather than c as most similar to a should be independent of whether the third element in the choice set is p or q. For example, the proportion of subjects who ranked England rather than Iran as most similar to Israel should be the same whether the third element in the choice set is Syria or France. In contrast, the diagnosticity hypothesis predicts that the replacement of Syria (which is grouped with Iran)
459 by France (which is grouped with England) will affect the ranking of similarity so that the proportion of subjects that ranked England rather than Iran as most similar to Israel is greater in Set 1 than in Set 2.

Let $b(p)$ denote the percentage of subjects who ranked country b as most similar to a when the odd element in the choice set is p, etc. Recall that b is generally grouped with q, and c is generally grouped with p. The differences $b(p) - b(q)$ and $c(q) - c(p)$, therefore, measure the effects of the odd elements, p and q, on the similarity of b and c to the target a. The value of these differences for all pairs of quadruples are presented in Table 4.4. In the absence of context effects, the differences should equal 0, while under the diagnosticity hypothesis, the differences should be positive. In Fig. 4.3, for example, $b(p) - b(q) = 37.5 - 24.2 = 13.3$, and $c(q) - c(p) = 45.5 - 37.5 = 8$. The average difference across all pairs of quadruples was 11%, which is significantly positive: $t = 6.37$, $df = 19$, $p < .01$.

An additional test of the diagnosticity hypothesis was conducted using a slightly different design. As in the previous study, we constructed pairs of sets that differ in one element only (p or q). Furthermore, the sets were constructed so that b is likely to be grouped with q, and c is likely to be grouped with p. Two groups of 29 subjects were presented with all sets of five countries in the format displayed in Fig. 4.4. These subjects were asked to select, for each set, the country in the choice set below that is most similar to the two target countries above. Each group received exactly one set of five countries from each pair. Thus the present study differs from the previous one in that: (1) the target consists of a pair of countries (a_1 and a_2) rather than of a single country; and (2) the subjects were instructed to select an element of the choice set that is most similar to the target rather than to rank all elements of the choice set.

b-q

c-p

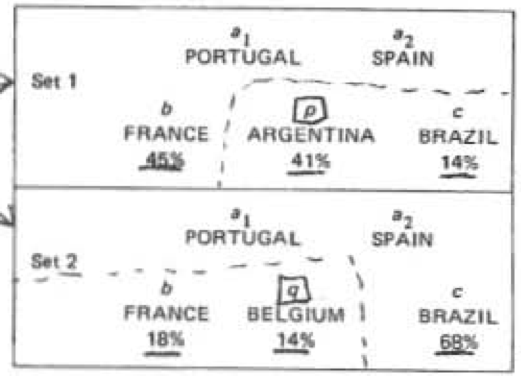

FIG. 4.4. Two sets of countries used to test the diagnosticity hypothesis. The percentage of subjects who selected each country (as most similar to the two target countries) is presented below the country.

The analysis follows the previous study. Specifically, let $b(p)$ denote the proportion of subjects who selected country b as most similar to the two target countries when the odd element in the choice set was p, etc. Hence, under the diagnosticity hypothesis, the differences $b(p) - b(q)$ and $c(q) - c(p)$ should both be positive, whereas under the assumption of context independence, both differences should equal 0. The values of these differences for all 12 pairs of sets are displayed in Table 4.5. The average difference across all pairs equals 10.9%, which is significantly positive: $t = 3.46$, $df = 11$, $p < .01$.

In Fig. 4.4, for example, France was selected, as most similar to Portugal and Spain, more frequently in Set 1 (where the natural grouping is: Brazil and Argentina vs. Portugal, Spain, and France) than in Set 2 (where the natural grouping is: Belgium and France vs. Portugal, Spain, and Brazil). Likewise, Brazil was selected, as most similar to Portugal and Spain, more frequently in Set 2 than in Set 1. Moreover, in this particular example, the replacement of p by q actually reversed the proximity order. In Set 1, France was selected more frequently than Brazil; in Set 2, Brazil was chosen more frequently than France.

There is considerable evidence that the grouping of objects is determined by the similarities among them. The preceding studies provide evidence for the converse (diagnosticity) hypothesis that the similarity of objects is modified by the manner in which they are grouped. Hence, similarity serves as a basis for the classification of objects, but it is also influenced by the adopted classification. The diagnosticity principle that underlies the latter process may provide a key to the understanding of the effects of context on similarity.

DISCUSSION

The investigations reported in this chapter were based on the contrast model according to which the similarity between objects is expressed as a linear combination of the measures of their common and distinctive features. The results provide support for the general hypothesis that the parameters of the contrast

TABLE 4.5
Similarity Data for the Test of the Diagnosticity Hypothesis

	a_1	a_2	b	c	p	q	$b(p) - b(q)$	$c(q) - c(p)$
1	China	U.S.S.R.	Poland	U.S.A.	England	Hungary	18.8	1.6
2	Portugal	Spain	France	Brazil	Argentina	Belgium	27.0	54.1
3	New Zealand	Australia	Japan	Canada	U.S.A.	Philippines	27.2	−12.4
4	Libya	Algeria	Syria	Uganda	Angola	Jordan	13.8	10.3
5	Australia	New Zealand	S. Africa	England	Ireland	Rhodesia	−0.1	13.8
6	Cyprus	Malta	Sicily	Crete	Greece	Italy	0.0	3.4
7	India	China	U.S.S.R.	Japan	Philippines	U.S.A.	−6.6	14.8
8	S. Africa	Rhodesia	Ethiopia	New Zealand	Canada	Zaire	33.4	5.9
9	Iraq	Syria	Lebanon	Libya	Algeria	Cyprus	9.6	20.3
10	U.S.A.	Canada	Mexico	England	Australia	Panama	6.0	13.8
11	Holland	Belgium	Denmark	France	Italy	Sweden	5.4	−8.3
12	Australia	England	Cyprus	U.S.A.	U.S.S.R.	Greece	5.4	5.1

model are sensitive to manipulations that make the subject focus on certain features rather than on others. Consequently, similarities are not invariant with respect to the marking of the attribute (similarity vs. difference), the directionality of the comparison [$s(a, b)$ vs. $s(b, a)$], and the context (i.e., the set of objects under consideration). In accord with the focusing hypothesis, Study 1 shows that the relative weight attached to the common features is greater in judgments of similarity than in judgments of difference (i.e., $\theta > \lambda$). Studies 2 and 3 show that people attach greater weight to the subject of a comparison than to its referent (i.e., $\alpha > \beta$). Studies 4 and 5 show that the salience of features is determined, in part, by their diagnosticity (i.e., by their classificatory significance).

What are the implications of the present findings to the analysis and representation of similarity relations? First, they indicate that there is no unitary concept of similarity that is applicable to all different experimental procedures used to elicit proximity data. Rather, it appears that there is a wide variety of similarity relations (defined on the same domain) that differ in the weights attached to the various arguments of the feature-matching function. Experimental manipulations that call attention to the common features, for example, are likely to increase the weight assigned to these features. Likewise, experimental manipulations (e.g., the introduction of a standard) that emphasize the directionality of the comparison are likely to produce asymmetry. Finally, changes in the natural clustering of the objects under study are likely to highlight those features on which the clusters are based.

Although the violations of complementarity, symmetry, and context independence are statistically significant and experimentally reliable in the sense that they were observed with different stimuli under different experimental conditions, the effects are relatively small. Consequently, complementarity, symmetry, or context independence may provide good first approximations to similarity data. Scaling models that are based on these assumptions, therefore, should not be rejected off-hand. A Euclidean map may provide a very useful and parsimonious description of complex data, even though its underlying assumptions (e.g., symmetry, or the triangle inequality) may be incorrect. At the same time, one should not treat such a representation, useful as it might be, as an adequate psychological theory of similarity. An analogy to the measurement of physical distance illustrates the point. The knowledge that the earth is round does not prevent surveyors from using plane geometry to calculate small distances on the surface of the earth. The fact that such measurements often provide excellent approximations to the data, however, should not be taken as evidence for the flat-earth model.

Finally, two major objections have been raised against the usage of the concept of similarity [see e.g., Goodman (1972)]. First, it has been argued that similarity is relative and variable: Objects can be viewed as either similar or dif-

EUCLIDEAN $\{$ $\delta_{ii} = 0 \quad \forall i$

$\delta_{ij} = \delta_{ji} \quad \forall i \neq j$

ferent depending on the context and frame of reference. Second, similarity often does not account for our inductive practice but rather is inferred from it; hence, the concept of similarity lacks explanatory power.

Although both objections have some merit, they do not render the concept of similarity empirically uninteresting or theoretically useless. The present studies, like those of Shepard (1964) and Torgerson (1965), show that similarity is indeed relative and variable, but it varies in a lawful manner. A comprehensive theory, therefore, should describe not only how similarity is assessed in a given situation but also how it varies with a change of context. The theoretical development, outlined in this chapter, provides a framework for the analysis of this process.

As for the explanatory function of similarity, it should be noted that similarity plays a dual role in theories of knowledge and behavior: It is employed as an independent variable to explain inductive practices such as concept formation, classification, and generalization; but it is also used as a dependent variable to be explained in terms of other factors. Indeed, similarity is as much a summary of past experience as a guide for future behavior. We expect similar things to behave in the same way, but we also view things as similar because they behave in the same way. Hence, similarities are constantly updated by experience to reflect our ever-changing picture of the world.

REFERENCES

Goodman, N. Seven strictures on similarity. In N. Goodman, *Problems and projects.* New York: Bobbs-Merril, 1972.

Hosman, J., and Kuennapas, T. On the relation between similarity and dissimilarity estimates. Report No. 354, Psychological Laboratories, The University of Stockholm, 1972.

Quine, W. V. Natural kinds. In W. V. Quine, *Ontological relativity and other essays.* New York: Columbia University Press, 1969.

Restle, F. *Psychology of judgment and choice.* New York: Wiley, 1961.

Rosch, E. Cognitive reference points. *Cognitive Psychology,* 1975, *7,* 532–547.

Shepard, R. N. Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology,* 1964, *1,* 54–87.

Shepard, R. N. Representation of structure in similarity data: Problems and prospects. *Psychometrika,* 1974, *39,* 373–421.

Sjöberg, L. A cognitive theory of similarity. *Goteborg Psychological Reports,* 1972, *2*(No. 10).

Torgerson, W. S. Multidimensional scaling of similarity. *Psychometrika,* 1965, *30,* 379–393.

Tversky, A. Features of similarity. *Psychological Review,* 1977, *84,* 327–352.

5 Aspects of a Stimulus: Features, Dimensions, and Configurations

W. R. Garner
Yale University

ABSTRACT

Several necessary distinctions about properties of stimuli are made. Component properties (attributes) consist of either dimensions or features. Dimensions are variables for which mutually exclusive levels exist, and quantitative dimensions are distinguished from qualitative dimensions on the basis of the role of zero: Zero is a positive value for quantitative dimensions but simply indicates absence of dimension for qualitative dimensions. Features are variables that exist or do not exist, so that zero is confounded as a level on a variable and as absence of the feature. Wholistic properties can be simple wholes, templates, or configurations, with simple wholes (and possibly templates) not being more than the sum of the parts; configural properties are emergent properties, thus other than the sum of the parts. Component and wholistic properties are different aspects of the same stimulus, because they coexist and are not independent. Implications of these distinctions for several cognitive tasks are made, including free classification, concept learning, decision and choice, and speed and accuracy of stimulus identification. Implications for modes of processing are also discussed.

ASPECTS OF A STIMULUS: FEATURES, DIMENSIONS, AND CONFIGURATIONS

In the beginning is a stimulus. At least a stimulus is at the beginning of nearly every experimental procedure designed to tell us about human information processing, or cognitive psychology, more broadly speaking. But a stimulus is