1 Features of Similarity

Amos Tversky

Similarity plays a fundamental role in theories of knowledge and behavior. It serves as an organizing principle by which individuals classify objects, form concepts, and make generalizations. Indeed, the concept of similarity is ubiquitous in psychological theory. It underlies the accounts of stimulus and response generalization in learning, it is employed to explain errors in memory and pattern recognition, and it is central to the analysis of connotative meaning.

Similarity or dissimilarity data appear in different forms: ratings of pairs, sorting of objects, communality between associations, errors of substitution, and correlation between occurrences. Analyses of these data attempt to explain the observed similarity relations and to capture the underlying structure of the objects under study.

The theoretical analysis of similarity relations has been dominated by geometric models. These models represent objects as points in some coordinate space such that the observed dissimilarities between objects correspond to the metric distances between the respective points. Practically all analyses of proximity data have been metric in nature, although some (e.g., hierarchical clustering) yield tree-like structures rather than dimensionally organized spaces. However, most theoretical and empirical analyses of similarity assume that objects can be adequately represented as points in some coordinate space and that dissimilarity behaves like a metric distance function. Both dimensional and metric assumptions are open to question.

It has been argued by many authors that dimensional representations are appropriate for certain stimuli (e.g., colors, tones) but not for others. It seems more appropriate to represent faces, countries, or personalities in terms of many qualitative features than in terms of a few quantitative dimensions. The assessment of similarity between such stimuli, therefore, may be better described as a comparison of features rather than as the computation of metric distance between points.

A metric distance function, δ, is a scale that assigns to every pair of points a non-negative number, called their distance, in accord with the following three axioms:

Minimality: $\delta(a, b) \geq \delta(a, a) = 0.$

Symmetry: $\delta(a, b) = \delta(b, a).$

The triangle inequality: $\delta(a, b) + \delta(b, c) \geq \delta(a, c).$

To evaluate the adequacy of the geometric approach, let us examine the validity of the metric axioms when δ is regarded as a measure of dissimilarity. The minimality axiom implies that the similarity between an object and itself is the same for all objects. This assumption, however, does not hold for some similarity measures. For example, the probability of judging two identical stimuli as "same" rather that "different" is not constant for all stimuli. Moreover, in recognition experiments the off-diagonal entries often exceed the diagonal entries; that is, an object is identified as another object more frequently than it is identified as itself. If identification probability is interpreted as a measure of similarity, then these observations violate minimality and are, therefore, incompatible with the distance model.

Similarity has been viewed by both philosophers and psychologists as a prime example of a symmetric relation. Indeed, the assumption of symmetry underlies essentially all theoretical treatments of similarity. Contrary to this tradition, the present paper provides empirical evidence for asymmetric similarities and argues that similarity should not be treated as a symmetric relation.

Similarity judgments can be regarded as extensions of similarity statements, that is, 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 subject and referent depends, at least in part, 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. 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." We say "an ellipse is like a circle," not "a circle is like an ellipse," and we say "North Korea is like Red China" rather than "Red China is like North Korea."

As will be 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. Likewise, an ellipse is more similar to a circle than a circle is to an ellipse. Apparently, the direction of asymmetry is determined by the relative salience of the stimuli; the variant is more similar to the prototype than vice versa.

The directionality and asymmetry of similarity relations are particularly noticeable in similies and metaphors. We say "Turks fight like tigers" and not "tigers fight like Turks." Since the tiger is renowned for its fighting spirit, it is used as the referent rather than the subject of the simile. The poet writes "my love is as deep as the ocean," not "the ocean is as deep as my love," because the ocean epitomizes depth. Sometimes both directions are used but they carry different meanings. "A man is like a tree" implies that man has roots; "a tree is like a man" implies that the tree has a life history. "Life is like a play" says that people play roles. "A play is like life" says that a play can capture the essential elements of human life. The relations between the interpretation of metaphors and the assessment of similarity are briefly discussed in the final section.

The triangle inequality differs from minimality and symmetry in that it cannot be formulated in ordinal terms. It asserts that one distance must be smaller than the sum of two others, and hence it cannot be readily refuted with ordinal or even interval data. However, the triangle inequality implies that if a is quite similar to b, and b is quite similar to c, then a and c cannot be very dissimilar from each other. Thus, it sets a lower limit to the similarity between a and c in terms of the similarities between a and b and between b and c. The following example (based on William James) casts some doubts on the psychological validity of this assumption. Consider the similarity between countries: Jamaica is similar to Cuba (because of geographical proximity); Cuba is similar to Russia (because of their political affinity); but Jamaica and Russia are not similar at all.

This example shows that similarity, as one might expect, is not transitive. In addition, it suggests that the perceived distance of Jamaica to Russia exceeds the perceived distance of Jamaica to Cuba, plus that of Cuba to Russia—contrary to the triangle inequality. Although such examples do not necessarily refute the triangle inequality, they indicate that it should not be accepted as a cornerstone of similarity models.

It should be noted that the metric axioms, by themselves, are very weak. They are satisfied, for example, by letting $\delta(a,b) = 0$ if $a = b$, and $\delta(a,b) = 1$ if $a \neq b$. To specify the distance function, additional assumptions are made (e.g., intradimensional subtractivity and interdimensional additivity) relating the dimensional structure of the objects to their metric distances. For an axiomatic analysis and a critical discussion of these assumptions, see Beals, Krantz, and Tversky (1968), Krantz and Tversky (1975), and Tversky and Krantz (1970).

In conclusion, it appears that despite many fruitful applications (see e.g., Carroll & Wish, 1974; Shepard, 1974), the geometric approach to the analysis of similarity faces several difficulties. The applicability of the dimensional assumption is limited, and the metric axioms are questionable. Specifically, minimality is somewhat problematic, symmetry is apparently false, and the triangle inequality is hardly compelling.

The next section develops an alternative theoretical approach to similarity, based on feature matching, which is neither dimensional nor metric in nature. In subsequent sections this approach is used to uncover, analyze, and explain several empirical phenomena, such as the role of common and distinctive features, the relations between judgrnents of similarity and difference, the presence of asymmetric similarities, and the effects of context on similarity. Extensions and implications of the present development are discussed in the final section.

Feature Matching

Let $\Delta = \{a, b, c, \ldots\}$ be the domain of objects (or stimuli) under study. Assume that each object in Δ is represented by a set of features or attributes, and let A, B, C denote the sets of features associated with the objects a, b, c, respectively. The features may correspond to components such as eyes or mouth; they may represent concrete properties such as size or color; and they may reflect abstract attributes such as quality or complexity. The characterization of stimuli as feature sets has been employed in the analysis of many cognitive processes such as speech perception (Jakobson, Fant, & Halle, 1961), pattern recognition (Neisser, 1967), perceptual learning (Gibson, 1969), preferential choice (Tversky, 1972), and semantic judgment (Smith, Shoben, & Rips, 1974).

Two preliminary comments regarding feature representations are in order. First, it is important to note that our total data base concerning a particular object (e.g., a person, a country, or a piece of furniture) is generally rich in content and complex in form. It includes appearance, function, relation to other objects, and any other property of the object that can be deduced from our general knowledge of the world. When faced with a particular task (e.g., identification or similarity assessment) we extract and compile from our data base a limited list of relevant features on the basis of which we perform the required task. Thus, the representation of an object as a collection of features is viewed as a product of a prior process of extraction and compilation.

Second, the term *feature* usually denotes the value of a binary variable (e.g., voiced vs. voiceless consonants) or the value of a nominal variable (e.g., eye color). Feature representations, however, are not restricted to binary or nominal variables; they are also applicable to ordinal or cardinal variables (i.e., dimensions). A series of tones that differ only in loudness, for example, could be represented as a sequence of nested sets where the feature set associated with each tone is included in the feature sets associated with louder tones. Such a representation is isomorphic to a directional unidimensional structure. A nondirectional unidimensional structure (e.g., a series of tones that differ only in pitch) could be represented by a chain of overlapping sets. The set-theoretical representation of qualitative and quantitative dimensions has been investigated by Restle (1959).

Let $s(a, b)$ be a measure of the similarity of a to b defined for all distinct a, b in Δ. The scale s is treated as an ordinal measure of similarity. That is, $s(a, b) > s(c, d)$ means that a is more similar to b than c is to d. The present theory is based on the following assumptions.

1. MATCHING:

$$s(a, b) = F(A \cap B, A - B, B - A).$$

The similarity of a to b is expressed as a function F of three arguments: $A \cap B$, the features that are common to both a and b; $A - B$, the features that belong to a but not to b; $B - A$, the features that belong to b but not to a. A schematic illustration of these components is presented in figure 1.1.

2. MONOTONICITY:

$$s(a, b) \geq s(a, c)$$

whenever

$$A \cap B \supset A \cap C, \quad A - B \subset A - C,$$

and

$$B - A \subset C - A.$$

Moreover, the inequality is strict whenever either inclusion is proper.

That is, similarity increases with addition of common features and/or deletion of distinctive features (i.e., features that belong to one object but not to the other). The monotonicity axiom can be readily illustrated with block letters if we identify their features with the component (straight) lines. Under this assumption, E should be more similar to F than to I because E and F have more common features than E and I. Furthermore, I should be more similar to F than to E because I and F have fewer distinctive features than I and E.

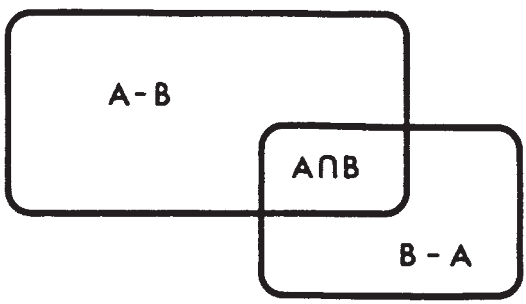

Figure 1.1
A graphical illustration of the relation between two feature sets.

Any function F satisfying assumptions 1 and 2 is called a *matching function*. It measures the degree to which two objects—viewed as sets of features—match each other. In the present theory, the assessment of similarity is described as a feature-matching process. It is formulated, therefore, in terms of the set-theoretical notion of a matching function rather than in terms of the geometric concept of distance.

In order to determine the functional form of the matching function, additional assumptions about the similarity ordering are introduced. The major assumption of the theory (independence) is presented next; the remaining assumptions and the proof of the representation theorem are presented in the appendix. Readers who are less interested in formal theory can skim or skip the following paragraphs up to the discussion of the representation theorem.

Let Φ denote the set of all features associated with the objects of Δ, and let $X, Y, Z, \ldots$ etc. denote collections of features (i.e., subsets of Φ). The expression $F(X, Y, Z)$ is defined whenever there exists a, b in Δ such that $A \cap B = X$, $A - B = Y$, and $B - A = Z$, whence $s(a, b) = F(A \cap B, A - B, B - A) = F(X, Y, Z)$. Next, define $V \simeq W$ if one or more of the following hold for some X, Y, Z: $F(V, Y, Z) = F(W, Y, Z)$, $F(X, V, Z) = F(X, W, Z)$, $F(X, Y, V) = F(X, Y, W)$.

The pairs (a, b) and (c, d) are said to *agree* on one, two, or three components, respectively, whenever one, two, or three of the following hold: $(A \cap B) \simeq (C \cap D)$, $(A - B) \simeq (C - D)$, $(B - A) \simeq (D - C)$.

3. INDEPENDENCE Suppose the pairs (a, b) and (c, d), as well as the pairs (a', b') and (c', d'), agree on the same two components, while the pairs (a, b) and (a', b'), as well as the pairs (c, d) and (c', d'), agree on the remaining (third) component. Then

$$s(a, b) \geq s(a', b') \quad \text{iff} \quad s(c, d) \geq s(c', d').$$

To illustrate the force of the independence axiom consider the stimuli presented in figure 1.2, where

$A \cap B = C \cap D =$ round profile $= X$,

$A' \cap B' = C' \cap D' =$ sharp profile $= X'$,

$A - B = C - D =$ smiling mouth $= Y$,

$A' - B' = C' - D' =$ frowning mouth $= Y'$,

$B - A = B' - A' =$ straight eyebrow $= Z$,

$D - C = D' - C' =$ curved eyebrow $= Z'$.

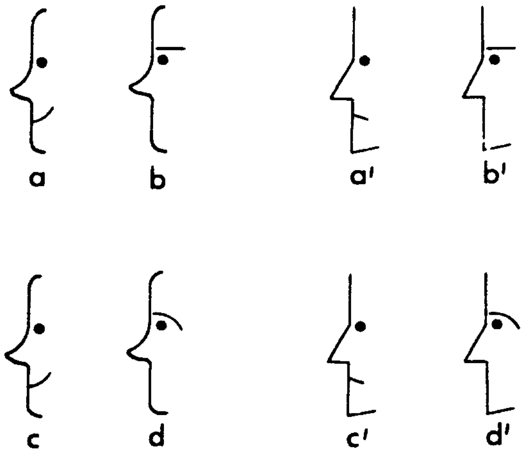

Figure 1.2
An illustration of independence.

By independence, therefore,

$$
\begin{aligned}
s(a,b) &= F(A \cap B, A - B, B - A) \\
&= F(X, Y, Z) \geq F(X', Y', Z) \\
&= F(A' \cap B', A' - B', B' - A') \\
&= s(a', b')
\end{aligned}
$$

if and only if

$$
\begin{aligned}
s(c,d) &= F(C \cap D, C - D, D - C) \\
&= F(X, Y, Z') \geq F(X', Y', Z') \\
&= F(C' \cap D', C' - D', D' - C') \\
&= s(c', d').
\end{aligned}
$$

Thus, the ordering of the joint effect of any two components (e.g., X, Y vs. X', Y') is independent of the fixed level of the third factor (e.g., Z or Z').

It should be emphasized that any test of the axioms presupposes an interpretation of the features. The independence axiom, for example, may hold in one interpretation and fail in another. Experimental tests of the axioms, therefore, test jointly the adequacy of the interpretation of the features and the empirical validity of the assumptions. Furthermore, the above examples should not be taken to mean that stimuli (e.g., block letters, schematic faces) can be properly characterized in terms of their components. To achieve an adequate feature representation of visual forms, more global properties (e.g., symmetry, connectedness) should also be introduced. For an interesting discussion of this problem, in the best tradition of Gestalt psychology, see Goldmeier (1972; originally published in 1936).

In addition to matching (1), monotonicity (2), and independence (3), we also assume solvability (4), and invariance (5). Solvability requires that the feature space under study be sufficiently rich that certain (similarity) equations can be solved. Invariance ensures that the equivalence of intervals is preserved across factors. A rigorous formulation of these assumptions is given in the Appendix, along with a proof of the following result.

Representation Theorem Suppose assumptions 1, 2, 3, 4, and 5 hold. Then there exist a similarity scale S and a nonnegative scale f such that for all a, b, c, d in Δ,

(i) $S(a, b) \geq S(c, d)$ iff $s(a, b) \geq s(c, d)$;

(ii) $S(a, b) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A)$, for some $\theta, \alpha, \beta \geq 0$;

(iii) f and S are interval scales.

The theorem shows that under assumptions 1–5, there exists an interval similarity scale S that preserves the observed similarity order and expresses similarity as a linear combination, or a contrast, of the measures of the common and the distinctive features. Hence, the representation is called the *contrast model*. In parts of the following development we also assume that f satisfies feature additivity. That is, $f(X \cup Y) = f(X) + f(Y)$ whenever X and Y are disjoint, and all three terms are defined.[1]

Note that the contrast model does not define a single similarity scale, but rather a family of scales characterized by different values of the parameters θ, α, and β. For example, if $\theta = 1$ and α and β vanish, then $S(a, b) = f(A \cap B)$; that is, the similarity between objects is the measure of their common features. If, on the other hand, $\alpha = \beta = 1$ and θ vanishes then $-S(a, b) = f(A - B) + f(B - A)$; that is, the dissimilarity between objects is the measure of the symmetric difference between the respective feature sets. Restle (1961) has proposed these forms as models of similarity and psychological distance, respectively. Note that in the former model ($\theta = 1$, $\alpha = \beta = 0$), similarity between objects is determined only by their common features, whereas in the latter model ($\theta = 0$, $\alpha = \beta = 1$), it is determined by their distinctive features only. The contrast model expresses similarity between objects as a weighted difference of the measures of their common and distinctive features, thereby allowing for a variety of similarity relations over the same domain.

The major constructs of the present theory are the contrast rule for the assessment of similarity, 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.

Let us recapitulate what is assumed and what is proven in the representation theorem. We begin with a set of objects, described as collections of features, and a similarity ordering which is assumed to satisfy the axioms of the present theory. From these assumptions, we derive a measure f on the feature space and prove that the similarity ordering of object pairs coincides with the ordering of their contrasts, defined as linear combinations of the respective common and distinctive features. Thus, the measure f and the contrast model are derived from qualitative axioms regarding the similarity of objects.

The nature of this result may be illuminated by an analogy to the classical theory of decision under risk (von Neumann & Morgenstern, 1947). In that theory, one starts with a set of prospects, characterized as probability distributions over some consequence space, and a preference order that is assumed to satisfy the axioms of the theory. From these assumptions one derives a utility scale on the consequence space and proves that the preference order between prospects coincides with the order of their expected utilities. Thus, the utility scale and the expectation principle are derived from qualitative assumptions about preferences. The present theory of similarity differs from the expected-utility model in that the characterization of objects as feature sets is perhaps more problematic than the characterization of uncertain options as probability distributions. Furthermore, the axioms of utility theory are proposed as (normative) principles of rational behavior, whereas the axioms of the present theory are intended to be descriptive rather than prescriptive.

The contrast model is perhaps the simplest form of a matching function, yet it is not the only form worthy of investigation. Another matching function of interest is the *ratio model*,

$$S(a, b) = \frac{f(A \cap B)}{f(A \cap B) + \alpha f(A - B) + \beta f(B - A)}, \quad \alpha, \beta \geq 0,$$

where similarity is normalized so that S lies between 0 and 1. The ratio model generalizes several set-theoretical models of similarity proposed in the literature. If $\alpha = \beta = 1$, $S(a, b)$ reduces to $f(A \cap B)/f(A \cup B)$ (see Gregson, 1975, and Sjöberg, 1972). If $\alpha = \beta = \frac{1}{2}$, $S(a, b)$ equals $2f(A \cap B)/(f(A) + f(B))$ (see Eisler & Ekman, 1959). If $\alpha = 1$ and $\beta = 0$, $S(a, b)$ reduces to $f(A \cap B)/f(A)$ (see Bush & Mosteller, 1951). The present framework, therefore, encompasses a wide variety of similarity models that differ in the form of the matching function F and in the weights assigned to its arguments.

In order to apply and test the present theory in any particular domain, some assumptions about the respective feature structure must be made. If the features associated with each object are explicitly specified, we can test the axioms of the theory directly and scale the features according to the contrast model. This approach, however, is generally limited to stimuli (e.g., schematic faces, letters, strings of symbols) that are constructed from a fixed feature set. If the features associated with the objects under study cannot be readily specified, as is often the case with natural stimuli, we can still test several predictions of the contrast model which involve only general qualitative assumptions about the feature structure of the objects. Both approaches were employed in a series of experiments conducted by Itamar Gati and the present author. The following three sections review and discuss our main findings, focusing primarily on the test of qualitative predictions. A more detailed description of the stimuli and the data are presented in Tversky and Gati (in press).

Asymmetry and Focus

According to the present analysis, similarity is not necessarily a symmetric relation. Indeed, it follows readily (from either the contrast or the ratio model) that

$$\begin{aligned} s(a, b) = s(b, a) \quad &\text{iff} \quad \alpha f(A - B) + \beta f(B - A) = \alpha f(B - A) + \beta f(A - B) \\ &\text{iff} \quad (\alpha - \beta) f(A - B) = (\alpha - \beta) f(B - A). \end{aligned}$$

Hence, $s(a, b) = s(b, a)$ if either $\alpha = \beta$, or $f(A - B) = f(B - A)$, which implies $f(A) = f(B)$, provided feature additivity holds. Thus, symmetry holds whenever the objects are equal in measure ($f(A) = f(B)$) or the task is nondirectional ($\alpha = \beta$). To interpret the latter condition, compare the following two forms:

(i) Assess the degree to which a and b are similar to each other.

(ii) Assess the degree to which a is similar to b.

In (i), the task is formulated in a nondirectional fashion; hence it is expected that $\alpha = \beta$ and $s(a, b) = s(b, a)$. In (ii), on the other hand, the task is directional, and hence α and β may differ and symmetry need not hold.

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., $\alpha > \beta$). Consequently, similarity is reduced more by the distinctive features of the subject than by the distinctive features of the referent. It follows readily that whenever $\alpha > \beta$,

$$s(a, b) > s(b, a) \quad \text{iff} \quad f(B) > f(A).$$

Thus, the focusing hypothesis (i.e., $\alpha > \beta$) 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 particular, the variant is more similar to the prototype than the prototype is to the variant, because the prototype is generally more salient than the variant.

Similarity of Countries

Twenty-one pairs of countries served as stimuli. The pairs were constructed so that one element was more prominent than the other (e.g., Red China–North Vietnam, USA–Mexico, Belgium–Luxemburg). To verify this relation, we asked a group of 69 subjects[2] to select in each pair the country they regarded as more prominent. The proportion of subjects that agreed with the a priori ordering exceeded $\frac{2}{3}$ for all pairs except one. A second group of 69 subjects was asked to choose which of two phrases they preferred to use: "country a is similar to country b," or "country b is similar to country a." In all 21 cases, most of the subjects chose the phrase in which the less prominent country served as the subject and the more prominent country as the referent. For example, 66 subjects selected the phrase "North Korea is similar to Red China" and only 3 selected the phrase "Red China is similar to North Korea." These results demonstrate the presence of marked asymmetries in the choice of similarity statements, whose direction coincides with the relative prominence of the stimuli.

To test for asymmetry in direct judgments of similarity, we presented two groups of 77 subjects each with the same list of 21 pairs of countries and asked subjects to rate their similarity on a 20-point scale. 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 the USSR is similar to Poland," whereas the second group was asked to assess "the degree to which Poland is similar to the USSR." The lists were constructed so that the more prominent country appeared about an equal number of times in the first and second positions.

For any pair (p, q) of stimuli, let p denote the more prominent element, and let q denote the less prominent element. The average s(q, p) was significantly higher than the average s(p, q) across all subjects and pairs: *t* test for correlated samples yielded $t(20) = 2.92$, $p < .01$. 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; that is, s(q, p), minus the average similarity for comparisons with a prominent subject, s(p, q). The average difference was significantly positive: $t(153) = 2.99$, $p < .01$.

The above study was repeated using judgments of difference instead of judgments of similarity. Two groups of 23 subjects each participated in this study. They received the same list of 21 pairs except that one group was asked to judge the degree to which country a differed from country b, denoted d(a, b), whereas the second group was asked to judge the degree to which country b was different from country a, denoted d(b, a). If judgments of difference follow the contrast model, and $\alpha > \beta$, then we expect the prominent stimulus p to differ from the less prominent stimulus q more than q differs from p; that is, $d(p, q) > d(q, p)$. This hypothesis was tested using the same set of 21 pairs of countries and the prominence ordering established earlier. The average d(p, q), across all subjects and pairs, was significantly higher than the average d(q, p): *t* test for correlated samples yielded $t(20) = 2.72$, $p < .01$. Furthermore, the average asymmetry score, computed as above for each subject, was significantly positive, $t(45) = 2.24$, $p < .05$.

Similarity of Figures

A major determinant of the salience of geometric figures is goodness of form. Thus, a "good figure" is likely to be more salient than a "bad figure," although the latter is generally more complex. However, when two figures are roughly equivalent with respect to goodness of form, the more complex figure is likely to be more salient. To investigate these hypotheses and to test the asymmetry prediction, two sets of eight pairs of geometric figures were constructed. 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 matched in 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 figure 1.3.

A group of 69 subjects was presented with the entire list of 16 pairs of figures, where the two elements of each pair were displayed side by side. For each pair, the subjects were asked to indicate which of the following two statements they preferred

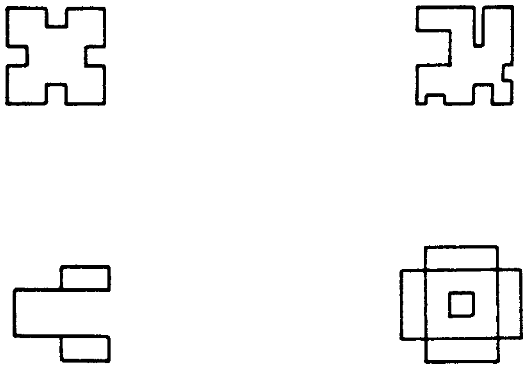

Figure 1.3
Examples of pairs of figures used to test the prediction of asymmetry. The top two figures are examples of a pair (from the first set) that differs in goodness of form. The bottom two are examples of a pair (from the second set) that differs in complexity.

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 stimuli were randomized so that p and q appeared an equal number of times on the left and on the right. The results showed that in each one of the pairs, most of the subjects selected the form "q is similar to p." Thus, the more salient stimulus was generally chosen as the referent rather than the subject of similarity statements.

To test for asymmetry in judgments of similarity, we presented two groups of 67 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 was 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 results showed 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 $t(15) = 2.94$, $p < .01$. Furthermore, in both sets the average asymmetry scores, computed as above for each subject, were significantly positive: In the first set $t(131) = 2.96$, $p < .01$, and in the second set $t(131) = 2.79$, $p < .01$.

Similarity of Letters

A common measure of similarity between stimuli is the probability of confusing them in a recognition or an identification task: The more similar the stimuli, the more likely they are to be confused. While confusion probabilities are often asymmetric (i.e., the probability of confusing a with b is different from the probability of confusing b with a), this effect is typically attributed to a response bias. To eliminate this interpretation of asymmetry, one could employ an experimental task where the subject merely indicates whether the two stimuli presented to him (sequentially or simultaneously) are identical or not. This procedure was employed by Yoav Cohen and the present author in a study of confusion among block letters.

The following eight block letters served as stimuli: ┌, ⊏, ⊓, □, ╒, ⋿, ⍐, ⊟. All pairs of letters were displayed on a cathode-ray tube, side by side, on a noisy background. The letters were presented sequentially, each for approximately 1 msec. The right letter always followed the left letter with an interval of 630 msec in between. After each presentation the subject pressed one of two keys to indicate whether the two letters were identical or not.

A total of 32 subjects participated in the experiment. Each subject was tested individually. On each trial, one letter (known in advance) served as the standard. For one half of the subjects the standard stimulus always appeared on the left, and for the other half of the subjects the standard always appeared on the right. Each one of the eight letters served as a standard. The trials were blocked into groups of 10 pairs in which the standard was paired once with each of the other letters and three times with itself. Since each letter served as a standard in one block, the entire design consisted of eight blocks of 10 trials each. Every subject was presented with three replications of the entire design (i.e., 240 trials). The order of the blocks in each design and the order of the letters within each block were randomized.

According to the present analysis, people compare the variable stimulus, which serves the role of the subject, to the standard (i.e., the referent). The choice of standard, therefore, determines the directionality of the comparison. A natural partial ordering of the letters with respect to prominence is induced by the relation of inclusion among letters. Thus, one letter is assumed to have a larger measure than another if the former includes the latter. For example, ⋿ includes ╒ and ┌ but not □. For all 19 pairs in which one letter includes the other, let p denote the more prominent letter and q denote the less prominent letter. Furthermore, let $s(a, b)$ denote the percentage of times that the subject judged the variable stimulus a to be the same as the standard b.

It follows from the contrast model, with $\alpha > \beta$, that the proportion of "same" responses should be larger when the variable is included in the standard than when the standard is included in the variable, that is, $s(q, p) > s(p, q)$. This prediction was borne out by the data. The average $s(q, p)$ across all subjects and trials was 17.1%, whereas the average $s(p, q)$ across all subjects and trials was 12.4%. To obtain a statistical test, we computed for each subject the difference between $s(q, p)$ and $s(p, q)$ across all trials. The difference was significantly positive, $t(31) = 4.41$, $p < .001$.

These results demonstrate that the prediction of directional asymmetry derived from the contrast model applies to confusion data and not merely to rated similarity.

Similarity of Signals

Rothkopf (1957) presented 598 subjects with all ordered pairs of the 36 Morse Code signals and asked them to indicate whether the two signals in each pair were the same or not. The pairs were presented in a randomized order without a fixed standard. Each subject judged about one fourth of all pairs.

Let s(a, b) denote the percentage of "same" responses to the ordered pair (a, b), i.e., the percentage of subjects that judged the first signal a to be the same as the second signal b. Note that a and b refer here to the first and second signal, and not to the variable and the standard as in the previous section. Obviously, Morse Code signals are partially ordered according to temporal length. For any pair of signals that differ in temporal length, let p and q denote, respectively, the longer and shorter element of the pair.

From the total of 555 comparisons between signals of different length, reported in Rothkopf (1957), s(q, p) exceeds s(p, q) in 336 cases, s(p, q) exceeds s(q, p) in 181 cases, and s(q, p) equals s(p, q) in 38 cases, $p < .001$, by sign test. The average difference between s(q, p) and s(p, q) across all pairs is 3.3%, which is also highly significant. A t test for correlated samples yields $t(554) = 9.17$, $p < .001$.

The asymmetry effect is enhanced when we consider only those comparisons in which one signal is a proper subsequence of the other. (For example, ·· is a subsequence of ··- as well as of ·--·). From a total of 195 comparisons of this type, s(q, p) exceeds s(p, q) in 128 cases, s(p, q) exceeds s(q, p) in 55 cases, and s(q, p) equals s(p, q) in 12 cases, $p < .001$ by sign test. The average difference between s(q, p) and s(p, q) in this case is 4.7%, $t(194) = 7.58$, $p < .001$.

A later study following the same experimental paradigm with somewhat different signals was conducted by Wish (1967). His signals consisted of three tones separated by two silent intervals, where each component (i.e., a tone or a silence) was either short or long. Subjects were presented with all pairs of 32 signals generated in this fashion and judged whether the two members of each pair were the same or not.

The above analysis is readily applicable to Wish's (1967) data. From a total of 386 comparisons between signals of different length, s(q, p) exceeds s(p, q) in 241 cases, s(p, q) exceeds s(q, p) in 117 cases, and s(q, p) equals s(p, q) in 28 cases. These data are clearly asymmetric, $p < .001$ by sign test. The average difference between s(q, p) and s(p, q) is 5.9%, which is also highly significant, $t(385) = 9.23$, $p < .001$.

In the studies of Rothkopf and Wish there is no a priori way to determine the directionality of the comparison, or equivalently to identify the subject and the referent. However, if we accept the focusing hypothesis ($\alpha > \beta$) and the assumption that longer signals are more prominent than shorter ones, then the direction of the observed asymmetry indicates that the first signal serves as the subject that is compared with the second signal that serves the role of the referent. Hence, the directionality of the comparison is determined, according to the present analysis, from the prominence ordering of the stimuli and the observed direction of asymmetry.

Rosch's Data

Rosch (1973, 1975) has articulated and supported the view that perceptual and semantic categories are naturally formed and defined in terms of focal points, or prototypes. Because of the special role of prototypes in the formation of categories, she hypothesized that (i) in sentence frames involving hedges such as "a is essentially b," focal stimuli (i.e., prototypes) appear in the second position; and (ii) the perceived distance from the prototype to the variant is greater than the perceived distance from the variant to the prototype. To test these hypotheses, Rosch (1975) used three stimulus domains: color, line orientation, and number. Prototypical colors were focal (e.g., pure red), while the variants were either non-focal (e.g., off-red) or less saturated. Vertical, horizontal, and diagonal lines served as prototypes for line orientation, and lines of other angles served as variants. Multiples of 10 (e.g., 10, 50, 100) were taken as prototypical numbers, and other numbers (e.g., 11, 52, 103) were treated as variants.

Hypothesis (i) was strongly confirmed in all three domains. When presented with sentence frames such as "____ is virtually ____," subjects generally placed the prototype in the second blank and the variant in the first. For instance, subjects preferred the sentence "103 is virtually 100" to the sentence "100 is virtually 103." To test hypothesis (ii), 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." As hypothesized, the measured distance between stimuli was significantly smaller when the prototype, rather than the variant, was fixed at the origin, in each of the three domains.

If focal stimuli are more salient than non-focal stimuli, then Rosch's findings support the present analysis. The hedging sentences (e.g., "a is roughly b") can be regarded as a particular type of similarity statements. Indeed, the hedges data are in perfect agreement with the choice of similarity statements. Furthermore, the observed asymmetry in distance placement follows from the present analysis of asymmetry and the natural assumptions that the standard and the variable serve, respectively, as referent and subject in the distance-placement task. Thus, the placement of b at distance t from a is interpreted as saying that the (perceived) distance from b to a equals t.

Rosch (1975) attributed the observed asymmetry to the special role of distinct prototypes (e.g., a perfect square or a pure red) in the processing of information. In the present theory, on the other hand, asymmetry is explained by the relative salience of the stimuli. Consequently, it implies asymmetry for pairs that do not include the prototype (e.g., two levels of distortion of the same form). If the concept of prototypicality, however, is interpreted in a relative sense (i.e., a is more prototypical than b) rather than in an absolute sense, then the two interpretations of asymmetry practically coincide.

Discussion

The conjunction of the contrast model and the focusing hypothesis implies the presence of asymmetric similarities. This prediction was confirmed in several experiments of perceptual and conceptual similarity using both judgmental methods (e.g., rating) and behavioral methods (e.g., choice).

The asymmetries discussed in the previous section were observed in *comparative* tasks in which the subject compares two given stimuli to determine their similarity. Asymmetries were also observed in *production* tasks in which the subject is given a single stimulus and asked to produce the most similar response. Studies of pattern recognition, stimulus identification, and word association are all examples of production tasks. A common pattern observed in such studies is that the more salient object occurs more often as a response to the less salient object than vice versa. For example, "tiger" is a more likely associate to "leopard" than "leopard" is to "tiger." Similarly, Garner (1974) instructed subjects to select from a given set of dot patterns one that is similar—but not identical—to a given pattern. His results show that "good" patterns are usually chosen as responses to "bad" patterns and not conversely.

This asymmetry in production tasks has commonly been attributed to the differential availability of responses. Thus, "tiger" is a more likely associate to "leopard" than vice versa, because "tiger" is more common and hence a more available response than "leopard." This account is probably more applicable to situations where the subject must actually produce the response (as in word association or pattern recognition) than to situations where the subject merely selects a response from some specified set (as in Garner's task).

Without questioning the importance of response availability, the present theory suggests another reason for the asymmetry observed in production tasks. Consider the following translation of a production task to a question-and-answer scheme.

Question: What is a like? Answer: a is like b. If this interpretation is valid and the given object a serves as a subject rather than as a referent, then the observed asymmetry of production follows from the present theoretical analysis, since $s(a,b) > s(b,a)$ whenever $f(B) > f(A)$.

In summary, it appears that proximity data from both comparative and production tasks reveal significant and systematic asymmetries whose direction is determined by the relative salience of the stimuli. Nevertheless, the symmetry assumption should not be rejected altogether. It seems to hold in many contexts, and it serves as a useful approximation in many others. It cannot be accepted, however, as a universal principle of psychological similarity.

Common and Distinctive Features

In the present theory, the similarity of objects is expressed as a linear combination, or a contrast, of the measures of their common and distinctive features. This section investigates the relative impact of these components and their effect on the relation between the assessments of similarity and difference. The discussion concerns only symmetric tasks, where $\alpha = \beta$, and hence $s(a,b) = s(b,a)$.

Elicitation of Features

The first study employs the contrast model to predict the similarity between objects from features that were produced by the subjects. The following 12 vehicles served as stimuli: bus, car, truck, motorcycle, train, airplane, bicycle, boat, elevator, cart, raft, sled. One group of 48 subjects rated the similarity between all 66 pairs of vehicles on a scale from 1 (no similarity) to 20 (maximal similarity). Following Rosch and Mervis (1975), we instructed a second group of 40 subjects to list the characteristic features of each one of the vehicles. Subjects were given 70 sec to list the features that characterized each vehicle. Different orders of presentation were used for different subjects.

The number of features per vehicle ranged from 71 for airplane to 21 for sled. Altogether, 324 features were listed by the subjects, of which 224 were unique and 100 were shared by two or more vehicles. For every pair of vehicles we counted the number of features that were attributed to both (by at least one subject), and the number of features that were attributed to one vehicle but not to the other. The frequency of subjects that listed each common or distinctive feature was computed.

In order to predict the similarity between vehicles from the listed features, the measures of their common and distinctive features must be defined. The simplest measure is obtained by counting the number of common and distinctive features produced by the subjects. The product-moment correlation between the (average) similarity of objects and the number of their common features was .68. The correlation between the similarity of objects and the number of their distinctive features was −.36. The multiple correlation between similarity and the numbers of common and distinctive features (i.e., the correlation between similarity and the contrast model) was .72.

The counting measure assigns equal weight to all features regardless of their frequency of mention. To take this factor into account, let X_a denote the proportion of subjects who attributed feature X to object a, and let N_X denote the number of objects that share feature X. For any a, b, define the measure of their common features by $f(A \cap B) = \sum X_a X_b / N_X$, where the summation is over all X in $A \cap B$, and the measure of their distinctive features by

$$f(A - B) + f(B - A) = \sum Y_a + \sum Z_b$$

where the summations range over all $Y \in A - B$ and $Z \in B - A$, that is, the distinctive features of a and b, respectively. The correlation between similarity and the above measure of the common features was .84; the correlation between similarity and the above measure of the distinctive features was −.64. The multiple correlation between similarity and the measures of the common and the distinctive features was .87.

Note that the above methods for defining the measure f were based solely on the elicited features and did not utilize the similarity data at all. Under these conditions, a perfect correlation between the two should not be expected because the weights associated with the features are not optimal for the prediction of similarity. A given feature may be frequently mentioned because it is easily labeled or recalled, although it does not have a great impact on similarity, and vice versa. Indeed, when the features were scaled using the additive tree procedure (Sattath & Tversky, in press) in which the measure of the features is derived from the similarities between the objects, the correlation between the data and the model reached .94.

The results of this study indicate that (i) it is possible to elicit from subjects detailed features of semantic stimuli such as vehicles (see Rosch & Mervis, 1975); (ii) the listed features can be used to predict similarity according to the contrast model with a reasonable degree of success; and (iii) the prediction of similarity is improved when frequency of mention and not merely the number of features is taken into account.

Similarity versus Difference

It has been generally assumed that judgments of similarity and difference are complementary; that is, judged difference is a linear function of judged similarity with a slope of -1. This hypothesis has been confirmed in several studies. For example, Hosman and Kuennapas (1972) obtained independent judgments of similarity and difference for all pairs of lowercase 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 sum of the two judgments for each pair was quite close to 20 in all cases. The product–moment correlation between the ratings was again $-.98$. This inverse relation between similarity and difference, however, does not always hold.

Naturally, an increase in the measure of the common features increases similarity and decreases difference, whereas an increase in the measure of the distinctive features decreases similarity and increases difference. However, the relative weight assigned to the common and the distinctive features may differ in the two tasks. In the assessment of similarity between objects the subject may attend more to their common features, whereas in the assessment of difference between objects the subject may attend more to their distinctive features. Thus, the relative weight of the common features will be greater in the former task than in the latter task.

Let $d(a, b)$ denote the perceived difference between a and b. Suppose d satisfies the axioms of the present theory with the reverse inequality in the monotonicity axiom, that is, $d(a, b) \leq d(a, c)$ whenever $A \cap B \supset A \cap C$, $A - B \subset A - C$, and $B - A \subset C - A$. Furthermore, suppose s also satisfies the present theory and assume (for simplicity) that both d and s are symmetric. According to the representation theorem, therefore, there exist a nonnegative scale f and nonnegative constants θ and λ such that for all a, b, c, e,

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

and

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

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

$$f(A \cap B) > f(C \cap E)$$

and

$$f(A - B) + f(B - A) > f(C - E) + f(E - C).$$

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.

To test this hypothesis, 20 sets of four countries were constructed on the basis of a pilot test. 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 our subjects (e.g., USA–USSR, Red China–Japan). The nonprominent pairs consisted of countries that were known to the subjects, but not as well as the prominent ones (e.g., Tunis–Morocco, Paraguay–Ecuador). All subjects were presented with the same 20 sets. One group of 30 subjects selected between the two pairs in each set the pair of countries that were more *similar*. Another group of 30 subjects selected between the two pairs in each set the pair of countries that were more *different*.

Let Π_s and Π_d denote, respectively, the percentage of choices where the prominent pair of countries was selected as more similar or as more different. If similarity and difference are complementary (i.e., $\theta = \lambda$), then $\Pi_s + \Pi_d$ should equal 100 for all pairs. On the other hand, if $\theta > \lambda$, then $\Pi_s + \Pi_d$ should exceed 100. The average value of $\Pi_s + \Pi_d$, across all sets, was 113.5, which is significantly greater than 100, $t(59) = 3.27$, $p < .01$.

Moreover, on the average, the prominent pairs were selected more frequently than the nonprominent pairs in both the similarity and the difference tasks. For example, 67% of the subjects in the similarity group selected West Germany and East Germany as more similar to each other than Ceylon and Nepal, while 70% of the subjects in the difference group selected West Germany and East Germany as more different from each other than Ceylon and Nepal. These data demonstrate how the relative weight of the common and the distinctive features varies with the task and support the hypothesis that people attend more to the common features in judgments of similarity than in judgments of difference.

Similarity in Context

Like other judgments, similarity depends on context and frame of reference. Sometimes the relevant frame of reference is specified explicitly, as in the questions, "How similar are English and French with respect to sound?" "What is the similarity of a pear and an apple with respect to taste?" In general, however, the relevant feature space is not specified explicitly but rather inferred from the general context.

When subjects are asked to assess the similarity between the USA and the USSR, for instance, they usually assume that the relevant context is the set of countries and that the relevant frame of reference includes all political, geographical, and cultural features. The relative weights assigned to these features, of course, may differ for different people. With natural, integral stimuli such as countries, people, colors, and sounds, there is relatively little ambiguity regarding the relevant feature space. However, with artificial, separable stimuli, such as figures varying in color and shape, or lines varying in length and orientation, subjects sometimes experience difficulty in evaluating overall similarity and occasionally tend to evaluate similarity with respect to one factor or the other (Shepard, 1964) or change the relative weights of attributes with a change in context (Torgerson, 1965).

In the present theory, changes in context or frame of reference correspond to changes in the measure of the feature space. When asked to assess the political similarity between countries, for example, the subject presumably attends to the political aspects of the countries and ignores, or assigns a weight of zero to, all other features. In addition to such restrictions of the feature space induced by explicit or implicit instructions, the salience of features and hence the similarity of objects are also influenced by the effective context (i.e., the set of objects under consideration). To understand this process, let us examine the factors that determine the salience of a feature and its contribution to the similarity of objects.

The Diagnosticity Principle

The salience (or the measure) of a feature is determined by two types of factors: intensive and diagnostic. The former refers to factors that increase intensity or signal-to-noise ratio, such as the brightness of a light, the loudness of a tone, the saturation of a color, the size of a letter, the frequency of an item, the clarity of a picture, or the vividness of an image. The diagnostic factors refer to the classificatory significance of features, that is, the importance or prevalence of the classifications that are based on these features. Unlike the intensive factors, the diagnostic factors are highly sensitive to the particular object set under study. For example, the feature "real" has no diagnostic value in the set of actual animals since it is shared by all actual animals and hence cannot be used to classify them. This feature, however, acquires considerable diagnostic value if the object set is extended to include legendary animals, such as a centaur, a mermaid, or a phoenix.

When faced with a set of objects, people often sort them into clusters to reduce information load and facilitate further processing. Clusters are typically selected so as to maximize the similarity of objects within a cluster and the dissimilarity of objects from different clusters. Hence, the addition and/or deletion of objects can alter the clustering of the remaining objects. A change of clusters, in turn, is expected to increase the diagnostic value of features on which the new clusters are based, and therefore, the similarity of objects that share these features. This relation between similarity and grouping—called the diagnosticity hypothesis—is best explained in terms of a concrete example. Consider the two sets of four schematic faces (displayed in figure 1.4), which differ in only one of their elements (p and q).

The four faces of each set were displayed in a row and presented to a different group of 25 subjects who were instructed to partition them into two pairs. The most frequent partition of set 1 was c and p (smiling faces) versus a and b (nonsmiling faces). The most common partition of set 2 was b and q (frowning faces) versus a and c (nonfrowning faces). Thus, the replacement of p by q changed the grouping of a: In set 1 a was paired with b, while in set 2 a was paired with c.

According to the above analysis, smiling has a greater diagnostic value in set 1 than in set 2, whereas frowning has a greater diagnostic value in set 2 than in set 1. By the diagnosticity hypothesis, therefore, similarity should follow the grouping. That is, the similarity of a (which has a neutral expression) to b (which is frowning) should be greater in set 1, where they are grouped together, than in set 2, where they are grouped separately. Likewise, the similarity of a to c (which is smiling) should be greater in set 2, where they are grouped together, than in set 1, where they are not.

To test this prediction, two different groups of 50 subjects were presented with sets 1 and 2 (in the form displayed in figure 1.4) and asked to select one of the three faces below (called the choice set) that was most similar to the face on the top (called the target). The percentage of subjects who selected each of the three elements of the choice set is presented below the face. The results confirmed the diagnosticity hypothesis: b was chosen more frequently in set 1 than in set 2, whereas c was chosen more frequently in set 2 than in set 1. Both differences are statistically significant, $p < .01$. Moreover, the replacement of p by q actually reversed the similarity ordering: In set 1, b is more similar to a than c, whereas in set 2, c is more similar to a than b.

A more extensive test of the diagnosticity hypothesis was conducted using semantic rather than visual stimuli. The experimental design was essentially the same,

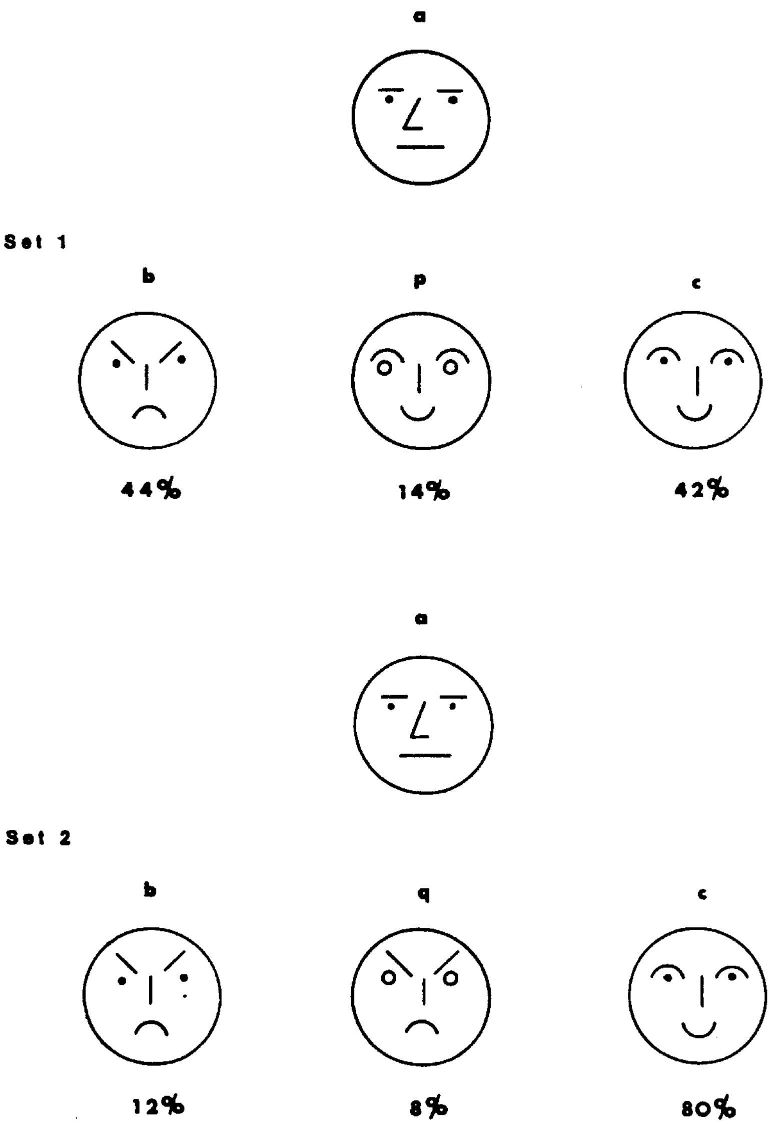

Figure 1.4
Two sets of schematic faces used to test the diagnosticity hypothesis. The percentage of subjects who selected each face (as most similar to the target) is presented below the face.

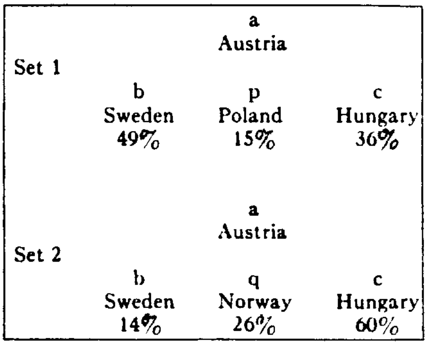

Figure 1.5
Two sets of countries used to test the diagnosticity hypothesis. The percentage of subjects who selected each country (as most similar to Austria) is presented below the country.

except that countries served as stimuli instead of faces. Twenty pairs of matched sets of four countries of the form $\{a, b, c, p\}$ and $\{a, b, c, q\}$ were constructed. An example of two matched sets is presented in figure 1.5.

Note that the two matched sets (1 and 2) differ only by one element (p and q). The sets were constructed so that a (in this case Austria) is likely to be grouped with b (e.g., Sweden) in set 1, and with c (e.g., Hungary) in set 2. To validate this assumption, we presented two groups of 25 subjects with all sets of four countries and asked them to partition each quadruple into two pairs. Each group received one of the two matched quadruples, which were displayed in a row in random order. The results confirmed our prior hypothesis regarding the grouping of countries. In every case but one, the replacement of p by q changed the pairing of the target country in the predicted direction, $p < .01$ by sign test. For example, Austria was paired with Sweden by 60% of the subjects in set 1, and it was paired with Hungary by 96% of the subjects in set 2.

To test the diagnosticity hypothesis, we presented two groups of 35 subjects with 20 sets of four countries in the format displayed in figure 1.5. These subjects were asked to select, for each quadruple, the country in the choice set that was most similar to the target country. 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 chose b rather than c as most similar to a should be the same regardless of whether the third element in the choice set is p or q. For example, the proportion of subjects who select Sweden rather than Hungary as most similar to Austria should be independent of whether the odd element in the choice set is Norway or Poland.

In contrast, the diagnosticity hypothesis implies that the change in grouping, induced by the substitution of the odd element, will change the similarities in a predictable manner. Recall that in set 1 Poland was paired with Hungary, and Austria with Sweden, while in set 2 Norway was paired with Sweden, and Austria with Hungary. Hence, the proportion of subjects who select Sweden rather than Hungary (as most similar to Austria) should be higher in set 1 than in set 2. This prediction is strongly supported by the data in figure 1.5, which show that Sweden was selected more frequently than Hungary in set 1, while Hungary was selected more frequently than Sweden in set 2.

Let b(p) denote the percentage of subjects who chose country b as most similar to a when the odd element in the choice set is p, and so on. As in the above examples, the notation is chosen so 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, reflect the effects of the odd elements, p and q, on the similarity of b and c to the target a. In the absence of context effects, both differences should equal 0, while under the diagnosticity hypothesis both differences should be positive. In figure 1.5, for example, $b(p) - b(q) = 49 - 14 = 35$, and $c(q) - c(p) = 60 - 36 = 24$. The average difference, across all pairs of quadruples, equals 9%, which is significantly positive, $t(19) = 3.65$, $p < .01$.

Several variations of the experiment did not alter the nature of the results. The diagnosticity hypothesis was also confirmed when (i) each choice set contained four elements, rather than three, (ii) the subjects were instructed to rank the elements of each choice set according to their similarity to the target, rather than to select the most similar element, and (iii) the target consisted of two elements, and the subjects were instructed to select one element of the choice set that was most similar to the two target elements. For further details, see Tversky and Gati (in press).

The Extension Effect

Recall that the diagnosticity of features is determined by the classifications that are based on them. Features that are shared by all the objects under consideration cannot be used to classify these objects and are, therefore, devoid of diagnostic value. When the context is extended by the enlargement of 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 will usually be smaller than their similarity in the extended context.

Essentially the same account was proposed and supported by Sjöberg[3] in studies of similarity between animals, and between musical instruments. For example, Sjöberg showed that the similarities between string instruments (banjo, violin, harp, electric guitar) were increased when a wind instrument (clarinet) was added to this set. Since the string instruments are more similar to each other than to the clarinet, however, the above result may be attributed, in part at least, to subjects' tendency to standardize the response scale, that is, to produce the same average similarity for any set of comparisons.

This effect can be eliminated by the use of a somewhat different design, employed in the following study. Subjects were presented with pairs of countries having a common border and assessed their similarity on a 20-point scale. Four sets of eight pairs were constructed. Set 1 contained eight pairs of European countries (e.g., Italy–Switzerland). Set 2 contained eight pairs of American countries (e.g., Brazil–Uruguay). Set 3 contained four pairs from set 1 and four pairs from set 2, while 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.

According to the diagnosticity hypothesis, the features "European" and "American" have no diagnostic value in sets 1 and 2, although they both have a 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). This prediction was confirmed by the data, $t(15) = 2.11$, $p < .05$.

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[3] study, which extended the context by introducing nonhomogeneous pairs, our experiment extended the context by constructing heterogeneous sets composed of homogeneous pairs. Hence, the increase of similarity with the enlargement of context, observed in the present study, cannot be explained by subjects' tendency to equate the average similarity for any set of assessments.

The Two Faces of Similarity

According to the present analysis, the salience of features has two components: intensity and diagnosticity. The intensity of a feature is determined by perceptual and cognitive factors that are relatively stable across contexts. The diagnostic value of a feature is determined by the prevalence of the classifications that are based on it, which change with the context. The effects of context on similarity, therefore, are treated as changes in the diagnostic value of features induced by the respective changes in the grouping of the objects.

This account was supported by the experimental finding that changes in grouping (produced by the replacement or addition of objects) lead to corresponding changes in the similarity of the objects. These results shed light on the dynamic interplay between similarity and classification. It is generally assumed that classifications are determined by similarities among the objects. The preceding discussion supports the converse hypothesis: that the similarity of objects is modified by the manner in which they are classified. Thus, similarity has two faces: causal and derivative. It serves as a basis for the classification of objects, but it is also influenced by the adopted classification. The diagnosticity principle which underlies this process may provide a key to the analysis of the effects of context on similarity.

Discussion

In this section we relate the present development to the representation of objects in terms of clusters and trees, discuss the concepts of prototypicality and family resemblance, and comment on the relation between similarity and metaphor.

Features, Clusters, and Trees

There is a well-known correspondence between features or properties of objects and the classes to which the objects belong. A red flower, for example, can be characterized as having the feature "red," or as being a member of the class of red objects. In this manner we associate with every feature in Φ the class of objects in Δ which possesses that feature. This correspondence between features and classes provides a direct link between the present theory and the clustering approach to the representation of proximity data.

In the contrast model, the similarity between objects is expressed as a function of their common and distinctive features. Relations among overlapping sets are often represented in a Venn diagram (see figure 1.1). However, this representation becomes cumbersome when the number of objects exceeds four or five. To obtain useful graphic representations of the contrast model; two alternative simplifications are entertained.

First, suppose the objects under study are all equal in prominence, that is, $f(A) = f(B)$ for all a, b in Δ. Although this assumption is not strictly valid in general, it may serve as a reasonable approximation in certain contexts. Assuming feature additivity and symmetry, we obtain

$$
\begin{aligned}
S(a,b) &= \theta f(A \cap B) - f(A - B) - f(B - A) \\
&= \theta f(A \cap B) + 2f(A \cap B) - f(A - B) - f(B - A) - 2f(A \cap B) \\
&= (\theta + 2) f(A \cap B) - f(A) - f(B) \\
&= \lambda f(A \cap B) + \mu,
\end{aligned}
$$

since $f(A) = f(B)$ for all a, b in Δ. Under the present assumptions, therefore, similarity between objects is a linear function of the measure of their common features.

Since f is an additive measure, $f(A \cap B)$ is expressible as the sum of the measures of all the features that belong to both a and b. For each subset Λ of Δ, let $\Phi(\Lambda)$ denote the set of features that are shared by all objects in Λ, and are not shared by any object that does not belong to Λ. Hence,

$$\begin{aligned} S(a,b) &= \lambda f(A \cap B) + \mu \\ &= \lambda\left(\sum f(X)\right) + \mu \quad X \in A \cap B \\ &= \lambda\left(\sum f(\Phi(\Lambda))\right) + \mu \quad \Lambda \supset \{a,b\}. \end{aligned}$$

Since the summation ranges over all subsets of Δ that include both a and b, the similarity between objects can be expressed as the sum of the weights associated with all the sets that include both objects.

This form is essentially identical to the additive clustering model proposed by Shepard and Arabie[4]. These investigators have developed a computer program, ADCLUS, which selects a relatively small collection of subsets and assigns weight to each subset so as to maximize the proportion of (similarity) variance accounted for by the model. Shepard and Arabie[4] applied ADCLUS to several studies including Shepard, Kilpatric, and Cunningham's (1975) on judgments of similarity between the integers 0 through 9 with respect to their abstract numerical character. A solution with 19 subsets accounted for 95% of the variance. The nine major subsets (with the largest weights) are displayed in table 1.1 along with a suggested interpretation. Note that all the major subsets are readily interpretable, and they are overlapping rather than hierarchical.

Table 1.1
ADCLUS Analysis of the Similarities among the Integers 0 through 9 (from Shepard & Arabie[4])

Rank	Weight	Elements of subset	Interpretation of subset
1st	.305	2 4 8	powers of two
2nd	.288	6 7 8 9	large numbers
3rd	.279	3 6 9	multiples of three
4th	.202	0 1 2	very small numbers
5th	.202	1 3 5 7 9	odd numbers
6th	.175	1 2 3	small nonzero numbers
7th	.163	5 6 7	middle numbers (largish)
8th	.160	0 1	additive and multiplicative identities
9th	.146	0 1 2 3 4	smallish numbers

The above model expresses similarity in terms of common features only. Alternatively, similarity may be expressed exclusively in terms of distinctive features. It has been shown by Sattath[5] that for any symmetric contrast model with an additive measure f, there exists a measure g defined on the same feature space such that

$$S(a,b) = \theta f(A \cap B) - f(A - B) - f(B - A)$$
$$= \lambda - g(A - B) - g(B - A) \quad \text{for some } \lambda > 0.$$

This result allows a simple representation of dissimilarity whenever the feature space Φ is a tree (i.e., whenever any three objects in Δ can be labeled so that $A \cap B = A \cap C \subset B \cap C$). Figure 1.6 presents an example of a feature tree, constructed by Sattath and Tversky (in press) from judged similarities between lowercase letters, obtained by Kuennapas and Janson (1969). The major branches are labeled to facilitate the interpretation of the tree.

Each (horizontal) arc in the graph represents the set of features shared by all the objects (i.e., letters) that follow from that arc, and the arc length corresponds to the measure of that set. The features of an object are the features of all the arcs which lead to that object, and its measure is its (horizontal) distance to the root. The tree distance between objects a and b is the (horizontal) length of the path joining them, that is, $f(A - B) + f(B - A)$. Hence, if the contrast model holds, $\alpha = \beta$, and Φ is a tree, then dissimilarity (i.e., $-S$) is expressible as tree distance.

A feature tree can also be interpreted as a hierarchical clustering scheme where each arc length represents the weight of the cluster consisting of all the objects that follow from that arc. Note that the tree in figure 1.6 differs from the common hierarchical clustering tree in that the branches differ in length. Sattath and Tversky (in press) describe a computer program, ADDTREE, for the construction of additive feature trees from similarity data and discuss its relation to other scaling methods.

It follows readily from the above discussion that if we assume both that the feature set Φ is a tree, and that $f(A) = f(B)$ for all a, b in Δ, then the contrast model reduces to the well-known hierarchical clustering scheme. Hence, the additive clustering model (Shepard & Arabie)[4], the additive similarity tree (Sattath & Tversky, in press), and the hierarchical clustering scheme (Johnson, 1967) are all special cases of the contrast model. These scaling models can thus be used to discover the common and distinctive features of the objects under study. The present development, in turn, provides theoretical foundations for the analysis of set-theoretical methods for the representation of proximities.

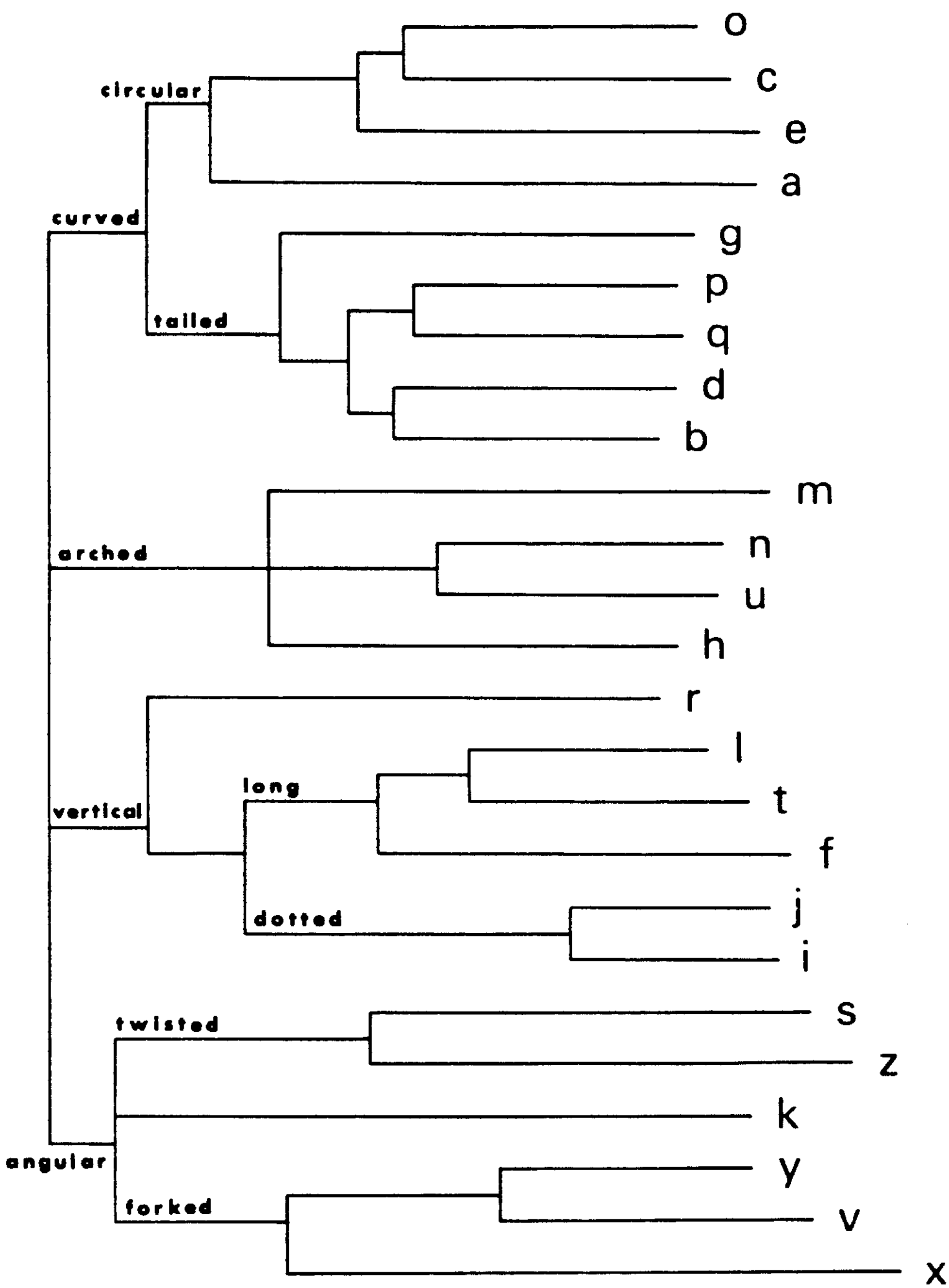

Figure 1.6
The representation of letter similarity as an additive (feature) tree. From Sattath and Tversky (in press).

Similarity, Prototypicality, and Family Resemblance

Similarity is a relation of proximity that holds between two objects. There exist other proximity relations such as prototypicality and representativeness that hold between an object and a class. Intuitively, an object is prototypical if it exemplifies the category to which it belongs. Note that the prototype is not necessarily the most typical or frequent member of its class. Recent research has demonstrated the importance of prototypicality or representativeness in perceptual learning (Posner & Keele, 1968; Reed, 1972), inductive inference (Kahneman & Tversky, 1973), semantic memory (Smith, Rips, & Shoben, 1974), and the formation of categories (Rosch & Mervis, 1975). The following discussion analyzes the relations of prototypicality and family resemblance in terms of the present theory of similarity.

Let $P(a, \Lambda)$ denote the (degree of) prototypicality of object a with respect to class Λ, with cardinality n, defined by

$$P(a, \Lambda) = p_n\Big(\lambda \sum f(A \cap B) - \sum (f(A - B) + f(B - A))\Big),$$

where the summations are over all b in Λ. Thus, $P(a, \Lambda)$ is defined as a linear combination (i.e., a contrast) of the measures of the features of a that are shared with the elements of Λ and the features of a that are not shared with the elements of Λ. An element a of Λ is a *prototype* if it maximizes $P(a, \Lambda)$. Note that a class may have more than one prototype.

The factor p_n reflects the effect of category size on prototypicality, and the constant λ determines the relative weights of the common and the distinctive features. If $p_n = 1/n$, $\lambda = \theta$, and $\alpha = \beta = 1$, then $P(a, \Lambda) = 1/n \sum S(a, b)$ (i.e., the prototypicality of a with respect to Λ equals the average similarity of a to all members of Λ). However, in line with the focusing hypotheses discussed earlier, it appears likely that the common features are weighted more heavily in judgments of prototypicality than in judgments of similarity.

Some evidence concerning the validity of the proposed measure was reported by Rosch and Mervis (1975). They selected 20 objects from each one of six categories (furniture, vehicle, fruit, weapon, vegetable, clothing) and instructed subjects to list the attributes associated with each one of the objects. The prototypicality of an object was defined by the number of attributes or features it shared with each member of the category. Hence, the prototypicality of a with respect to Λ was defined by $\sum N(a, b)$, where $N(a, b)$ denotes the number of attributes shared by a and b, and the summation ranges over all b in Λ. Clearly, the measure of prototypicality employed by Rosch and Mervis (1975) is a special case of the proposed measure, where λ is large and $f(A \cap B) = N(a, b)$.

These investigators also obtained direct measures of prototypicality by instructing subjects to rate each object on a 7-point scale according to the extent to which it fits the "idea or image of the meaning of the category." The rank correlations between these ratings and the above measure were quite high in all categories: furniture, .88; vehicle, .92; weapon, .94; fruit, .85; vegetable, .84; clothing, .91. The rated prototypicality of an object in a category, therefore, is predictable by the number of features it shares with other members of that category.

In contrast to the view that natural categories are definable by a conjunction of critical features, Wittgenstein (1953) argued that several natural categories (e.g., a game) do not have any attribute that is shared by all their members, and by them alone. Wittgenstein proposed that natural categories and concepts are commonly characterized and understood in terms of family resemblance, that is, a network of similarity relations that link the various members of the class. The importance of family resemblance in the formation and processing of categories has been effectively underscored by the work of Rosch and her collaborators (Rosch, 1973; Rosch & Mervis, 1975; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). This research demonstrated that both natural and artificial categories are commonly perceived and organized in terms of prototypes, or focal elements, and some measure of proximity from the prototypes. Furthermore, it lent substantial support to the claim that people structure their world in terms of basic semantic categories that represent an optimal level of abstraction. Chair, for example, is a basic category; furniture is too general and kitchen chair is too specific. Similarly, car is a basic category; vehicle is too general and sedan is too specific. Rosch argued that the basic categories are selected so as to maximize family resemblance—defined in terms of cue validity.

The present development suggests the following measure for family resemblance, or category resemblance. Let Λ be some subset of Δ with cardinality n. The category resemblance of Λ denoted $R(\Lambda)$ is defined by

$$R(\Lambda) = r_n\Big(\lambda \sum f(A \cap B) - \sum (f(A - B) + f(B - A))\Big),$$

the summations being over all a, b in Λ. Hence, category resemblance is a linear combination of the measures of the common and the distinctive features of all pairs of objects in that category. The factor r_n reflects the effect of category size on category resemblance, and the constant λ determines the *relative* weight of the common and the distinctive features. If $\lambda = \theta$, $\alpha = \beta = 1$, and $r_n = 2/n(n-1)$, then

$$R(\Lambda) = \frac{\sum S(a,b)}{\binom{n}{2}},$$

the summation being over all a, b in Λ; that is, category resemblance equals average similarity between all members of Λ. Although the proposed measure of family resemblance differs from Rosch's, it nevertheless captures her basic notion that family resemblance is highest for those categories which "have the most attributes common to members of the category and the least attributes shared with members of other categories" (Rosch et al., 1976, p. 435).

The maximization of category resemblance could be used to explain the formation of categories. Thus, the set Λ rather than Γ is selected as a natural category whenever $R(\Lambda) > R(\Gamma)$. Equivalently, an object a is added to a category Λ whenever $R(\{\Lambda \cup a\}) > R(\Lambda)$. The fact that the preferred (basic) categories are neither the most inclusive nor the most specific imposes certain constraints on r_n.

If $r_n = 2/n(n-1)$ then $R(\Lambda)$ equals the average similarity between all members of Λ. This index leads to the selection of minimal categories because average similarity can generally be increased by deleting elements. The average similarity between sedans, for example, is surely greater than the average similarity between cars; nevertheless, car rather than sedan serves as a basic category. If $r_n = 1$ then $R(\Lambda)$ equals the sum of the similarities between all members of Λ. This index leads to the selection of maximal categories because the addition of objects increases total similarity, provided S is nonnegative.

In order to explain the formation of intermediate-level categories, therefore, category resemblance must be a compromise between an average and a sum. That is, r_n must be a decreasing function of n that exceeds $2/n(n-1)$. In this case, $R(\Lambda)$ increases with category size whenever average similarity is held constant, and vice versa. Thus, a considerable increase in the extension of a category could outweigh a small reduction in average similarity.

Although the concepts of similarity, prototypicality, and family resemblance are intimately connected, they have not been previously related in a formal explicit manner. The present development offers explications of similarity, prototypicality, and family resemblance within a unified framework, in which they are viewed as contrasts, or linear combinations, of the measures of the appropriate sets of common and distinctive features.

Similes and Metaphors

Similes and metaphors are essential ingredients of creative verbal expression. Perhaps the most interesting property of metaphoric expressions is that despite their novelty and nonliteral nature, they are usually understandable and often informative. For example, the statement that Mr. X resembles a bulldozer is readily understood as saying that Mr. X is a gross, powerful person who overcomes all obstacles in getting a job done. An adequate analysis of connotative meaning should account for man's ability to interpret metaphors without specific prior learning. Since the message conveyed by such expressions is often pointed and specific, they cannot be explained in terms of a few generalized dimensions of connotative meaning, such as evaluation or potency (Osgood, 1962). It appears that people interpret similes by scanning the feature space and selecting the features of the referent that are applicable to the subject (e.g., by selecting features of the bulldozer that are applicable to the person). The nature of this process is left to be explained.

There is a close tie between the assessment of similarity and the interpretation of metaphors. In judgments of similarity one assumes a particular feature space, or a frame of reference, and assesses the quality of the match between the subject and the referent. In the interpretation of similes, one assumes a resemblance between the subject and the referent and searches for an interpretation of the space that would maximize the quality of the match. The same pair of objects, therefore, can be viewed as similar or different depending on the choice of a frame of reference.

One characteristic of good metaphors is the contrast between the prior, literal interpretation, and the posterior, metaphoric interpretation. Metaphors that are too transparent are uninteresting; obscure metaphors are uninterpretable. A good metaphor is like a good detective story. The solution should not be apparent in advance to maintain the reader's interest, yet it should seem plausible after the fact to maintain coherence of the story. Consider the simile "An essay is like a fish." At first, the statement is puzzling. An essay is not expected to be fishy, slippery, or wet. The puzzle is resolved when we recall that (like a fish) an essay has a head and a body, and it occasionally ends with a flip of the tail.

Notes

This paper benefited from fruitful discussions with Y. Cohen, I. Gati, D. Kahneman, L. Sjöberg, and S. Sattath.

1. To derive feature additivity from qualitative assumptions, we must assume the axioms of an extensive structure and the compatibility of the extensive and the conjoint scales; see Krantz et al. (1971, Section 10.7).

2. The subjects in all our experiments were Israeli college students, ages 18–28. The material was presented in booklets and administered in a group setting.

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

4. Shepard, R. N., & Arabie, P. Additive cluster analysis of similarity data. *Proceedings of the U.S.–Japan Seminar on Theory, Methods, and Applications of Multidimensional Scaling and Related Techniques*. San Diego, August 1975.

5. Sattath, S. *An equivalence theorem*. Unpublished note, Hebrew University, 1976.

References

Beals, R., Krantz, D. H., & Tversky, A. Foundations of multidimensional scaling. *Psychological Review*, 1968, *75*, 127–142.

Bush, R. R., & Mosteller, F. A model for stimulus generalization and discrimination. *Psychological Review*, 1951, *58*, 413–423.

Carroll, J. D., & Wish, M. Multidimensional perceptual models and measurement methods. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception*. New York: Academic Press, 1974.

Eisler, H., & Ekman, G. A mechanism of subjective similarity. *Acta Psychologica*, 1959, *16*, 1–10.

Garner, W. R. *The processing of information and structure*. New York: Halsted Press, 1974.

Gibson, E. *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts, 1969.

Goldmeier, E. Similarity in visually perceived forms. *Psychological Issues*, 1972, *8*, 1–136.

Gregson, R. A. M. *Psychometrics of similarity*. New York: Academic Press, 1975.

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

Jakobson, R., Fant, G. G. M., & Halle, M. *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge, Mass.: MIT Press, 1961.

Johnson, S. C. Hierarchical clustering schemes. *Psychometrika*, 1967, *32*, 241–254.

Kahneman, D., & Tversky, A. On the psychology of prediction. *Psychological Review*, 1973, *80*, 237–251.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. *Foundations of measurement* (Vol. 1). New York: Academic Press, 1971.

Krantz, D. H., & Tversky, A. Similarity of rectangles: An analysis of subjective dimensions. *Journal of Mathematical Psychology*, 1975, *12*, 4–34.

Kuennapas, T., & Janson, A. J. Multidimensional similarity of letters. *Perceptual and Motor Skills*, 1969, *28*, 3–12.

Neisser, U. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1967.

Osgood, C. E. Studies on the generality of affective meaning systems. *American Psychologist*, 1962, *17*, 10–28.

Posner, M. I., & Keele, S. W. On the genesis of abstract ideas. *Journal of Experimental Psychology*, 1968, *77*, 353–363.

Reed, S. K. Pattern recognition and categorization. *Cognitive Psychology*, 1972, *3*, 382–407.

Restle, F. A metric and an ordering on sets. *Psychometrika*, 1959, *24*, 207–220.

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

Rosch, E. On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language*. New York: Academic Press, 1973.

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

Rosch, E., & Mervis, C. B. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 1975, *7*, 573–603.

Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Braem, P. Basic objects in natural categories. *Cognitive Psychology*, 1976, *8*, 382–439.

Rothkopf, E. Z. A measure of stimulus similarity and errors in some paired-associate learning tasks. *Journal of Experimental Psychology*, 1957, *53*, 94–101.

Sattath, S., & Tversky, A. Additive similarity trees. *Psychometrika*, in press.

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.

Shepard, R. N., Kilpatric, D. W., & Cunningham, J. P. The internal representation of numbers. *Cognitive Psychology*, 1975, *7*, 82–138.

Smith, E. E., Rips, L. J., & Shoben, E. J. Semantic memory and psychological semantics. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8). New York: Academic Press, 1974.

Smith, E. E., Shoben, E. J., & Rips, L. J. Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, 1974, *81*, 214–241.

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

Tversky, A. Elimination by aspects: A theory of choice. *Psychological Review*, 1972, *79*, 281–299.

Tversky, A., & Gati, I. Studies of similarity. In E. Rosch & B. Lloyd (Eds.), *On the nature and principle of formation of categories*. Hillsdale, N.J.: Erlbaum, in press.

Tversky, A., & Krantz, D. H. The dimensional representation and the metric structure of similarity data. *Journal of Mathematical Psychology*, 1970, *7*, 572–597.

von Neumann, J., & Morgenstern, O. *Theory of games and economic behavior*. Princeton, N.J.: Princeton University Press, 1947.

Wish, M. A model for the perception of Morse Code-like signals. *Human Factors*, 1967, *9*, 529–540.

Wittgenstein, L. *Philosophical investigations*. New York: Macmillan, 1953.

Appendix: An Axiomatic Theory of Similarity

Let $\Delta = \{a, b, c, \ldots\}$ be a collection of objects characterized as sets of features, and let A, B, C, denote the sets of features associated with a, b, c, respectively. Let $s(a, b)$ be an ordinal measure of the similarity of a to b, defined for all distinct a, b in Δ. The present theory is based on the following five axioms. Since the first three axioms are discussed in the paper, they are merely restated here; the remaining axioms are briefly discussed.

1. MATCHING: $s(a, b) = F(A \cap B, A - B, B - A)$ where F is some real-valued function in three arguments.

2. MONOTONICITY: $s(a, b) \geq s(a, c)$ whenever $A \cap B \supset A \cap C$, $A - B \subset A - C$, and $B - A \subset C - A$. Moreover, if either inclusion is proper then the inequality is strict.

Let Φ be the set of all features associated with the objects of Δ, and let X, Y, Z, etc. denote subsets of Φ. The expression $F(X, Y, Z)$ is defined whenever there exist a, b in Δ such that $A \cap B = X$, $A - B = Y$, and $B - A = Z$, whence $s(a, b) = F(X, Y, Z)$. Define $V \simeq W$ if one or more of the following hold for some X, Y, Z: $F(V, Y, Z) = F(W, Y, Z)$, $F(X, V, Z) = F(X, W, Z)$, $F(X, Y, V) = F(X, Y, W)$. The pairs (a, b) and (c, d) *agree* on one, two, or three components, respectively, whenever one, two, or three of the following hold: $(A \cap B) \simeq (C \cap D)$, $(A - B) \simeq (C - D)$, $(B - A) \simeq (D - C)$.

3. INDEPENDENCE: Suppose the pairs (a, b) and (c, d), as well as the pairs (a', b') and (c', d'), agree on the same two components, while the pairs (a, b) and (a', b'), as well as the pairs (c, d) and (c', d'), agree on the remaining (third) component. Then

$$s(a, b) \geq s(a', b') \quad \text{iff} \quad s(c, d) \geq s(c', d').$$

4. SOLVABILITY:

(i) For all pairs (a, b), (c, d), (e, f), of objects in Δ there exists a pair (p, q) which agrees with them, respectively, on the first, second, and third component, that is, $P \cap Q \simeq A \cap B$, $P - Q \simeq C - D$, and $Q - P \simeq F - E$.

(ii) Suppose $s(a, b) > t > s(c, d)$. Then there exist e, f with $s(e, f) = t$, such that if (a, b) and (c, d) agree on one or two components, then (e, f) agrees with them on these components.

(iii) There exist pairs (a, b) and (c, d) of objects in Δ that do not agree on any component.

Unlike the other axioms, solvability does not impose constraints on the similarity order; it merely asserts that the structure under study is sufficiently rich so that certain equations can be solved. The first part of axiom 4 is analogous to the existence of a factorial structure. The second part of the axiom implies that the range of s is a real interval: There exist objects in Δ whose similarity matches any real value that is bounded by two similarities. The third part of axiom 4 ensures that all arguments of F are essential.

Let Φ_1, Φ_2, and Φ_3 be the sets of features that appear, respectively, as first, second, or third arguments of F. (Note that $\Phi_2 = \Phi_3$.) Suppose X and X′ belong to Φ_1, while Y and Y′ belong to Φ_2. Define $(X, X')_1 \simeq (Y, Y')_2$ whenever the two intervals are matched, that is, whenever there exist pairs (a, b) and (a', b') of equally similar objects in Δ which agree on the third factor. Thus, $(X, X')_1 \simeq (Y, Y')_2$ whenever

$$s(a, b) = F(X, Y, Z) = F(X', Y', Z) = s(a', b').$$

This definition is readily extended to any other pair of factors. Next, define $(V, V')_i \simeq (W, W')_i$, $i = 1, 2, 3$ whenever $(V, V')_i \simeq (X, X')_j \simeq (W, W')_i$, for some $(X, X')_j$, $j \neq i$. Thus, two intervals on the same factor are equivalent if both match the same interval on another factor. The following invariance axiom asserts that if two intervals are equivalent on one factor, they are also equivalent on another factor.

5. INVARIANCE: Suppose V, V', W, W' belong to both Φ_i and Φ_j, $i, j = 1, 2, 3$. Then

$$(V, V')_i \simeq (W, W')_i \quad \text{iff} \quad (V, V')_j \simeq (W, W')_j.$$

REPRESENTATION THEOREM

Suppose axioms 1–5 hold. Then there exist a similarity scale S and a nonnegative scale f such that for all a, b, c, d in Δ

(i) $S(a,b) \geq S(c,d)$ iff $s(a,b) \geq s(c,d)$,

(ii) $S(a,b) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A)$, for some $\theta, \alpha, \beta \geq 0$.

(iii) f and S are interval scales.

While a self-contained proof of the representation theorem is quite long, the theorem can be readily reduced to previous results.

Recall that Φ_i is the set of features that appear as the *i*th argument of F, and let $\Psi_i = \Phi_i/\simeq$, $i = 1, 2, 3$. Thus, Ψ_i is the set of equivalence classes of Φ_i with respect to $\simeq$. It follows from axioms 1 and 3 that each Ψ_i is well defined, and it follows from axiom 4 that $\Psi = \Psi_1 \times \Psi_2 \times \Psi_3$ is equivalent to the domain of F. We wish to show that Ψ, ordered by F, is a three-component, additive conjoint structure, in the sense of Krantz, Luce, Suppes, and Tversky (1971, Section 6.11.1).

This result, however, follows from the analysis of decomposable similarity structures, developed by Tversky and Krantz (1970). In particular, the proof of part (c) of theorem 1 in that paper implies that, under axioms 1, 3, and 4, there exist nonnegative functions f_i defined on Ψ_i, $i = 1, 2, 3$, so that for all a, b, c, d in Δ

$$s(a,b) \geq s(c,d) \quad \text{iff} \quad S(a,b) \geq S(c,d)$$

where

$$S(a,b) = f_1(A \cap B) + f_2(A - B) + f_3(B - A),$$

and f_1, f_2, f_3 are interval scales with a common unit.

According to axiom 5, the equivalence of intervals is preserved across factors. That is, for all V, V', W, W' in $\Phi_i \cap \Phi_j$, $i, j = 1, 2, 3$,

$$f_i(V) - f_i(V') = f_i(W) - f_i(W') \quad \text{iff} \quad f_j(V) - f_j(V') = f_j(W) - f_j(W').$$

Hence by part (i) of theorem 6.15 of Krantz et al. (1971), there exist a scale f and constants θ_i such that $f_i(X) = \theta_i f(X)$, $i = 1, 2, 3$. Finally, by axiom 2, S increases in f_1 and decreases in f_2 and f_3. Hence, it is expressible as

$$S(a,b) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A),$$

for some nonnegative constants θ, α, β.