Similarity and Difference Judgments Under Perceptual and Non-Perceptual Conditions

Uri Hasson (uhasson@princeton.edu)
Department of Psychology, Princeton University
Princeton, NJ 08544, USA

Vladimir Sloutsky (Sloutsky.1@osu.edu)
Center for Cognitive Science & School of Teaching and Learning, The Ohio State University
21 Page Hall, 1810 College Road, Columbus, OH 43210, USA

Abstract

It has recently been suggested that knowledge is represented in the form of perceptual symbol systems (Barsalou, 1999). According to this view, perceptual states may be used to support higher cognitive processes without being transduced into a representational language. Since the ability to recognize difference and similarity is fundamental for cognition, we examined to what extent it might be based on perceptual information. In three experiments, participants made judgments of similarity and difference for simple items under three presentation conditions: Words Only, Words and Pictures, and Pictures Only. Reaction times for judgments in the Words-Only condition were consistently slower than in the other presentation conditions. However, judgments of perceived similarity and perceived difference did not markedly differ between presentation conditions. The results suggest that participants recruited perceptual information when evaluating similarity in the Word-Only condition. An analysis of those items for which similarity was found, demonstrated that the presentation condition had an effect on the relation between the similarity and difference scales: for a given degree of similarity, more extreme difference judgments were found under those conditions where words were displayed. We offer an explanation for this effect, and present a further research program.

Introduction

Similarity, or psychological resemblance of entities, is a fundamental aspect of cognition. Similarity plays a critical role in perception, memory, learning and transfer, categorization, analogical reasoning, problem solving, and language comprehension. It has also been suggested that recognizing differences between entities is fundamental to cognition; for instance, in distinguishing between basic-level categories. An extensive body of research then has outlined the processes for which similarity and difference are important. However, much less is known about what information is used for evaluating similarity and difference themselves.

Judgments of similarity and difference are taken to be processes in which knowledge plays an important role. According to some positions, conceptual knowledge plays such an important role in similarity, that similarity is taken to be akin to analogy; operating on an interconnected system of relations and their arguments (Gentner & Markman, 1997). The crux of this view is that processes of similarity, as well as difference (Markman, 1996), operate by aligning the structures of the entities which are being compared. The structures themselves are described as a system of relational predicates and their attributes. The research program of structural alignment (see Gentner & Markman, 1997) convincingly demonstrated that the use of structural knowledge is an integral part of making similarity judgments. Other theories of similarity hypothesize a process that takes as input representations in the form of feature lists (Nosofsky, 1986; Tversky, 1977). It is an open question, however, whether the *only* kinds of information used for these judgments are in the form of such a-modal representations as relations or feature lists.

A recent proposal (Barsalou, 1999), suggests that perceptual states - modal and analog forms of representation - may *also* be used to support higher cognitive processes. Perceptual states are taken to maintain a part of the perceptual nature of their referents, without being transduced into a representational language. If this is the case, judgments of similarity and/or difference might also make use of perceptual information, in addition to other forms of knowledge.

The purpose of the present study is to examine whether perceptual information is recruited when making judgments of similarity and difference. We examined this issue by evaluating the effects of various types of presentation conditions on similarity and difference judgments, which were made for a variety of natural stimuli. We presented participants with pairs of items for similarity and difference judgments under three presentation conditions which were manipulated between groups: as Words-Only (WO), as words

accompanied by pictures (WP) or as pictures only (PO). The instructions given to all three groups were the same, and did not make any reference to speed of response. Most critically, the instructions given to participants in the Word-Only condition did *not* mention envisioning the objects depicted by the words. If participants in this group took longer to reach their decision, but ultimately made judgments resembling those made in the picture conditions, this would suggest that perceptual information was recruited when making these judgments.

The reported research has two major goals: (1) to estimate effects of perceptual and non-perceptual aspects of the stimuli on making judgments of similarity and difference; and (2) to examine whether perceptual and non-perceptual aspects of the stimuli affect the relation between the similarity and difference scales.

Calibration Experiment

The purpose of the calibration study was to choose pictorial material that would be easily and unambiguously recognized as typical examples of the intended items.

Method

Participants Participants were 27 undergraduate students from the Ohio State University that participated in the study for course credit.

Materials Materials consisted of 100 items, taken from 10 categories of Battig and Montague's (1969) category norms with 10 items selected from each category. The ten categories belonged to two ontological domains, living things and artifacts, with five categories in each domain. In addition to the experimental items, 25 items that appeared to be bad examples of their types were also added as negative anchors. The pictures were chosen by the experimenters from the Corel Gallery software.

Procedure Participants were presented with pictures of objects, followed by words denoting these objects. For example, a picture of a car was displayed, followed by a blank screen and then by the word CAR. Participants were told that their task was to decide whether the photograph was a good example of the word that followed it. A good example was defined as "something that immediately reminds you strongly of that thing". Participants were told to press 1 if the photo was indeed a good example and 0 if it is not a good example.

Results

The proportions rating for the experimental items was 0.85. We analyzed the data by item to see whether certain items were particularly bad examples. We set a cutoff point at 67%, which meant that two thirds of the subjects agreed that the picture was a good example of the item. Only 10 of the 100 items fell below this criterion, and were not used in the experiment itself.

Experiment 1a: Judging Similarity and Difference of Objects Denoted by Words

Method

Participants There were 29 participants in the similarity-judgment condition and 25 participants in the difference-judgment condition. The participants in both conditions, and in all following experiments, were undergraduate students from the Ohio State University who participated to fulfill a psychology course requirement or in return for payment.

Design and Materials The experiment had a mixed design with Judgment type (similarity judgment or difference judgment) as a between-subjects variable and Pair-type (same-superordinate, within-ontological domain or across-ontological domains) as a within-subjects variable. The 90 items chosen from the calibration test were used. Items were paired to construct 240 pairs of items such that: (a) 80 pairs were of items from the same superordinate category, (b) 80 pairs were of items from different superordinate categories but from the same ontological domain (i.e., both were either artifacts or living things) and (c) 80 pairs were of items belonging to different ontological domains. We decided to use such pairs since there was ground to suppose that they considerably differ in their perceived similarity (see Markman & Wisniewski, 1997).

Procedure The participants worked alone in an experimental room. Participants making similarity judgments were told that their task was to decide how similar these items were. They were instructed to press 3 if items were very similar, 2 if they were not so similar, and 1 if they were not similar at all. Participants making difference judgments were told that their task was to decide how different these items were. They were instructed to press 3 if items were very different, 2 if they were not so different, and 1 if they were not different at all. The experiment was run in three blocks, with a 1-minute break between blocks. The experiment began with a few practice pairs, which were followed by 14 hidden practice pairs whose purpose was to bring participants up to speed. Practice items were not analyzed.

Results and Discussion

Reaction time data and ratings for similarity and difference judgments are presented in Table 1. Note that ANOVA results will be presented in the overall-analysis section, whereas the results of t-tests are presented in the Table 1.

Table 1. Reaction times and ratings for similarity and difference judgments in the Word-Only condition

	Reaction Time		Ratings	
Pair type	Sim	Dif	Sim [x]	Dif [y]
Across-domain	1232_a	1245_a	1.04_a	2.99_a
Within-domain	1315_b	1317_b	1.20_b	2.90_b
Within-superordinate	1400_c	1511_c	2.32_c	2.06_c
Average	1315	1358	1.52	2.65

Note. Numbers with different subscripts in a given column differ at $p < .0001$.
[x] High numbers reflect greater similarity
[y] High numbers reflect greater difference

As expected, similarity and difference ratings for the different pair types were different, with Within-Superordinate pairs (e.g., dog – cat) being more similar than Within-Domain (e.g., shark – horse) pairs, and the latter being more similar than Across-Domain (e.g., cow – spoon) pairs. The difference ratings showed the same pattern. More importantly, reaction times in both the similarity and difference judgment groups were fastest for Across-Domain pairs, slower for Within-Domain pairs and slowest for Within-Superordinate pairs. As will be shown, this pattern of results replicated in the Word-Picture and Picture-Only presentation conditions as well.

Experiment 1b: Judging Similarity and Difference of Objects Depicted by Words and Pictures

Method

Participants There were 18 participants in the similarity-judgment condition and 18 participants in the difference-judgment condition.

Design, Materials and Procedure Design, Materials, and Procedure were identical to those in Experiment 1a, except that items were denoted by words and by pictures, such that the words were printed above the pictures.

Results and Discussion

Reaction time data and ratings for similarity and difference judgments are given in Table 2. The overall pattern of reaction time and rating results was similar to that in experiment 1a.

Table 2. Reaction times and ratings for similarity and difference judgments in the Picture-Word condition

	Reaction Time		Ratings	
Pair type	Sim	Dif	Sim [x]	Dif [y]
Across-domain	1002_a	1180_a	1.04_a	2.96_a
Within-domain	1123_b	1177_b	1.23_b	2.84_b
Within-superordinate	1270_c	1376_c	2.37_c	1.82_c
Average	1131	1211	1.55	2.54

Note. Numbers with different subscripts in a given column differ at $p < .0001$.
[x] High numbers reflect greater similarity
[y] High numbers reflect greater difference

Initial inspection demonstrates that reaction times in this presentation mode were faster than in the Word-Only presentation condition. Though several theoretical explanations for this may exist, it was important to examine one in particular; namely, that participants were ignoring the words and basing their decisions on other criteria, which did not involve processing of the words.

To examine whether participants were reading the words, we modified the experiment slightly to include 12 inconspicuous spelling mistakes. If participants were paying attention to the words, reaction times for these changed items should be higher than reaction times to the same items in the original experiment. Thirteen participants participated in the modified version of the similarity-judgment condition. The average reaction time for the modified items (1453 ms) was significantly higher than for the original, unmodified, items (1170 ms), one-tailed t-test, $t_{22} = 2.37$, $p < .01$. In addition, we asked participants whether they had noticed something during the study. All but two noticed a few spelling mistakes. The results indicate that participants were indeed reading the words presented above the pictures, and that the reduced reaction times could not be attributed to such neglect.

Experiment 1c: Judging Similarity and Difference of Objects Depicted by Pictures

Method

Participants There were 25 participants in the similarity-judgment condition and 26 participants in the difference-judgment condition.

Design, Materials and Procedure Design, Materials, and Procedure were identical to those in Experiment 1a and 1b, except that items were now depicted only by pictures. Participants were told that they would be presented with pictures referring to objects in the world, and that their task is to determine how similar these objects are.

Results and Discussion

Reaction time data and ratings for similarity and difference judgments are given in Table 3. The overall pattern of results was as in experiments 1a and 1b. The fact that the similarity ratings increased and difference ratings decreased the more conceptually related the objects were, testifies to the fact that participants used conceptual knowledge in their judgments of pictorial material, and did not rely exclusively on perceptual information. Reaction time data and ratings of similarity and difference are given in Table 3.

Table 3. Reaction times and ratings for similarity and difference judgments in the Picture-Only condition

	Reaction Time		Ratings	
Pair type	Sim	Dif	Sim [x]	Dif [y]
Across-domain	1067_a	1159_a	1.03_a	2.92_a
Within-domain	1154_b	1171_b	1.21_b	2.74_b
Within-superordinate	1334_c	1316_c	2.22_c	1.85_c
Average	1185	1215	1.49	2.50

Note. Numbers with different subscripts in a given column differ at $p < .0001$.

[x] High numbers reflect greater similarity

[y] High numbers reflect greater difference

Combined Analysis of experiments 1a – 1c

Analysis of Response Times for Difference and Similarity Judgments Under Three Presentation Conditions.

The data from all experiments were combined to assess the effects of the presentation-mode on reaction times in the similarity- and difference-judgment tasks. Average response times were calculated and entered into a 3 (Presentation) X 2 (Judgment) X 3 (Pair-Type) mixed ANOVA, with Presentation (Word-only, Word-Picture and Picture-only) and Judgment (Similarity and Difference) as between subjects variables, and Pair-Type (Across-domain, Within-domain, Within-superordinate) as a within subjects variables.

As expected, the main effect of Pair-Type was significant, $F(2, 250) = 251.5$, $p < .0001$. More importantly, the main effect of Presentation was significant, $F(2,125) = 10.91$, $p < .0001$. Reaction times in the Word-Picture condition (1138 ms) were faster than in the Picture-Only condition (1183 ms) and the latter were faster than reaction times in the Word-Only condition (1337 ms). Scheffe's post hoc analysis revealed that reaction times in the WO condition were significantly slower than reaction times in the PO and WP presentation conditions ($p < .001$). Why were judgments in the WO condition significantly slower than in the other conditions? This result is somewhat counterintuitive given that in the WP condition participants were presented with more information, and that in the PO condition they were confronted with what amounted to a naming task. There are several possibilities: it might be that under the different presentation conditions, participants made differential use of perceptual and conceptual information in their judgments; relying more heavily on pictorial data whenever it was available, thus being faster in those conditions that contained pictures. The second possibility, which is consistent with the perceptual-symbols hypothesis, is that participants under the three presentation conditions eventually constructed the same representation, but that this process took longer in the Word-Only condition.

If participants relied more heavily on pictorial information in the PO and WP presentation conditions, but more on conceptual knowledge in the WO condition, then similarity and difference judgments in the PO and WP conditions should be more similar to each other than to the WO condition. However, if similarity operated on similar representations which were constructed under all three presentation-modes, judgments in all three modes are expected to be quite similar. An analysis of the similarity and difference judgments was conducted to test these hypotheses.

Analysis of Similarity and Difference Judgments Under Three Presentation Conditions.

We now examined the effects of presentation mode on (a) judgments of similarity and difference and (b) the relation between the similarity and difference scales. We separately analyzed similarity and difference judgments given under the three presentation conditions. Due to the ordinal nature of the response scale, we used the variance of response proportions as the measure of test in the following ANOVAs. However, since the data are more easily encapsulated in the form of averages, we present them in Table 4.

Table 4: Mean similarity and difference ratings under three presentation conditions.

Condition	Mean Similarity	Mean Difference
Picture Only	1.489	2.507
Word Picture	1.548	2.543
Word Only	1.522	2.654

We first analyzed responses given in the *similarity-judgment* groups. The proportions of 1, 2 and 3 responses were analyzed in 3 separate one-way ANOVAs, with Presentation as a grouping factor and response proportion of each response as the dependent variable. No effect was found in any of these analyses ($Fs < 1$). The important result here is that there were practically no differences in the response proportions in the three presentation conditions. Proportions of the

three responses varied minimally between presentation conditions; in the range of 1.5%.

A similar analysis was performed on the response proportions in the *difference-judgment* groups. The proportions of 1 and 2 responses did not vary significantly between presentation conditions. However, an ANOVA of '3' responses was significant, $F(2,62) = 5.08$, $p < 0.01$. In the WO condition, there were more '3' responses (74%) than in the other WP and PO presentation conditions (69%, 68% respectively).

The similarity-rating data are congruent with the hypothesis that participants in all presentation conditions performed the similarity judgment on the same representation. Note however that while judgments across all conditions were similar, participants in the WO condition, who were not presented with the perceptual component in the stimuli, were slower to make their judgments.

In addition to this converging measure, the data that are perhaps most suggestive of the use of perceptual information in the WO condition is the fact that participants in the WO condition were also slowest in judgments of Across-domain items (e.g., Pan-Dog). One does not *need* to imagine a Pan and a Dog in order to determine that they are not similar at all, or very different. Such a decision could easily be made on the basis of categorical knowledge alone. However, judgments of similarity and difference for such pairs in the WO condition were slower than such judgments in the WP and PO conditions (1232, 1002, 1067 and 1245,1080,1059 ms for similarity and difference judgments respectively). The difference-judgment data are also congruent with the possibility that participants evoked perceptual data. As reported, difference judgments to Across-domain items in the WO condition were slower than in the other presentation conditions, suggesting that perceptual information was used here as well. However, the slight tendency to find more items 'very different' in the WO conditions may suggest that another mechanism was also at work here. We discuss the issue subsequently.

Analysis of Asymmetry Between the Scales Under Three Presentation Conditions

We also computed a measure relating the similarity and difference ratings under the three presentation conditions. To recap, in this analysis we are interested in the relation between an items' perceived similarity and its perceived difference under the three presentation modes. Since it is only sensible to talk of a degree of similarity in cases where similarity was found, we included in this analysis only items whose average similarity was greater than 1.

To compute this measure, for each item whose average similarity score was greater than 1, we compared its location on the similarity scale to its location on the difference scale. Location was defined as the distance of an item, in units of standard deviation, from the extreme end of the scale. That is, for items in the similarity scale we calculated the distance of each item, in units of standard deviation of the similarity scale, from the "least similar" end. For items in the difference scale we calculated the distance of each item, in units of standard deviation of the difference scale, from the "very different" end.

Let S_i and D_i denote, respectively, the distance of an item from the "least similar" end of the similarity scale and the "most different" end of the difference scale, in units of standard deviation. Delta, a measure of asymmetry, is the average of the differences between these two parameters across all items. The total asymmetry between the scales then is $[\sum (S_i - D_i)]/N$. Delta measures for all items, under the three presentation conditions were subjected to a one-way ANOVA with Presentation as a between-groups factor. The main effect of Presentation was significant, $F(2,511) = 40.3$, $p < .001$. Scheffe's post hoc comparisons revealed that all Delta measures differed significantly both from each other and from zero, p's $< .05$. Delta measures, the number of pairs on which Delta was greater than zero and the number of pairs on which Delta was lower than zero for each presentation condition are given in Table 5.

Table 5: Delta ratings and proportions for three presentation conditions.

Condition	Delta	Pairs where Delta > 0	Pairs where Delta < 0
Picture-Only	-0.08_a	57	108
Word-Picture	0.09_b	109	54
Word-Only	0.16_c	146	38

Note. Numbers with different subscripts in a given column differ at $p < .05$.

In the Picture-Only condition, items tended to be located 'farther' from the "very different" end of the difference scale than they were from the "not similar" end of the similarity scale. However, for the Word-Picture condition, and especially the Word-Only condition, items tended to be 'closer' to the "very different" end than they were to the "not similar" end. In short, negative Deltas in the Picture-Only conditions stem from more conservative ratings of difference compared to the similarity ratings given.

A possible explanation is that the presence of words resulted in more weight being given to conceptual knowledge in judgments. It has been suggested that in certain conditions, difference judgments are based on a comparison that involves aligning the structure of the items in the pair (Markman, 1996). Since conceptual

knowledge affords a basis for structural alignment, it might be that the presentation of words resulted in enhanced attention to these differences. For instance, a pair such as 'dog – cat' might be judged highly similar when presented under all presentation conditions. However, when greater weight is given to conceptual knowledge, one might remember that dogs bark and cats meow, that dogs can become attached to people and cats are territorial, and that dogs bite and cats scratch. Minding these differences could lead to more extreme difference ratings.

General Discussion

We set out to examine the effects of various presentation modes on perception of similarity and difference in order to establish what information is used in such judgments, and how the presentation modes affect the similarity and difference scales themselves. We used Reaction times, and similarity and difference ratings to address the first issue, and used Delta, a measure of asymmetry between the scales to address the second.

In short, two major findings stem from the three reported experiments: (1) When pictures were not present, reaction times were slower than in the other presentation conditions, but the judgments were comparable with them and (2) Deltas were positive whenever words were present, and were negative whenever words were absent. These data may be indicative of several important regularities.

First, people may rely on both perceptual and conceptual input when making similarity/difference judgments. In particular, their responses are faster when words are accompanied by pictures than when words are presented alone. It seems reasonable to hypothesize that they try to envision objects when those are depicted by words alone. Particularly suggestive of this, is the fact that compared to the Picture-Only and Picture-Word conditions, participants in the Word-Only condition were also slower in responding to Across-domain pairs. Other research on the role of perceptual information in higher-level cognitive tasks supports this possibility. Recent studies have demonstrated that accessing perceptual knowledge is an integral part of such tasks as property generation and property verification (Solomon & Barsalou, 2000; Wu & Barsalou, 2001; see also Barsalou, 1999) . Though the data support the hypothesis that people did use perceptual information, the issue can be conclusively resolved by obtaining measures comparing activity in respective areas of the brain.

Second, the Delta measure refers to what may be different processing considerations under the Word-Only and Word-Picture conditions, on the one hand, and the Picture-Only condition on the other. In particular, in the former two conditions, Deltas were largely positive, whereas in the latter condition they were largely negative. Recall that the more that difference is underestimated per a given similarity rating, the larger the Delta. Therefore, negative Deltas in the Picture-Only condition stemmed from more conservative ratings of difference compared to the similarity ratings given. These data may point to the effect of knowledge on the perception of difference. They also indicate that regardless of the pair type, judgments of similarity and difference are affected by the modality of input.

Overall faster response times for different pairs than for similar pairs are indicative of the fact that judgments "not similar" or "very different" are made whenever no sufficient similarity is found. At the same time, computation of the degree of similarity is a more lengthy process. The results reported here seem to establish a boundary condition for situations where alignment is used in similarity and difference judgments. Participants were faster to decide that objects are not similar than to decide that they are; suggesting that judgment of difference was not based on finding differences. We are currently in the process of setting up research for examining brain activity under the conditions reported here.

References

Barsalou, L. W. (1999). Perceptual symbol systems. Behavioral and Brain Sciences, 22(4), 577-660.

Battig, W. F., & Montague, W. E. (1969). Category norms of verbal items in 56 categories: A replication and extension of the Connecticut category norms. Journal of Experimental Psychology, 80(3/2), 1-46.

Gentner, D., & Markman, A. B. (1997). Similarity mapping in analogy and similarity. American Psychologist, 52(1), 45-56.

Markman, A. B. (1996). Structural alignment in similarity and difference judgments. Psychonomic Bulletin and Review, 3(2), 227-230.

Markman, A. B., & Wisniewski, E. J. (1997). Similar and different: The differentiation of basic-level categories. Journal of Experimental Psychology: Learning, Memory and Cognition, 23(1), 54-70.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. Journal of Experimental Psychology: General, 115(1), 39-57.

Solomon, K. O., & Barsalou, L. W. (2000). Grounding concepts in perceptual simulation: II. Evidence from property verification, Under Review

Tversky, A. (1977). Features of similarity. Psychological Review, 84(4), 327-352.

Wu, L. L., & Barsalou, L. W. (2001). Grounding concepts in perceptual simulation: I. Evidence from property generation, Under Review