Psychological Review
2000, Vol. 107, No. 2, 345–357

0033-295X/00/$5.00 DOI: 10.1037//0033-295X.107.2.345

The Role of Expectations in Comparisons

Ilana Ritov
Hebrew University of Jerusalem

Expectations are proposed as a unifying framework for explaining a variety of findings on comparative judgment. Expectations may pertain to either the features of each separate object or their simultaneous occurrence in all relevant objects. Two experiments demonstrate that inducing expectations by varying the frequency with which a component occurs, as well as the frequency with which it is shared by the compared objects, results in augmented weighting of unexpected components. Expectations can arise not only from experienced frequencies but also from an inferred set of alternatives. Features of cohesive stimuli are shown to be more expected than features of noncohesive stimuli, in the latter sense. The relatively higher weighting of distinctive features in cohesive stimuli, as well as the higher weight assigned to structurally aligned differences, could thus reflect modifications in feature expectedness. Further experiments show how task-dependent weighting interacts with expectations to produce asymmetries between similarity and difference judgments. Finally, it is argued that changes in expectations regarding objects' attributes can explain context effects across different domains, including some types of preference reversals.

Comparisons form a major part of humans' cognitive endeavors. We compare objects to evaluate relations between them, such as similarity or dissimilarity; we compare objects in forming classifications; and we also compare objects to establish ordering between them, such as preference order on which choice may be based. Indeed, recent research on both decision making and similarity judgment has emphasized the crucial role of the comparative process (Medin, Goldstone, & Gentner, 1993; Medin, Goldstone, & Markman, 1995; Shafir, 1993; Simonson & Tversky, 1992). The fundamental building blocks of a comparative process are the commonalities and differences between the objects. In similarity judgment, for example, Tversky's (1977) contrast model formally expressed perceived similarity between two objects as a weighted linear combination, or a contrast, between the features shared by the two objects and those features that distinguish them. The determinants of the relative weight of shared and distinctive components are critical to understanding any comparison process.

Earlier studies demonstrated that the relative weight of shared and distinctive components is affected by the nature of the task. Thus, for example, shared features are weighted more heavily in similarity judgment, whereas distinctive features carry relatively more weight in judging dissimilarity or distance between objects (Gati & Tversky, 1984; Tversky & Gati, 1978). Establishing preference and assessing the distance between objects involve relatively higher weighting of distinctive components (Gati & Ashkenazi, 1996; Houston, Sherman, & Baker, 1991).

Beyond the influence of the task in directing one's attention to shared or distinctive components, the nature of the objects themselves may affect the weighting of features in the comparative process. Thus, in comparisons of pictorial stimuli, distinctive features are more salient, whereas in comparisons of verbal descriptions, shared features carry more weight (Gati & Tversky, 1984). Ritov, Gati, and Tversky (1990) proposed that the effect of stimulus cohesiveness may explain the modality effect in similarity judgment. In their study, reducing the cohesiveness of the objects decreased the relative weight of distinctive components.

More recent research has emphasized the role of structural alignment in similarity judgment (Gentner & Markman, 1994, 1995; Markman & Gentner, 1993). According to this view, judgment follows a process by which a structural alignment is established. Thus, the hierarchically structured representations of the objects are aligned in a way that maximizes the structural matches between them. On the basis of the established alignment, distinctive features can be classified as alignable or nonalignable. In the former case, the feature of one of the objects has a corresponding but different element in the other object, whereas in the latter case, the distinctive element does not have an aligned match in the other object. The aforementioned studies showed that in similarity judgment, alignable distinctive features are more heavily weighted than nonalignable ones. Markman and Medin (1995) showed a similar effect in choice. In their study, both the selection of options and the justification provided reflected the prominent role of comparable properties (alignable differences).

The present study integrated some of the aforementioned findings. Factors pertaining to the compared objects themselves (e.g., modality, cohesiveness, and structural alignment) involve changes in the perceiver's expectations regarding the inclusion of components in one or both of the compared objects. The realization or violation of those expectations affects feature weighting, thus causing the documented effects discussed earlier.

Expectations play a major role in different domains. As a general rule, the weight of an observation in forming a judgment is inversely related to one's prior expectations. Thus, for example, the emotional impact of an event is more poignant the more unexpected it is (Kahneman & Miller, 1986). Winning a small

I thank Jonathan Baron, Dedre Gentner, Douglas Medin, and Barbara Mellers for their helpful comments.

Correspondence concerning this article should be addressed to Ilana Ritov, School of Education, Hebrew University of Jerusalem, Jerusalem 91905, Israel. Electronic mail may be sent to msiritov@mscc.huji.ac.il.

amount of money in one gamble may be more elating than winning a larger amount of money in another gamble, if the smaller win is less expected than the larger one (Mellers, Schwartz, Ho, & Ritov, 1997). Even in the domain of public opinion, concern about social problems such as inflation and unemployment is amplified when the present level of the problem does not meet the projected level of the problem (Loewenstein & Mather, 1990).

Whereas in some cases expectations are stated explicitly (as, e.g., a projection of the unemployment level), in other situations expectations are often created by frequency of prior exposure to relevant stimuli. In the domain of category learning, the classification of new stimuli into previously learned categories was shown to be disproportionally affected by features of rare categories relative to features of more frequently occurring ones (Medin & Edelson, 1988). A recent model explaining this "inverse base rate" effect (Kruschke, 1996) is based on the assumption that rare categories are encoded primarily in terms of their distinctive features. The reason for differential encoding of frequent and rare categories lies in the temporal order of encounter: Rare categories are learned after the common features of the frequent categories have already been encoded. Possibly, learners attend more to the distinctive features of the rare categories because those features are less expected than the already encoded features shared by rare and frequent categories.

What factors determine people's expectations about objects' components? Clearly, as indicated earlier, prior experience with relevant objects, either in a natural setting or in the laboratory plays an important role. However, expectations are often formed during the judgment process itself on the basis of an evoked set of alternatives. This idea goes back to the gestalt concept of pattern goodness and was further developed in Garner's (1974) research on the processing of information and structure. According to this theory, as an organism perceives a stimulus, an inferred set of stimuli, including the target stimulus, is constructed. By this account, the stimulus properties are significant only insofar as they are perceived as the dimensions that define a set of stimuli. In the absence of an explicitly defined set, the perceiver infers it. Consequently, the inferred set reflects one's encoding of the stimulus properties, which are based on one's expectations regarding the possible variations in these properties. Kahneman and Miller (1986) extended the notion of inferred set beyond the domain of perception and argued that most judgments involve an implicit comparison between the stimulus and its spontaneously evoked set of alternatives, or counterfactuals. Although in many cases the generation and use of counterfactuals involve conscious thought processes, which are often motivated by affective goals, in many other cases counterfactuals are automatically and unconsciously generated (Kahneman, 1995).

Garner's (1974) research on structure processing demonstrated the link between the goodness of a pattern and the size of the perceived or inferred subset of related patterns. According to this theory, pattern goodness is associated with a correlational structure, in which some properties are redundant. The subset of related patterns reflects this redundancy. Consequently, better patterns evoke smaller sets. For example, when participants were presented with a set of 90 dot patterns, subjective goodness rating of the patterns was negatively correlated with the size of the group in which each pattern was placed in a free-classification task. Furthermore, when predictability was assessed directly, by requiring participants to guess the location of a missing dot, good patterns, from small subsets, resulted in much higher accuracy (Bear, 1973). Thus, properties of good patterns appear to be more expected than properties of poor patterns.

The essence and role of structure is a momentous issue in cognitive psychology, which extends far beyond the scope of this article. The present discussion focuses primarily on the role of structure as a determinant of expectations. Although the research described above (Garner, 1974) highlighted the correlations between properties, research on structure in verbal processing and in visual-perceptual analogy offered the notion of interconnected predicates governed by higher order relations. This type of structure was termed *systematicity* (Bowdle & Gentner, 1997; Gentner, 1983; Markman & Gentner, 1993; Medin et al., 1993), and it is closely related to the concept of coherence used in theories of text processing (Shank & Abelson, 1977). Although Ritov et al. (1990) did not define the term *cohesiveness,* it is essentially congruous to the concept of systematicity. In particular, cohesive stimuli are clearly characterized by a complex, hierarchical system of constraints. For example, in a picture of an office scene, there is no place for a bed. Furthermore, a chair cannot appear larger than a desk, and a bookcase cannot be positioned under the desk. If the same objects were presented in a noncohesive array, these constraints would not necessarily apply. In that sense, the inferrable sets of cohesive stimuli are smaller, and the stimuli's components are more expected than the components of noncohesive stimuli.

The aforementioned findings suggest that in explicit as well as implicit comparisons, the weight of a component is augmented by its surprisingness. Similarly, the absence of a component may be augmented by its expectedness. Expectations in comparative judgment, however, may involve not only the components of each separate object but also the possibility or likelihood that a certain component is included in both objects. Consider, for example, a comparison of Objects A and B. Although a Component x may rarely appear in objects like A and B, expectations are that, to the extent that it appears at all, it will be included in both A and B. In this case, the fact that x is indeed shared by both objects is not particularly surprising. If, in contrast, x is actually included only in A, its absence from B can be regarded, to some extent, as surprising. The impact of x as a shared component relative to its impact as a distinctive one will consequently be affected by these expectations. More generally, the higher the expectations that a feature will be included in both members of a compared pair, the lower its weight as a shared feature relative to its weight as a distinctive feature.

Consider, for example, a consumer who is considering buying a camera. The feature "all-weather protection" is rather rare among point-and-shoot cameras. It is generally unrelated to other camera features, except price: It is mostly available in upper price range cameras. If the cameras were compared in random pairs, one would not expect all-weather protection to be available in any of the cameras, much less so in both of them. If, however, the comparison pairs were made up of cameras in the same price level, then the availability of the feature in one member of the pair would raise expectations that it be available in the other member as well. Consequently, having all-weather protection available in both cameras would likely be more highly weighted in the former comparison procedure than in the latter.

Expectations concerning shared and distinctive features could be based on frequencies, as in the preceding example, but they also could stem from the relations between the compared objects. In particular, if A and B are structurally aligned and x, a component of A, is related to the common structure, then expectations concerning the inclusion of x in B may be high. If, in contrast, x is not related to the common structure, its inclusion in A does not raise the expectation that it be included in B as well. For example, in a comparison of pictures of two office scenes, if a bookcase appears in one of the scenes, it increases the expectation for a bookcase to be included in the second scene as well. However, a bird sitting on the window of one of these offices would not necessarily raise expectations that the other office picture will also feature a bird.

Expectations pertaining to a particular object feature in comparative judgment may be induced both by factors related to the single object, such as cohesiveness, and by factors related to the pair, such as structural alignment. The two sources of expectations may affect expectations synonymously. Thus, a distinctive, rare component of a cohesive stimulus that has an alignable match in the compared stimulus would be heavily weighted for both reasons. Indeed, one may argue that, in demonstrating the increased weighting of distinctive components in cohesive stimuli, Ritov et al. (1990) used mostly structurally aligned stimulus pairs. Along the same line, a distinctive component that is an ordinary element of a cohesive stimulus but is unrelated to the common structure would be the most compatible with expectations. As Gentner and her associates (Bowdle & Gentner, 1997; Markman & Gentner, 1993) found, such distinctive components are often hardly weighted at all. The two sources of expectations can, however, work in opposite directions. Thus, for example, a rare, distinctive component that is not related to the common structure would be surprising as a feature of the single object, but its absence from the compared object may be perfectly compatible with expectations. The question of which of these factors exert higher influence on overall expectations is considered, at present, largely as an empirical question.

In summary, the preceding discussion proposed a twofold account of the factors affecting the relative weight of shared and distinctive components in comparisons. Limiting constraints imposed by the cohesiveness of single objects and structural mapping between the compared objects are factors affecting expectations that a component will play the role of a shared feature. Such expectations, in turn, augment (or diminish) the weight of the component as shared relative to its weight as distinctive. This account further implies that if changes in expectations concerning the inclusion of components occurred independently of cohesiveness and structural alignment, due to frequency, they would still be associated with similar changes in the relative weight of shared and distinctive components.

In the first part of this article, I present two experiments documenting changes in the features' relative weight as a function of their frequency (Experiments 1 and 2). In the second part, I review earlier results demonstrating cohesiveness effects in similarity judgment. Some of the stimuli used in those studies are then used here to show the influence of cohesiveness on expectations regarding feature inclusion (Experiments 3 and 4). Finally, in the third part, I discuss a possible consequence of the impact of expectations on relative weight. A stimulus is shown to be judged as both more similar to and more different from the very same alternative (Experiments 5 and 6).

Frequency

Generally, rare components are more surprising than common ones. Assuming the inclusion of components in each stimulus is independent of their inclusion in the other, the inclusion of the same rare component in both members of a stimulus pair is especially surprising. For the same reason, exclusion of a component from only one member of the pair is less surprising if the component is rare rather than common. Two predictions follow from the argument developed here: First, the weight of rare components is higher than that of common ones. Second, the weight of shared rare components relative to their weight as distinctive ones is particularly high. In other words, an interaction between the component's frequency and the role it plays in the comparison is anticipated. These predictions were tested in Experiment 1.

Experiment 1

In this experiment, feature frequency was varied, and the weight of the critical components as shared and as distinctive features was estimated. The stimuli used here were descriptions of ice-cream cones, each containing two or three scoops of different flavors. In one condition, the cones came from a shop featuring only 5 flavors (limited selection), whereas in the other condition, the cones came from an ice-cream parlor featuring 17 different flavors (broad selection).

Naturally, the expected frequency of occurrence of a particular flavor in a list containing descriptions of ice-cream cones would be much higher in the first condition (5 available flavors) than in the second condition (17 available flavors). Hence, it was predicted that in participants' comparison of two cones, the same flavor would be more heavily weighted if it was shared by cones that came from the broad-selection parlor than if the cones were from the shop featuring only a limited selection.

To estimate the weight of shared and distinctive components, I followed the basic procedure used in previous studies of similarity judgment (Gati & Tversky, 1984; Ritov et al., 1990). The procedure involves a series of similarity judgments between pairs of stimuli. A separable component (x) is added to a basic pair of stimuli (p, q) to create two other pairs: one pair in which the component is added to both members (px, qx) and another pair in which the component is added to only one member (px, q). The increase in perceived similarity resulting from adding x to both members of a basic pair, denoted $C(x)$, provides an estimate of the weight of x as a shared feature. Similarly, the decrease in similarity resulting from adding x to only one member of the pair provides an estimate of the weight of x as a distinctive feature, denoted $D(x)$.[1]

[1] Gati and Tversky (1984) used a more complicated procedure including tests of component independence. One of these tests was exchangeability: $s(px, q) = s(p, qx)$. With the present stimuli, consisting of a list of ice-cream flavors, independence was assumed to at least roughly hold. Hence, for assessing $D(x)$, x was randomly added to one of the cones. The same procedure was followed in Experiment 2 as well. The feature-weighting procedure can be extended to substitutive, rather than additive, features (Gati & Tversky, 1984).

As noted earlier, the overall weight of x is assumed to be higher for rare features than for frequent features. Thus, I predicted a main effect of experimental condition (limited selection vs. broad selection) on feature weight. However, beyond the main effect, the preceding hypothesis implies an interaction of experimental conditions with feature role ($C[x]$ vs. $D[x]$).

Method

Participants. Forty-eight students at the University of Pennsylvania participated in this experiment. They were run individually and were paid for participating.

Stimuli and design. Three sets of pairs were used in each of the two conditions. Every set was made up of a basic pair (p, q) of two-flavored ice-cream cones and two additional pairs: In one pair, a third flavor (the additive component) was added to one of the cones, and in the other pair, the same flavor was added to both cones.

The first two sets in the two conditions were made up of totally different flavors, and the third set was identical in both conditions. The pairs comprising the identical set were presented last. Separate questionnaires were prepared for the two conditions. A list of the different flavors available at the shop (5 flavors in Condition 1 vs. 17 flavors in Condition 2) was presented at the top of each questionnaire. Nine pairs of cones followed. Participants were instructed to assess the similarity between the two members of each pair on a scale ranging from 1 (*not similar*) to 20 (*very similar*). Participants were randomly assigned to one of the two conditions.

Results

For each of the three additive components, x, $C(x)$, and $D(x)$ were computed separately for each participant, following the same procedure as that used by Ritov et al. (1990).

Thus, $C(x) = s(px, qx) - s(p, q)$ and $D(x) = s(p, q) - s(px, q)$. I then computed for each participant the average of $C(x)$ across the three sets, denoted C. Similarly, D, the average of $D(x)$ across the three sets, is a measure of the average weight of a distinctive additive component.

The mean estimates of C and D across participants in each of the conditions are presented in Table 1. Table 1 also includes the mean estimates of $C(x)$ and $D(x)$ for the third set for each of the conditions. The latter measures are especially interesting because the third set was identical in both conditions. Hence, any differences between the two groups in those measures cannot be attributed to the stimuli themselves.

Although overall the additive components (across the three sets) received greater weight in the low-frequency condition (17 flavors) than in the high-frequency condition (5 flavors), the relative size of C and D was also affected, as predicted, by frequency: Analysis of variance of C and D (as a within-subject variable) by condition (as a between-subject variable) revealed a main effect of condition, $F(1, 46) = 7.64$, $p < .01$, as well as an interaction between component type (C vs. D) and condition, $F(1, 46) = 6.21$, $p < .05$. As Table 1 shows, the same pattern held even when the components themselves were kept constant and only their frequency of occurrence changed. Analysis of variance of $C(x)$ and $D(x)$ for the third set only (by condition) yielded a significant main effect of condition, $F(1, 46) = 4.84$, $p < .05$, as well as the predicted interaction of component type by condition, $F(1, 46) = 4.44$, $p < .05$.

Table 1
Experiment 1: Mean Estimates of Component Weights as Shared and Distinctive Features

Component role	Component frequency: Frequent (5-flavors set)	Rare (17-flavors set)
Across all sets		
Shared	0.61	3.27
Distinctive	1.80	1.58
Identical set only		
Shared	1.12	3.33
Distinctive	0.87	0.83

Note. Similarity was rated on a 20-point scale ranging from 1 (*not similar*) to 20 (*very similar*). The weights of the additive component (x) in each set were computed by adding x to one or both members of a basic stimulus pair (p, q). The three similarity judgments—$s(p, q)$, $s(px, q)$, and $s(px, qx)$—were used in computing the component's weight as a shared ($C[x] = s[px, qx] - s[p, q]$) and as a distinctive ($D[x] = s[p, q] - s[px, q]$) feature.

Experiment 2

Whereas Experiment 1 used stimulus pairs whose features could reasonably be assumed to be independent, in Experiment 2, this assumption no longer held. As I argued earlier, expectations pertaining to the joint occurrence of a component can be affected not only by the general frequency of the component but also by the experienced frequency of the component's previous occurrence in the role of shared or distinctive features. If, for example, a specific component tends to occur in both compared objects (to the extent that it occurs at all), its inclusion in only one of the two objects may be regarded as surprising. Consequently, the weight of this component as a distinctive feature would be expected to be higher than its weight as a shared feature. More generally, it is predicted that the relative weight of shared and distinctive components will vary as a function of the frequency with which they previously served in these roles, independently of their overall frequency of occurrence. This prediction was tested in this experiment.

In this experiment, the relative weight of an additive component as a shared feature and a distinctive feature of the compared stimuli was estimated, as in Experiment 1. Unlike Experiment 1, however, in this experiment, the critical comparisons were preceded by a sequence of other comparisons (the training sequence). In the training sequence, the relative frequency with which a component played the role of shared versus distinctive feature and the overall frequency of its occurrence were varied orthogonally. As indicated earlier, the relative frequencies during training were predicted to affect the relative weight of the component in the critical comparisons.

Method

Participants. One hundred and thirty-eight students at Ben-Gurion University in Israel participated in this experiment. Questionnaires were run in a classroom setting.

Stimuli and design. The questionnaire presented the participants with 17 pairs of blobs representing human cells, as viewed under a

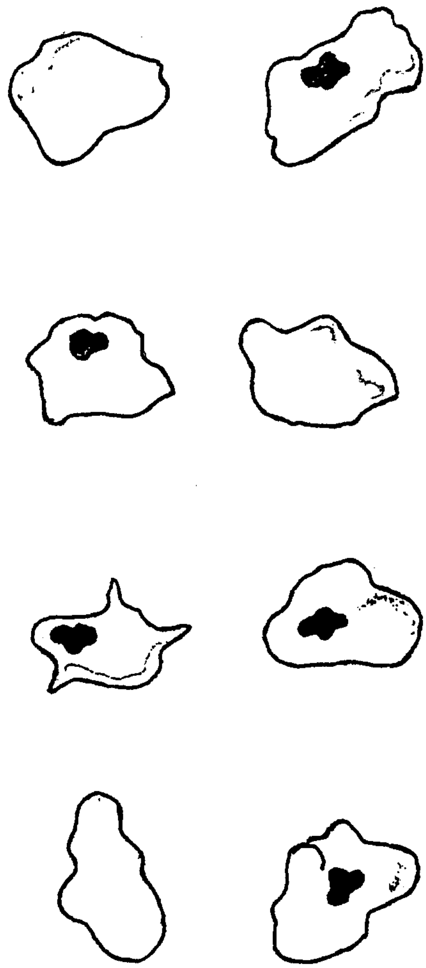

Figure 1. Experiment 2: Examples of stimulus pairs.

microscope. Some of the blobs had a small spot on them. Figure 1 depicts a number of these pairs. The participants were asked to rate the similarity between the 2 blobs in each pair on a 20-point scale. The motivation for the experiment was described as learning about the way that people make such judgments, which are often required in medical diagnosis (e.g., the diagnosis of malignant tumors).

The target judgments consisted of the 12th, 15th, and 17th pairs, and these were identical for all participants. The first 14 pairs made up the training set and were manipulated between participants. The participants were unaware of the distinction between the training and target pairs.

The spot appearing on some of the blobs served as the additive component. Two variables were orthogonally manipulated between participants in the training sequence: the frequency of the spot's occurrence in any of the blobs (8 or 16 out of 28 blobs) and the percentage of occurrences in which it appeared in both members of the pair (25% or 75% shared). In addition, two different orders of each training set were used. The participants were randomly assigned to one cell in this 2 × 2 × 2 between-subjects design. The target pairs included the basic pair of blobs (p, q); a second pair of blobs identical in shape to the basic pair, with a spot added on only 1 blob (px, q); and a third identically shaped pair with a spot added on each blob (px, qx). To preserve the "organic-form look" of the stimuli, the spots were never exactly identical.

Results

$C(x)$ and $D(x)$ were computed for each participant as they were in Experiment 1. Order did not significantly affect any of the effects; hence, the data from the two orders were collapsed. Table 2 presents the means of these measures for each group.

The relative size of C and D differed across the different training conditions. In particular, higher relative frequency of shared components during training led to higher relative weight of the additive component when it was not shared by the compared objects. Analysis of variance of C versus D (as a within-subject variable) by frequency of occurrence and percentage shared (as between-subject variables) yielded a significant interaction of C versus D with the percentage of shared occurrence during training, $F(1, 134) = 8.60$, $p < .005$. Furthermore, this effect was particularly strong when the overall frequency of the additive components during training was high, $F(1, 134) = 3.91$, $p < .05$, for the triple interaction (C vs. D × Frequency × Percentage Shared). The absolute frequency of occurrence did not yield a significant main effect, $F(1, 134) = 0.14$, $p = .709$, nor did it significantly interact with the additive components' role, $F(1, 134) = 0.71$, $p = .400$.

Discussion of Experiments 1 and 2

Experiments 1 and 2 demonstrated that the different patterns of the relative weight of additive shared and distinctive components, associated with modality, cohesiveness, and structural alignment, can be generated by varying the components' frequency. In Experiment 1, the absolute frequency of occurrence was manipulated, but the simultaneous occurrence of the same additive component was assumed to be random. In this case, the less frequent components were more highly weighted than the more frequent ones, particularly when they were shared by the two compared stimuli. In Experiment 2, when the percentage of joint occurrence was varied independently of the overall frequency, the relative weight of shared and distinctive components was primarily determined by the percentage of simultaneous occurrence rather than by the overall frequency. Because the overall frequency with which a component occurs is not highly informative under these conditions, it is conceivable that expectations focus on the role a

Table 2
Experiment 2: Mean Estimates of Component Weights

		Component frequency	
% shared	Component role	8/28	16/28
25	Shared	2.24	1.79
	Distinctive	4.00	2.27
75	Shared	1.61	0.78
	Distinctive	4.10	6.83

Note. Similarity was rated on a 20-point scale ranging from 1 (*not similar*) to 20 (*very similar*). The weights of the additive component (x) were computed by adding x to one or both members of a basic stimulus pair (p, q). The three similarity judgments—$s(p, q)$, $s(px, q)$, and $s(px, qx)$—were used in computing the component's weight as a shared ($C[x] = s[px, qx] - s[p, q]$) and as a distinctive ($D[x] = s[p, q] - s[px, q]$) feature.

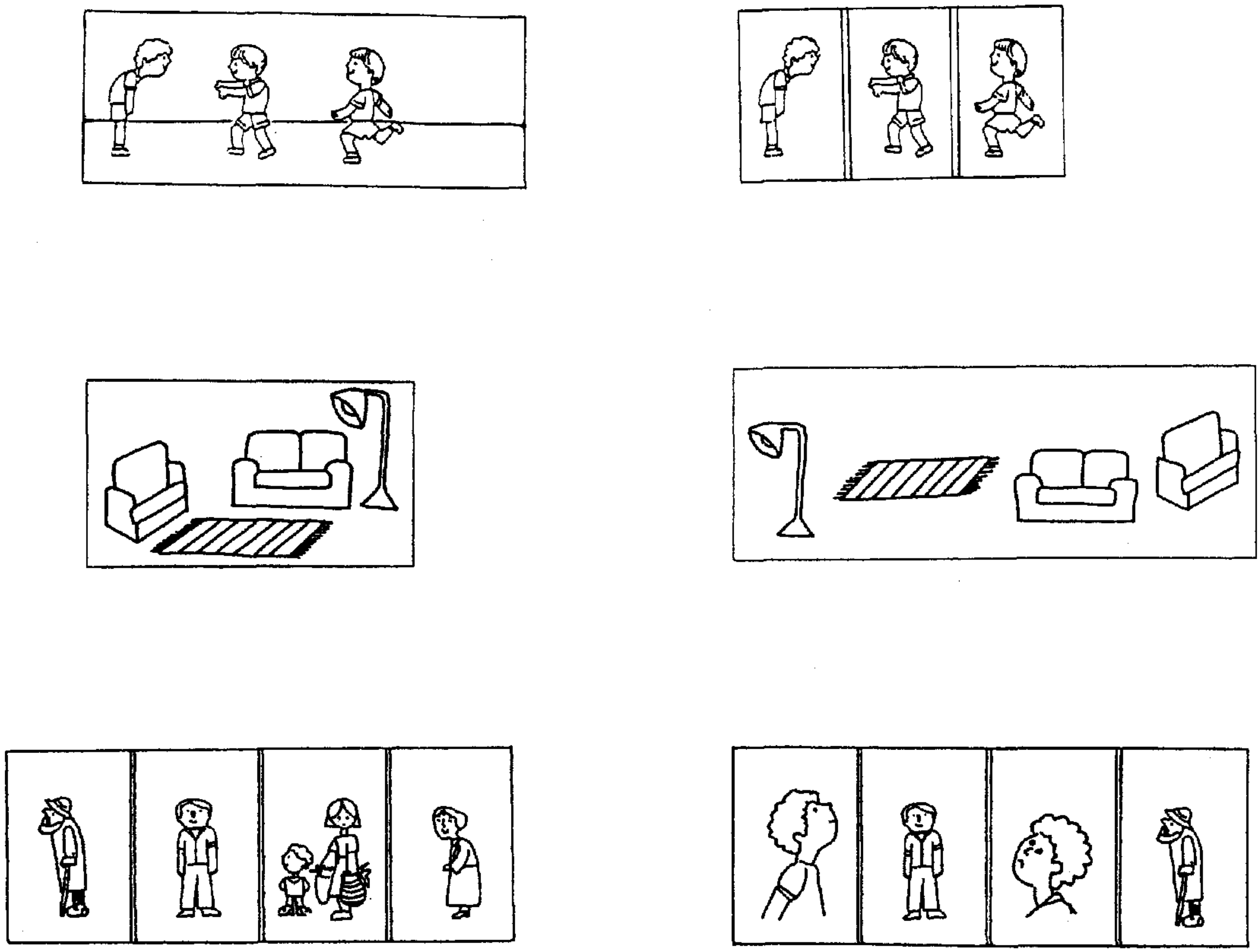

Figure 2. Cohesiveness-reducing transformations used in Ritov et al. (1990). From "Differential Weighting of Common and Distinctive Components," by I. Ritov, I. Gati, and A. Tversky, 1990, *Journal of Experimental Psychology: General, 119,* pp. 34, 36, 37. Copyright 1990 by the American Psychological Association. Reprinted with permission.

component might play, rather than simply on whether it appears at all.[2]

Cohesiveness

Ritov et al. (1990) examined the impact of cohesiveness on feature weighting by using various transformations that increased or decreased the cohesiveness of stimuli. Figure 2 displays examples of three cohesiveness-reducing transformations of visual stimuli. In the upper panels, the same collection of separable visual components is depicted either against a common background (high cohesiveness) or as isolated elements separated by vertical lines (low cohesiveness). The middle panels show a different transformation: scrambling of the natural order of components in commonly depicted scenes. Finally, in the lower panels, the manipulation involved composite displays of elements belonging either to a uniform set (high cohesiveness) or to a mixed set (low cohesiveness). In each of these experiments, the contribution of the same additive component was assessed under conditions of high and low cohesiveness. As predicted, the impact of adding the component to both stimuli, relative to its impact when added to only one of the compared stimuli, was higher in the low-cohesiveness conditions than in the high-cohesiveness conditions.

I suggested earlier that the change in relative weight associated with a decrease in cohesiveness results from a change in expectations concerning the relevant components. In these experiments,

[2] The sharing of an additive component could, alternatively, be regarded as a relation between the compared objects (both have a blob in them). From this perspective, the results are compatible with earlier findings that relations are weighted more highly than simple attributes in similarity judgment (Gentner, Rattermann, & Forbus, 1993; Medin et al., 1993). By this interpretation, the learning sequence, particularly in the high relative frequency of the shared-component condition, enhances expectations concerning the relationship between the two stimuli.

unlike Experiments 1 and 2 of the present research, expectations were based on inferred sets, rather than experienced frequencies. As detailed in the introduction, more cohesive stimuli typically involve a more complex and hierarchical system of interconnected predicates and relations. Such complex structures impose more strict constraints on the type and position of possible components. These constraints, in turn, affect the observer's expectations that a particular element will be included or excluded from the stimulus. To support the expectations account of cohesiveness, the impact of cohesiveness on expectations regarding particular components was tested explicitly. The very same components whose weight was estimated by Ritov et al. (1990) in similarity judgments were used here. In a paradigm similar to the one used in earlier studies of structure processing, the target component was eliminated from the display, and the participants were asked to guess the identity of the missing component. Self-rated confidence served as the primary indicator of the strength of expectations.

Experiment 3

This experiment used the same stimuli as those used in Ritov et al.'s (1990) Experiment 3. In this experiment, sets of stimuli were constructed using an initial group of objects, such as furniture in a living room, with some additive components, such as a lamp, a rug, and a houseplant. Each picture in a set included the initial group and some of the additive components. The objects included in each picture were displayed either in a natural ordering (high cohesiveness) or scrambled (low cohesiveness).

For the purpose of this experiment, an additive component was eliminated from each picture by placing a dark patch over it. The participants were asked to guess which of two given additive components (both were included in the original group) was covered by the patch. They also were asked to rate how confident they felt in their response. Figure 3 presents one of the stimuli under both conditions. Given the preceding discussion, I expected that

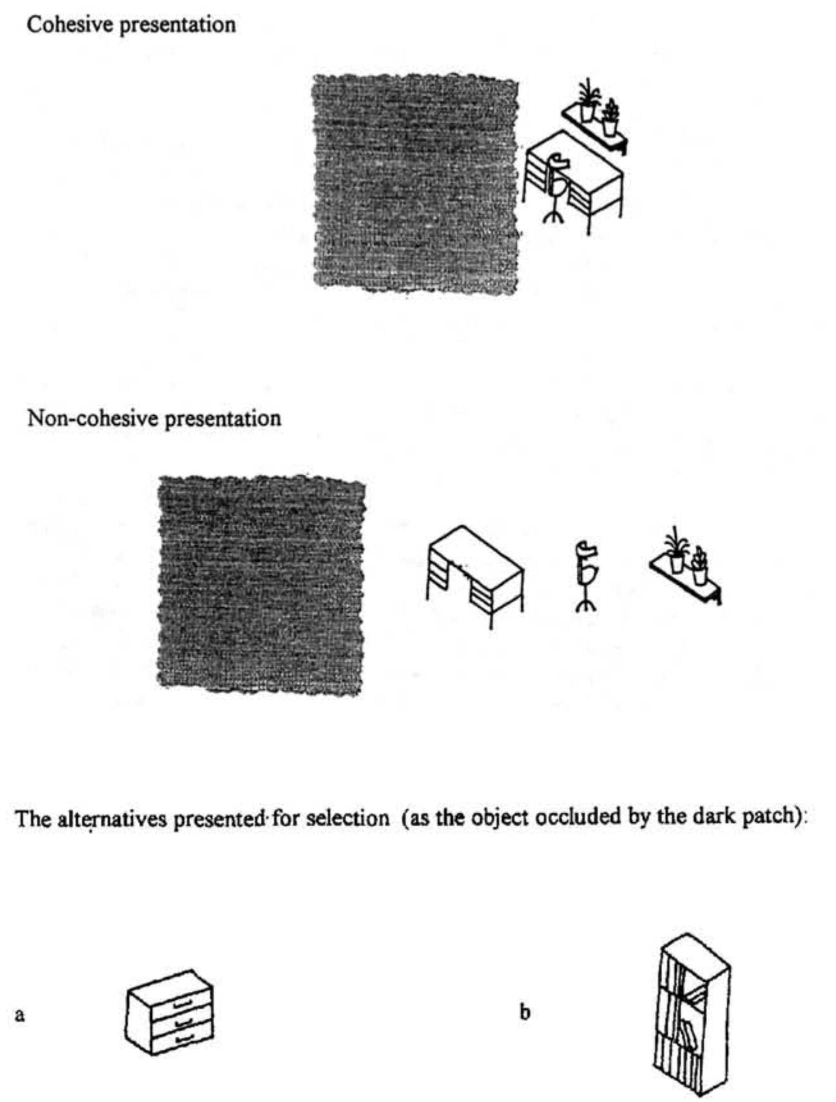

Figure 3. Experiment 3: An example of a stimulus in cohesive and noncohesive presentation.

the participants' confidence would be higher when the ordering of elements in the picture was natural than when the ordering was scrambled.

Method

Participants. Sixty-three students at Ben-Gurion University in Israel participated in this experiment. Questionnaires were run in a classroom setting.

Stimuli and design. The five sets used were flatware, office furniture, a living room, a street scene, and an airport. One picture was presented from each set, including the initial components and a dark patch covering the location of an additive component. The patches were square-shaped and were of identical size to avoid hinting at the shape of the missing component. The question following the picture read, "What do you think is hidden by the patch?" The participants were instructed to select their best guess from two additive components that were displayed directly below. Finally, the participants were asked to rate their confidence in their response by using a scale ranging from 1 (*low confidence*) to 10 (*high confidence*). In half of the questionnaires, components were presented in their natural order, and in the other half of the questionnaires, the order was scrambled. Participants were randomly assigned one of the questionnaires. Thus, cohesiveness was manipulated between participants. The five stimuli were presented in two different random orders.

Results

The percentage of correct responses in the high-cohesiveness presentation exceeded the percentage of correct responses in the low-cohesiveness presentation for four out of five sets. Mean confidence rating was higher in the high-cohesiveness condition than in the low-cohesiveness condition for all five sets. For a within-subject analysis, the mean percentage correct and the mean confidence rating were computed for each participant. Participants in the high-cohesiveness condition selected the correct response 67% of the time, whereas participants in the low-cohesiveness condition were correct only 57% of the time, $t(60) = 2.10, p < .05$. However, guessing right may not be the most informative indication of a cohesiveness effect. It is possible that, in some cases, after the order of the components had been scrambled, the other given component appeared more suitable than the original one. Confidence in one's choice provided a better measure of the degree to which cohesiveness uniquely determined expectations with respect to the missing component. Analysis of the confidence ratings supported the prediction that high cohesiveness would yield higher confidence (7.8 vs. 6.6), $t(61) = 2.27, p < .05$.

Experiment 4

In this experiment, the same task of selecting the missing component was used to test the effect of cohesiveness with the stimuli of Ritov et al.'s (1990) Experiment 4. In their experiment, the stimuli, all consisting of a series of pictures, were either uniform or mixed. The components of the uniform stimuli were consistent with respect to content and scale. The mixed stimuli were constructed by mixing the components of two uniform stimuli. The missing components were the additive components used by Ritov et al. to assess the relative weight of a feature in similarity judgment.

Method

Participants. Sixty-one students at Ben-Gurion University in Israel participated in this experiment. Questionnaires were run in a classroom setting.

Stimuli and design. Six sets were used in the uniform condition: a playground, a building, children, adults, caged animals, and dogs. The mixed stimuli were obtained by mixing the components from adjacent pairs of sets. One stimulus was presented from each set. The stimuli did not include the additive components that served as test features. The choice set for selecting the missing component included the two additive components of the adjacent sets. Thus, the same pair of components was presented as a choice set for two uniform stimuli and for the two mixed stimuli obtained from them. Uniformness was manipulated between participants. About half of the participants were presented with the uniform stimuli, and the remainder of the participants were presented with the mixed stimuli.

Results

The percentage of correct responses in the high-cohesiveness presentation exceeded the percentage of correct responses in the low-cohesiveness presentation for all six sets. Mean confidence rating was higher in the high-cohesiveness condition than in the low-cohesiveness condition for five out of six sets. For a within-subjects analysis, the mean percentage correct and the mean confidence rating were computed for each participant. Participants in the uniform (high-cohesiveness) condition selected the correct response 79% of the time, whereas participants in the mixed (low-cohesiveness) condition were correct only 61% of the time, $t(58) = 2.95, p < .01$. In an analysis of the confidence ratings, I found, as predicted, that uniform stimuli yielded higher confidence than mixed ones (8.3 vs. 6.8), $t(59) = 4.10, p < .01$.

Discussion of Experiments 3 and 4

Experiments 3 and 4 demonstrated that breaking the cohesiveness of stimuli decreased the probability of correctly identifying the missing components and lowered judged confidence. Thus, selecting the missing components of presented stimuli can be done with higher confidence if the stimuli are cohesive or with lower confidence if the stimuli are incohesive. To the extent that confidence reflects one's expectations, these findings imply that cohesive stimuli evoke more distinct expectations regarding their components. Hence, the results of Experiments 3 and 4 support the proposal that the documented change in the relative weight of the components, associated with a decrease in cohesiveness (Ritov et al., 1990), is due to modified expectations concerning these components.

Similarity Versus Difference

As noted earlier, the nature of the comparison task affects the relative weight of features by directing attention to either shared or distinctive components. Thus, shared features loom larger and consequently are more heavily weighted in similarity judgment than in dissimilarity judgment. As a result, similarity and dissimilarity may not always yield complementary results. Indeed, when participants were presented with two pairs of countries, one more prominent than the other, the pair of more prominent countries was selected both as more similar to each other and as more different from each other than the other pair (Tversky, 1977). This apparent inconsistency reflects an interaction between the task and the nature of the stimuli. The more prominent pair of countries calls to mind more attributes—both more shared and more distinctive ones. Because similarity judgment involves greater weighting of shared features, the more prominent pair is judged as being more similar than the other pair because of the larger set of shared attributes. Dissimilarity, in contrast, involves higher weighting of distinctive features; hence, the more prominent pair is judged to be more dissimilar than the other pair because of the larger set of distinctive attributes.

A similar mechanism has been shown to generate inconsistencies in choice. Shafir (1993) asked participants either to choose or to reject an option from a pair of given options, one having both more positive and more negative features than the other. The enriched option (having more features) tended to be both chosen and rejected relatively more often than the impoverished option. Presumably, the inconsistency results from the increased weight of features that are compatible with the task: Choice increases the weight of positive features, whereas rejection focuses attention on the negative features.

The impact of expectations on the weighting of features may interact with the nature of the task in a similar way. A shared unexpected component is likely to make a great contribution if the comparison task assigns increased weight to shared features. If, in contrast, the task focuses attention on distinctive features, the aforementioned commonality would have only a minor impact. Similarly, when an ordinary feature, consistent with expectations, is included in only one of the compared objects, its impact will be higher if the task calls for increased weighting of distinctive features.

Consider, for example, a task in which a target stimulus, T, is presented along with two alternatives, A and B. T shares an unusual (unexpected) feature with Alternative A and an ordinary (expected) feature with Alternative B. Figure 4 provides illustrations of such triads. Now imagine that participants are asked to judge which of the alternatives (A or B) either is more similar to the target or is more different from the target.

In assessing the similarity between each alternative and the target, shared features carry extra weight. This effect, combined with the relatively higher weight of shared unexpected features, will make the sharing of an unexpected feature particularly salient. Hence, Alternative A is likely to be perceived as more similar to T than B. In assessing dissimilarity, in contrast, distinctive features are more heavily weighted. Combining this effect with the relatively higher weight of distinctive expected features will make the exclusion of an expected feature from Alternative A particularly salient. As a result, Alternative A is likely to appear as more different from T than Alternative B. In sum, for triads such as the ones displayed in Figure 4, similarity and dissimilarity will not be complementary: Alternative A will tend to be selected both as more similar to T and as more different from it. This hypothesis was tested in Experiments 5 and 6.

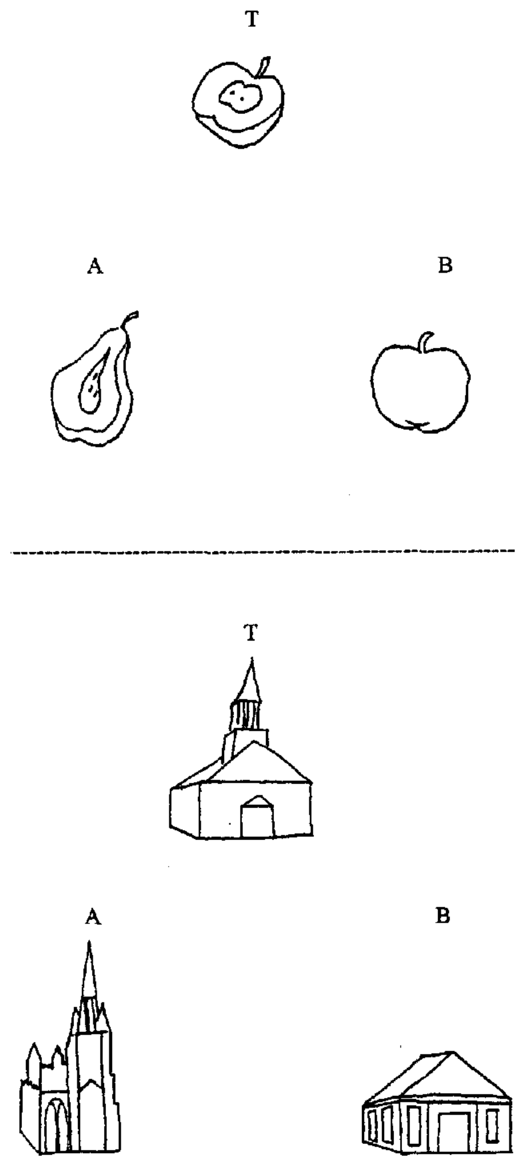

Figure 4. Experiment 5: Examples of pictorial triads.

Experiment 5

Method

Participants. Eighty-eight students participated in this experiment. They were run individually and were paid for participating.

Stimuli and design. Nine pictorial triads were used, including the ones presented in Figure 4. Each triad contained a target stimulus, T, and two alternative stimuli, A and B. T shared a rare feature with A and an ordinary feature with B. Each triad appeared on a separate page, with T at the top and A and B side by side below T. The left–right order of A and B was randomized between triads.

Four questionnaire versions were used, differing only in the instructions given to the participants. In one version, participants were instructed to decide, for each triad of pictures, whether the picture on the top was more similar to the picture on the left or to the picture on the right. In a second version, the words *more similar to* were replaced with *more different from.* A third and a fourth version of the questionnaire corresponded to the first and second versions, with one change: Instead of being asked about the similarity (difference) of the test picture to the other pictures (A and B), participants were asked about the similarity (difference) of A and B to the test picture. Participants were randomly assigned to one of the questionnaire versions.

Table 3
Experiment 5 (Pictorial Stimuli): Mean Percentages of Selecting the Shared Rare-Component Alternative (Alternative A)

	Comparison direction	
Task	Target to alternatives	Alternatives to target
Similarity	59	56
Difference	55	63

Results

For each participant, the percentage of triplets (*P*) in which he or she selected A over B was computed. The mean percentage of A selection across all participants in each condition is presented in Table 3. As predicted, A was selected overall 58% of the time, significantly more than 50%, $t(87) = 4.30$, $p < .01$. Analysis of variance of the percentage of A selection by the two manipulated variables (similarity vs. difference and comparison of target to alternatives vs. alternatives to target) yielded no significant effects.

Experiment 6

The design of this experiment replicated the design of Experiment 5 but with verbal stimuli.

Method

Participants. Eighty-six students participated in this experiment. They were run individually and were paid for participating.

Stimuli and design. All stimuli were descriptions of college students, each in terms of two features. Eight triads of descriptions were used. The design was identical to that used in Experiment 5. An example of a triad of descriptions is presented in Figure 5.

Results

For each participant, the percentage of triplets (*P*) in which he or she chose A as more similar to (or different from) T than B was computed. Overall, A was selected 53% of the time, significantly more than 50%, $t(87) = 1.85$, $p < .05$, one-tailed. The small size of the effect can be explained by examining the mean percentage of A selection in the different conditions. These are presented in Table 4. As one can clearly see, A was selected as more different from T more frequently than it was selected as more similar to T. This fact indicates that, at least for some triplets, one of the features was clearly weighted more heavily than the other. Thus, the stimulus that shared this feature with T was judged as more similar to T, and the stimulus that lacked the feature was judged as more different from T. In cases like this, in which one feature is considerably more salient than the other features, it is indeed predicted that similarity and difference judgment will complement each other.

The difference between the percentage of participants who selected A as more different from T and the percentage of participants who selected A as more similar to T can be taken as a measure of the imbalance between the features. Excluding from the analysis three triplets for which the aforementioned difference exceeded 50%, I found that, overall, A selections were made on 55% of the choices.

T

A geophysics major
who is a member of
the local public
television station

A

A geophysics major
who is a member of
the boating club

B

An economics major
who is a member of
the local public
television station

Figure 5. Experiment 6: A triad of descriptions.

Discussion of Experiments 5 and 6

Inconsistencies between similarity and difference judgments may result from interaction of the task with the nature of the attributes. Previous research (Markman, 1996; Medin, Goldstone, & Gentner, 1990) uncovered special kinds of inconsistencies, resulting from higher weighting of relational matches, relative to attributional ones, in similarity judgment. Experiments 5 and 6 of the present research provide evidence of a new class of comparisons for which judgments of similarity and dissimilarity are not necessarily complementary. The lack of complementarity in these experiments stemmed from the impact of expectations on the relative weight of shared and distinctive components. As predicted, the picture or verbal description that shared an unusual feature with the test stimulus was perceived both as more similar to it and as more different from it than the other alternative.

General Discussion

In this article, I have proposed that expectations play a major role in determining the relative weight of shared and distinctive components. This general principle accounts for earlier findings in the domain of comparative judgment involving the effects of modality, cohesiveness, and structural alignment. To support this account, expectations were manipulated by changing the frequency with which a component occurred, without changing cohesiveness

Table 4
Experiment 6 (Verbal Stimuli): Mean Percentages of Selecting the Shared Rare-Component Alternative (Alternative A)

	Comparison direction	
Task	Target to alternatives	Alternatives to target
Similarity	48	43
Difference	61	61

or alignment. Results of Experiments 1 and 2 showed the predicted pattern of changes in the components' weights, whereby sharing of an unexpected (infrequent) component was highly weighted when the components of the compared objects were reasonably assumed to be independent. When independence did not hold, higher relative frequency of shared components led to higher relative weight of distinctive (nonshared) components. Next, increased confidence in expectations in cohesive stimuli relative to noncohesive stimuli (Experiments 3 and 4) provided further support for the notion that the impact of cohesiveness stemmed from changes in expectations.

The judged frequencies of independent as well as joint occurrence may depend on the context in which the comparison takes place. As in Experiments 1 and 2, participants have been locally exposed to frequencies different from universally relevant ones, or to frequencies different from the frequencies characterizing other local contexts. In those cases, the effect of frequency on the relative weight of shared and distinctive components can be regarded as a context effect. Some other context effects, particularly those involving perceived range of variation, are also compatible with the expectations account delineated here. Some examples of such context effects are briefly discussed below.

Mellers and Cooke (1994) showed that in judgments of multiattribute objects, trade-offs between attributes depend on the attributes' range. In one of their studies, for example, participants judged the attractiveness of apartments varying in monthly rent and distance to campus. When the range of rent within an experimental set was narrow, the attractiveness of low rent had a greater impact. People tended to favor cheaper apartments over closer apartments. When the range of rent was wide, the preference for closer apartments increased. Thus, a fixed difference in rent had a greater effect when the range of rent was narrow. This effect produced rank-order shifts when the same pair of apartments was embedded in different context sets. I propose that the impact of expectations can also account for range effects of the kind described above (Mellers & Cooke, 1994). In those cases, the local context defines the range of possible values and evokes expectations with respect to differences along an attribute. To the extent that a narrow range engenders expectations for small differences, the actual difference between two apartments is more likely to exceed expectations in the narrow-range condition than in the wide-range condition.

Local context is not, however, the only determinant of range expectations. The stimuli included in a particular experimental context may invoke a range of possibilities beyond those actually presented. The impact of this invoked range could exceed that of the actually encountered one. For example, earlier studies involving paired-associates learning (Garner, 1974) suggested that the size of the stimulus-inferred subset is at least as important to improved memory as the size of the actual subset used. Thus, memory for one dot pattern would be superior to memory for a second dot pattern, although each belongs to an actually presented subset of two patterns, because the former's inferred subset is smaller than the latter's. More generally, to the extent that judged stimuli spontaneously invoke an expected range beyond the one characterizing the encountered set, experimentally defined range is less prone to play a major role. Hence, attribute range effects of the kind documented by Mellers and Cooke (1994) are more likely to occur with stimuli for which the invoked expectations concerning attribute range are largely congruent with experimentally defined range.

Beattie and Baron (1991) presented a series of studies in which no range effects were found. For example, the ranges of midterm and final-exam grades did not affect the overall ratings assigned to students on the basis of these grades. In this case, it is likely that participants' expectations were based on their preconceived ideas with respect to the grades' ranges, rather than on the actual ranges presented in the experimental context. Consequently, the actual range of the experimental set did not play a role in this judgment. Beattie and Baron found range effects in one experiment (Experiment 6) that required participants to judge hypothetical job candidates on the basis of stimulus scales with which the participants had no previous experience. In this case, clearly the expectations were wholly determined by the presented distribution.

A related context effect, the extension effect (Tversky, 1977), occurs when the set embedding the objects of comparison is enlarged, so that an attribute that had formerly been perceived as single-valued becomes variable. As a result, the weight of this attribute in the comparison increases, often substantially changing the judged similarity between the original object pair. A possible account for this effect involves changes in expectations regarding the objects' features. When a feature is common to all objects in the relevant set, the fact that it is shared by the compared stimuli is undoubtedly fully consistent with expectations. Its being shared by the compared objects can be surprising, to any extent, only if one can conceivably imagine a different possibility. Medin et al.'s (1993) finding that some properties are not considered unless there is variation in the property is compatible with this assertion. If the weight assigned to a feature is inversely related to its surprisingness, as I have proposed here, then enlarging the set should have the documented effect. The diagnosticity principle proposed by Tversky, attributing the extension effect to a change in the diagnostic value of the feature, can be viewed as a special case of the expectations account proposed here.

Goldstone, Medin, and Halberstadt (1997) extended earlier findings by exploring context effects that occur within a single comparison rather than across several trials. They showed that the alternatives presented during a particular comparison influenced which dimensions were foregrounded and, hence, more heavily weighted. Their results are consistent with a strategy of placing emphasis on dimensions that show a large amount of variability within a stimulus set on dimension values that are rare. Furthermore, Goldstone et al. demonstrated that these context effects can occur even when other alternatives are not present to provide a context for the judgment. Building on the research of Garner (1970) and Kahneman and Miller (1986), Goldstone et al. accounted for these findings by assuming that isolated comparisons are made by recruiting standards of comparison that define the dimensions and alternatives relevant to the particular comparison.

In the domain of semantic processing, research on recall of learned word pairs provides an interesting example of a context effect related to expectations. Although the semantic relationship among to-be-remembered items typically facilitates memory for those items, Hirshman (1988) reported a series of memory experiments showing superior recall of weakly related word pairs that were learned in the same study list as strongly related word pairs. The advantage of the weakly related pairs disappeared when the list of learned word pairs wholly consisted of either weakly related or strongly related pairs. Furthermore, the effect depended on the relative frequency of weakly and strongly related pairs in the study list: Weakly related words were recalled better than strongly related words only if the list was mostly made up of strongly related words. Hirshman proposed an account based on the assumption that the weakly related word pairs, when appearing in the context of strongly related pairs, violated the participants' expectations. As a result, the participants engaged in "blind-alley" searches, which, in turn, served as cues for recall. If the study list largely or entirely consists of weakly related pairs, the lack of a strong relationship is hardly surprising. No expectations are violated, and hence the weakly related pairs lose their relative advantage.

A similar effect has been extensively studied by social psychologists. The concept of "expectancy," referring to beliefs about a future state of affairs, forms a major explanatory construct in this line of research (Olson, Roese, & Zanna, 1996). In particular, studies show that, under some broad conditions, information that violates prior expectations ("expectancy-incongruent") is processed and recalled better than information that is consistent with expectations (Higgins & Bargh, 1987; Stangor & McMillan, 1992).

Finally, a number of known preference reversals in the decision-making literature have recently been attributed to differences between independent and comparative elicitation procedures (Bazerman, Moore, Tenbrunsel, & Wade-Benzoni, 1999; Hsee, 1996; Kahneman & Ritov, 1994; Nowlis & Simonson, 1997; Ritov & Kahneman, 1997). In many of the observed reversals, it can be argued that when each object is judged separately, the evoked comparison set is different for various reasons from the one evoked by simultaneous presentation of both objects. For example, in the domain of public problems, when considering the problem of the threatened dolphin population, one spontaneously thinks of threats to other animal species, whereas being informed of a public health problem, such as the incidence of skin cancer among farmworkers, naturally brings to mind other public health problems. Thus, all of the alternatives evoked by the dolphin problem share the attribute "animal species problem," whereas the alternatives evoked by the skin-cancer problem share the attribute "human health problem."

If the evoked alternatives share an attribute with the judged object, the attribute's value can be viewed as consistent with expectations. As demonstrated by the extension effect, when a feature is shared by all relevant objects, its saliency decreases, and hence the attribute's weight in independent judgments is diminished. The variable "human versus animal problem" will, consequently, have little impact on separate judgments of the public problems described above. In comparing the two problems, however, the human versus animal problem attribute is no longer single-valued, and hence its weight increases. Furthermore, the difference between the two values is alignable and thus is heavily weighted in the comparison process. As a result, although public support for saving dolphins was rated higher than support for addressing the farmworkers' skin-cancer problem when the two problems were judged separately, the order was reversed when the two problems were considered simultaneously (Ritov & Kahneman, 1997).

Another cluster of preference reversals in which isolated stimuli evoke different sets of alternatives involves objects that are clearly defective. For example, Bazerman, Schroth, Shah, Diekman, and Tenbrunsel (1994) presented master of business administration students with two hypothetical job offers: one job with a higher but inequitable salary and another job with an equitable but somewhat lower salary. Because defects are generally mutated in the direction of the norm (Kahneman & Miller, 1986), the first offer is likely to evoke alternatives with equitable salary. Consequently, the inequitable pay is inconsistent with expectations and is heavily weighted. The inferred set is less likely to include variation in absolute pay because this attribute may be harder to evaluate on its own (Hsee, 1996). The amount of salary is, thus, consistent with expectations; its weight diminished as a result. When the two job offers are compared, however, the range of absolute salary becomes apparent, and the weight of the attribute consequently increases.

Conclusion

The role of expectations, delineated above, provides a unifying framework for different findings related to comparative judgment, including the effects of modality, cohesiveness, structural alignment, and some forms of preference reversals. Expectations, often in the form of inferred sets of alternatives, pertain either to the features of each separate object or to their simultaneous occurrence in all compared objects. The augmented weighting of unexpected features explains why distinctive features are relatively more highly weighted in cohesive stimuli; why aligned differences are more highly weighted than nonaligned features; why trade-offs in preference depend on attribute range; and why, under some conditions, judgments of similarity and dissimilarity are asymmetric. Finally, diverse effects in domains ranging from learning (improved memory for semantically unrelated word pairs) to choice (preference reversal between isolated and comparative evaluations of public goods) may, likewise, result from the greater impact of inconsistent than consistent-with-expectations attributes.

References

Bazerman, M. H., Moore, D., Tenbrunsel, A. E., & Wade-Benzoni, K. A. (1999). Explaining how preferences change across joint versus separate evaluations. *Journal of Economic Behavior and Organizations, 39,* 41–58.

Bazerman, M. H., Schroth, H., Shah, P. P., Diekman, K., & Tenbrunsel, A. E. (1994). The inconsistent role of comparison others and procedural justice in reactions to hypothetical job descriptions: Implications for job acceptance decisions. *Organizational Behavior and Human Decision Processes, 60,* 326–352.

Bear, G. (1973). Figural goodness and the predictability of figural elements. *Perception & Psychophysics, 13,* 32–40.

Beattie, J., & Baron, J. (1991). Investigating the effect of stimulus range on attribute weight. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 571–585.

Bowdle, B. F., & Gentner, D. (1997). Informativity and asymmetry in comparisons. *Cognitive Psychology, 34,* 244–286.

Garner, W. R. (1970). Good patterns have few alternatives. *American Scientist, 58,* 34–42.

Garner, W. R. (1974). *The processing of information and structure.* Potomac, MD: Erlbaum.

Gati, I., & Ashkenazi, T. (1996). *Factors affecting the relative weight of common to distinctive features in similarity and preference judgments.*

Unpublished manuscript, Hebrew University of Jerusalem, Jerusalem, Israel.

Gati, I., & Tversky, A. (1984). Weighting common and distinctive features in perceptual and conceptual judgments. *Cognitive Psychology, 16,* 341–370.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7,* 155–170.

Gentner, D., & Markman, B. A. (1994). Structural alignment in comparison: No difference without similarity. *Psychological Science, 5,* 152–158.

Gentner, D., & Markman, A. B. (1995). Similarity is like analogy: Structural alignment in comparison. In C. Cacciari (Ed.), *Similarity in language, thought and perception* (pp. 111–147). Brussels, Belgium: BREPOLS.

Gentner, D., Rattermann, M. J., & Forbus, K. D. (1993). The roles of similarity in transfer: Separating retrievability from inferential soundness. *Cognitive Psychology, 25,* 524–575.

Goldstone, R. L., Medin, D. L., & Halberstadt, J. (1997). Similarity in context. *Memory & Cognition, 25,* 237–255.

Higgins, E. T., & Bargh, J. A. (1987). Social cognition and social perception. *Annual Review of Psychology, 38,* 369–425.

Hirshman, E. (1988). The expectation–violation effect: Paradoxical effects of semantic relatedness. *Journal of Memory and Language, 27,* 40–58.

Houston, D. A., Sherman, S. J., & Baker, S. M. (1991). Feature matching, unique features, and dynamics of the choice process: Predecision conflict and postdecision satisfaction. *Journal of Experimental Social Psychology, 27,* 411–430.

Hsee, C. K. (1996). The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes, 67,* 247–257.

Kahneman, D. (1995). Varieties of counterfactual thinking. In N. J. Roese & J. M. Olson (Eds.), *What might have been: The social psychology of counterfactual thinking* (pp. 375–396). Mahwah, NJ: Erlbaum.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93,* 136–153.

Kahneman, D., & Ritov, I. (1994). Determinants of stated willingness to pay for public goods: A study in the headline method. *Journal of Risk and Uncertainty, 9,* 5–38.

Kruschke, J. K. (1996). Base rates in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 3–26.

Loewenstein, G., & Mather, J. (1990). Dynamic processes in risk perception. *Journal of Risk and Uncertainty, 3,* 155–175.

Markman, A. B. (1996). Structural alignment in similarity and difference judgments. *Psychonomic Bulletin & Review, 3,* 227–230.

Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology, 25,* 431–467.

Markman, A. B., & Medin, D. L. (1995). Similarity and alignment in choice. *Organizational Behavior and Human Decision Processes, 63,* 117–130.

Medin, D. L., & Edelson, S. M. (1988). Problem structure and the use of base-rate information from experience. *Journal of Experimental Psychology: General, 117,* 68–85.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1990). Similarity involving attributes and relations: Judgments of similarity and difference are not inverses. *Psychological Science, 1,* 64–69.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100,* 254–278.

Medin, D. L., Goldstone, R. L., & Markman, A. B. (1995). Comparison and choice: Relations between similarity processes and decision processes. *Psychonomic Bulletin & Review, 2,* 1–19.

Mellers, B. A., & Cooke, A. D. J. (1994). Trade-offs depend on attribute range. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 1055–1067.

Mellers, B. A., Schwartz, A., Ho, K., & Ritov, I. (1997). Elation and disappointment: Emotional responses to risky options. *Psychological Science, 8,* 423–429.

Nowlis, S. M., & Simonson, I. (1997). Attribute–task compatibility as a determinant of consumer preference reversals. *Journal of Marketing Research, 34,* 205–218.

Olson, J. M., Roese, N. J., & Zanna, M. P. (1996). Expectancies. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 211–238). New York: Guilford Press.

Ritov, I., Gati, I., & Tversky, A. (1990). Differential weighting of common and distinctive components. *Journal of Experimental Psychology: General, 119,* 30–41.

Ritov, I., & Kahneman, D. (1997). How people value the environment: Attitudes versus economic values. In M. Bazerman, D. Messick, A. Tenbrunsel, & K. Wade-Benzoni (Eds.), *Psychological perspectives to environmental and ethical issues* (pp. 33–51). San Francisco: New Lexington Press.

Shafir, E. (1993). Choosing versus rejecting: Why some options are both better and worse than others. *Memory & Cognition, 21,* 546–556.

Shank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding.* Hillsdale, NJ: Erlbaum.

Simonson, I., & Tversky, A. (1992). Choice in context: Tradeoff contrast and extremeness aversion. *Journal of Marketing Research, 29,* 281–295.

Stangor, C., & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social developmental literatures. *Psychological Bulletin, 111,* 42–61.

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

Tversky, A., & Gati, I. (1978). Studies of similarity. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 79–98). Hillsdale, NJ: Erlbaum.

Received December 26, 1997
Revision received July 12, 1999
Accepted July 12, 1999 ■