On putting milk in coffee: The effect of thematic relations on similarity judgments.

Edward Wisniewski
Department of Psychology
Northwestern University
2029 Sheridan Road
Evanston, IL 60208
(847) 467-1624
edw@nwu.edu

Miriam Bassok
Department of Psychology
University of Chicago
5458 S. University Avenue
Chicago, IL 60637
(312) 702-1962
m-bassok@uchicago.edu

Abstract

All existing accounts of similarity assume that it is a function of matching and mismatching attributes between mental representations. However, Bassok and Medin (1996) found that the judged similarity of sentences does not necessarily reflect the degree of overlap between the properties of paired stimuli. Rather, similarity judgments are often mediated by a process of thematic integration and reflect the degree to which stimuli can be integrated into a common thematic scenario. We present results of a study which extend this surprising finding by showing that it also applies to similarity ratings of objects and occurs whether or not subjects explain their judgments. Also, consistent with the Bassok and Medin findings, the tendency towards thematic integration was more pronounced when the paired stimuli shared few attributes--but was still an important factor in similarity judgments between objects which shared many attributes. We discuss the implications of these findings for models of cognitive processes which use similarity as an explanatory construct.

Introduction

It is taken as self evident that the similarity between two things is a function of their matching and mismatching features. For example, in judging how similar a robin is to a canary a person might identify matching features such as "they are birds," "they have wings," "they fly," "they have beaks, " "they have feathers," and mismatching features such as "one is red and the other is yellow," "one eats worms and the other eats seeds. " This sense of similarity has played an extremely important role in theories of analogy, metaphor, categorization, concept learning, induction, and conceptual combination and is a major component of formal models of these processes. For example, concept learning models compute the likelihood that a novel instance belongs to category based primarily on how well the features of the instance match those in the category's representation. As a second example, in models of induction, a major factor which determines whether a feature associated with one category will be generalized to a second category is the degree of feature overlap between the category representations. Finally, in models of analogy, interpretation and inference are largely determined by the overlap between relational features in the base and target domains.

Much recent work on similarity judgments has investigated how people determine and weight these matching and mismatching features. Inspired by structural-alignment models of analogy (e.g., Gentner, 1983), this research has characterized similarity as a comparison process in which people attempt to align or put into correspondence mental representations (Goldstone, 1994a; Markman, A. & Gentner, 1993; Medin, Goldstone, & Gentner, 1993). This work has identified important constraints on how this alignment is carried out.

This view of similarity notwithstanding, Bassok and Medin (1996) found that similarity judgments do not necessarily reflect the degree of overlap between the properties of the paired stimuli. Rather, contrary to every existing account of similarity, similarity judgments are often mediated by a process of thematic integration and reflect the degree to which paired stimuli can be integrated into a common thematic scenario. Specifically, these researchers asked undergraduate students to rate the degree of similarity between various pairs of simple noun-verb-noun statements and to explain their ratings. When the statements had common verbs (e.g., "The carpenter fixed the chair" and "The electrician fixed the radio") the explanations were consistent with the prevalent definition of similarity (e.g., "Similar because in both statements a professional is doing his job). However, when the statements had common nouns (e.g., "The carpenter fixed the chair" and "The carpenter sat on the chair"), 61% of the explanations generated by the participants were causal or temporal scenarios (e.g., "Similar because the carpenter sat on the chair to see whether he had fixed it well").

Thematic similarity has been previously documented only in young children (see E. Markman, 1989, for a review) and in adults from illiterate cultures (Luria, 1976). In one study, Luria (1976) had illiterate adults compare pairs of objects, asking them what the objects had in common and in what way they were alike. These adults sometimes responded by noting thematic relations between the objects. For example, when one illiterate adult from Uzbekistan was asked by Luria, "what do water and blood have in common?" he responded: "What's alike about them is that water washes off all sorts of dirt, so it can wash off blood too" (p. 82). Given that the undergraduate students in the Bassok and Medin (1995) study were neither young nor illiterate, their results are extremely surprising. That is, such results cannot be simply explained away by lack of knowledge or insufficient understanding of the task at hand. If these results reflect a robust and prevalent tendency to use a process of thematic integration for computing similarity then they should be seriously considered by theoretical accounts of similarity and cognitive processes which rely on

similarity as an explanatory construct. Our study follows up on these findings in order to examine their validity and generality. It was designed to answer three interrelated follow-up questions:

(1). Are thematic scenarios unique to comparison of statements ? It is possible that thematic integration mediates similarity judgments only when the stimuli are statements, because statements may induce a tendency for story construction. Our study examined whether thematic integration also mediates similarity judgments for simple pairs of object (e.g., milk-coffee; milk-lemonade). That is, we examined whether and to what extent people consider two things to be more similar if, in addition to having matching attributes (e.g., both are beverages), they are thematically related (e.g., one puts milk into coffee).

(2). Is thematic intergation more likely when it is difficult to align stimuli? It is possible that thematic integration affects similarity judgments only when the paired stimuli cannot be aligned in a satisfactory way. Medin and Bassok (1996) found that thematic scenarios were very rare when the paired statements could be structurally aligned (6% for matching verbs). Extending this logic to similarity between objects predicts a greater tendency for thematic integration when the stimuli can only be poorly aligned (e.g., milk-cow share very few common attributes) than when they can be readily aligned (e.g., milk-coffee share many common attributes). To test this possibility, our study examined the relative effect of thematic integration for object pairs that shared either many or few common attributes.

(3). Does thematic integration affect similarity judgments when people do not explain their ratings? The results of Bassok and Medin (1996) are based on an analysis of explanations that accompanied similarity ratings. However, explanations may change the process by which people arrive at similarity judgments and therefore either over or underestimate the prevalence of thematic scenarios. Alternatively, one could argue that thematic scenarios appear only in post-hoc explanations of judgments but do not affect the process by which people construe similarity (e.g., Nisbett & Wilson, 1977). Our study examined whether thematic integration actually affects similarity ratings, and whether such effects differ when people explain and do not explain their ratings.

Method

Participants. 128 Northwestern University undergraduates participated as part of a course requirement.

Materials. The stimuli were 12 quintuplets of objects, each consisting of a base and four targets. Two targets shared many attributes with the base (M) and the other two shared few attributes with the base (F). The M and T targets in each pair were chosen to be equally similar to the base in terms of their common attributes. However, one of the targets in each pair was thematically related to the base (T) and the other was not. That is, the four targets in each quintuplet had the following structure: MT, M, FT, F. The 12 quintuplets used in our study appear in Table 1.

Four lists of 12 base-target pairs were created by randomly selecting three base-target pairs of each of the four types (MT, M, FT, F), subject to the constraint that only one base-target pair from each quintuplet appeared in a list. The 12 base-target pairs of each list were then typed on sheets of paper, each above a seven-pont similarity rating scale, two per page, in a random order. This procedure yielded one six-page rating booklet per list. In addition, the pages of each form were put in reverse order to yield two alternative booklets per list.

Procedure. Participants read instructions telling them that they would see some pairs of common, everyday things, and that they would have to rate how similar the two things are. If a subject thought a pair of things was very similar, they should circle a 7. If they thought they were not at all similar they should circle a 1. Subjects were instructed to use the other numbers between 1 and 7 to indicate in between degrees of similarity. Participants in the Explanation condition were further instructed to write down an explanation for their rating in the space below the rating scale, i.e., to explain why they thought the two things had the degree of similarity that they did.

Participants were run in groups of 2-4, and each group was randomly assigned to either the Explanation (N = 64) or No Explanation condition (N = 64). Each of the four base-target lists was rated by 16 participants in each condition (8 participants per each order of the rating booklet). The task took about 15 minutes to complete.

Base	MT target	M target	FT target	F target
milk	coffee	lemonade	cow	horse
ship	lifeboat	canoe	sailor	soldier
car	tow truck	pickup truck	mechanic	plumber
chair	table	bed	carpenter	electrician
telephone	ans. machine.	tape recorder	receptionist	waitress
tie	suit	dress	man	woman
chisel	hammer	screwdriver	sculpture	painting
cat	mouse	hamster	veternarian	pediatrician
cup	kettle	pan	tea	wine
fly	spider	beetle	screen	curtain
peanut butter	jelly	cream cheese	knife	fork
apple pie	ice cream	jello	baker	tailor

Table 1: Quintuplets of base and target objects used in the study

Results

Similarity ratings. We performed a 2 X 2 X 2 mixed ANOVA with condition (explanation versus no explanation) as the between-subjects factor and thematic relation (absent versus present) and attribute overlap (many versus few) as the within-subjects factors. Table 2 shows the average similarity ratings in the Explanation and No Explanation conditions for the four types of base-target pairs (MT, M, FT, F).

As one would expect, similarity ratings for targets with many common attributes (M and MT) were significantly higher than for targets with few common attributes (F and FT), $F(1, 127) = 399.34$, $MSE = 642.2$, $p < .001$. Importantly, consistent with the findings of Bassok and Medin (1996), the existence of a thematic relation between the base and target signifcantly affected similarity ratings (MT + FT > M + T), $F(1, 127) = 66.63$, $MSE = 81.21$, $p < .001$.

The interaction between attribute overlap and thematic relation was also highly significant ($F[1, 127] = 75.87$, $MSE = 43.41$, $p < .001$). That is, the effect of thematic relations on similarity judgments was significantly larger when the targets had few than when they had many attributes in common with the base ([FT-F] > [MT-M]), although even the difference between the ratings for the MT and M pairs was statistically significant ($t[127] = 2.38$, $SME = .089$, $p < .02$). These results indicate that thematic integration can mediate similarity judgments even when the base and target can be readily aligned in terms of their common attributes.

As can be seen by comparing the two columns of Table 2, there was no reliable difference between the Explanation and No Explanation conditions ($F < 1$) and condition did not interact with either attribute overlap, $F(1, 127) = 1.64$, $MSE = 2.64$, $p < .21$, or with thematic relations ($F < 1$). Of importance, there was no three way interaction between these variables ($F < 1$). That is, explanations did not affect either the pattern or the magnitude of similarity ratings.

In general, the findings held across the particular items. An ANOVA on the item means revealed the same pattern of significant and nonsignificant findings as in the subject ANOVA. We also compared the rating of each FT pair with its corresponding F pair and each MT pair with its corresponding M pair. Every FT pair had a higher similarity rating than its F pair. However, only half of the MT pairs had higher ratings than their corresponding M pairs. Thus, unlike the subject mean ratings, the difference between the item mean ratings for the MT and M pairs failed to reach statistical significance, ($t[11] = 1.35$, $SME = .163$, $p < .21$).

Because the presence versus absence of thematic relations was a within-subjects factor, it is possible that the results could be explained by a demand characteristic. In particular, subjects may have noticed obvious thematic relations between some object pairs and not others and assumed that they should give higher ratings to those items which shared the thematic relation. If subjects were using this strategy, then the difference between the ratings of the thematic and non-thematic pairs (e.g., FT vs. F, between subjects) should be relatively higher in the second than in the first half of the rating form. This finding, however, was not obtained: There was no statistically significant interaction between presentation order (first versus second half) and thematic relation (absent versus present).

Justifications. Given that there was no difference between the Explanation and No Explanation conditions, the explanations generated by participants can shed further light on the effect of thematic integration on similarity judgments. Below we report results from a preliminary analysis of the explanations. This preliminary analysis was performed by one of the authors and awaits validation by independent judges.

The explanations were coded into three categories: Thematic, Attributional, and Uninformative. Explanations were coded as Thematic if they explicitly referred to a thematic relation between the base and the target, regardless of whether they also included references to attributional matches and mismatches. Examples included: "some people put milk in their coffee" and "you don't milk a horse." In some cases, thematic justifications also included attributional matches (e.g., "milk is produced from a cow and cows can be whitish like milk is in color"). In the present analysis we do not distinguish between purely thematic explanations and mixtures of thematic and attributional explanations. Explanations were coded as Attributional if they only included references to attributional matches and mismatches. Examples included: "they (milk and coffee) are both liquids that one can consume and digest," "both (milk and lemonade) are refreshing drinks," "both (milk and cow) can be consumed and both can be white," "(milk and horse) are not similar, one is liquid the other solid, one is living the other inanimate." Explanations were coded as Uninformative when they included general statements such as "there are no similarities I can see," "they are related." Blanks (i.e., no explanation) were also coded as Uninformative. Table 3 shows the distribution of these three types of explanations for the four types of base-target pairs (MT, M, FT, F).

Base-Target correspondence	Explanation	No Explanation
MT: Many attr. + Thematic relation	4.88	4.83
M: Many attributes	4.71	4.54
FT: Few attrib. + Thematic relation	3.12	3.29
F: Few attributes	1.79	1.84

Table 2: Average similarity ratings for the four types of base-target correspondence.

Base-Target correspondence	Thematic	Attributional	Uninformative
MT: Many attr. + Thematic	29%	69%	2%
M: Many attributes	2%	96%	2%
FT: Few attrib. + Thematic	71%	19%	10%
F: Few attributes	32%	40%	28%

Table 3: Percentage of explanation types generated for the four types of base-target pairs.

As shown in Table 3, the distribution of the explanations is consistent with the pattern of thematic effects documented by the ratings (see Table 2). First, there were higher proportions of thematic explanations for items which shared a thematic relation (50% for MT and FT) compared to those which did not (17% for M and F). In corresponding fashion, the similarity ratings were higher for items which shared a thematic relation. There also was a general tendency towards thematic integration when items shared few (52% for FT and F) versus many attributes (16% for MT and M). Finally, the distribution of thematic explanations for targets with many vs. few attributes mirrored the interaction between attribute overlap and thematic relations observed for the ratings. Specifically, the difference in frequency of thematic explanations between the FT and F targets (39%) was greater than that between the MT and M targets (27%).

Of interest, when the base and target could not be readily aligned in terms of common attributes, participants generated many thematic explanations even for the F targets that did not bear an obvious thematic relation to the base (32%). Participants were sometimes quite creative in thematically relating these items to explain why they were similar (e.g., "a fancy ship could employ a tailor"). These explanations were the closest to those found by Bassok and Medin (1996), because the statements used in their study were not chosen to be thematically related (i.e., did not fit familiar scripts). Other thematic explanations for the F targets seemed to reflect an implicit contrast with familiar thematic relations (e.g., "you don't milk horses," "women do not usually wear ties").

Of importance, note that thematic explanations were not a peculiarity of a few participants. Rather, 89% of the 64 participants in the Explanation condition generated at least one thematic relation in explaining their similarity ratings. The median number of thematic explanations per participant (out of 12 responses in their rating booklets) was 3.5, and the average was 5.1. Moreover, only a handful of these mature and educated participants (about 5%) explicitly pointed out that these thematic relations do not make the paired objects similar to each other.

Discussion

Our findings support and extend Bassok and Medin's (1996) surprising finding that, in addition to comparison of common and distinctive attributes, similarity judgments are mediated by the thematic integration of the paired stimuli into a common scenario. First, the current study shows that the effect of thematic relations on similarity judgments is not unique to ratings of sentences but applies to ratings of objects as well. Moreover, thematic integration is not unique to situations in which people explain their ratings. In fact, participants who did not explain their judgments gave nearly identical ratings to the same items (see Table 2), indicating that thematic relations affected their judgments as well. The effect of thematic integration on similarity judgments was apparent both in the magnitude of similarity ratings (Table 2) and in the explanations that accompanied these ratings (Table 3).

Also, consistent with the results which Bassok and Medin (1996) found for sentences, thematic integration was much higher when the stimuli could not be readily aligned. Specifically, thematic relations carried more weight in similarity judgments when the paired objects shared few than when they shared many common attributes. Nevertheless, our findings argue against the view that thematic relations only play a role in similarity judgments when there is little attributional similarity. Even among items that shared many attributes, almost one-third of the subjects listed thematic relations, and their ratings were higher for items which shared many attributes and a thematic relation compared to those which only shared many attributes (i.e., the MT versus M pairs). This finding is especially striking when one notes that the MT items (e.g., milk-coffee) shared many attributes but only a single thematic relation. Yet, this single thematic relation influenced similarity ratings and explanations.

One could argue that, despite our efforts to equate the number of common attributes in the paired targets (MT and M, FT and F), those which were thematically related to the base also had more common attributes with the base. We are currently collecting data to test this possibility, but we believe that it is very unlikely. We have divided each explanation into specific reasons given to explain why the pairs were similar or different (excluding uninformative reasons, we coded 1171 such reasons). The average number of attributional matches for the MT targets was actually slightly lower than for the M targets (1.2 vs. 1.5 matches per item, such as "both liquids"). Similarly, the average number of attributional matches for the FT targets was slightly lower than for the F targets (.19 vs. .23 matches per item, such as "both can be white,"). Thus, if there were differences in shared attributes between items, it would appear that they would work against our hypothesis. Obviously, it is difficult to explain why the ratings of the FT items (e.g., milk cow) were approximately 1.75 times higher than those of the F items (e.g., milk-horse) by suggesting that it was only due to differences in the number of common attributes. Recall also that the rating of every FT item was higher than its corresponding F item. It is unlikely that we unintentionally chose every FT item to have more common attributes when we purposely attempted to avoid such discrepancies.

These findings raise two important questions. First, what is the best processing level account of these results? One possibility is that people easily notice the salient, pre-existing thematic relations between our stimuli. These relations then influence judgments (perhaps because the meaning of the word "similarity" is ambiguous). However, Bassok and Medin (1996) showed that in explaining their similarity ratings, subjects actually constructed novel thematic relations between sentences. Similarly, novel thematic integrations were generated for the F targets in our study. Another possibility is that people's responses reflect an interaction between processes that are stimulus-driven and those that are driven by the type of task. That is, properties of the stimuli (e.g., having common attributes, being thematically related) may be difficult to override even though the task instructions are incompatible with those properties. In a recent sudy of ours, subjects who were asked to list the commonalities and differences between pairs of objects sometimes listed thematic relations even though the instructions described an example in which the commonalities and differences only involved attributes.

Second, what are the theoretical implications of these results? Most researchers are interested in factors which influence similarity judgments to the extent that these factors are relevant to cognitive processes which rely on similarity, such as similarity-based learning, categorization, conceptual combination, probabilistic reasoning, etc. Currently, we are examining whether and how thematic integration affects such processes. Previous work suggests that, indeed, probability judgments (Tversky & Kahneman, 1983) and conceptual combination (Wisniewski, in press) are sometimes mediated by thematic associations rather than by feature comparison. Clearly, future work needs to address these questions in order to understand the role that thematic relations play in processes that use similarity and to clarify the extent to which models of these processes must be extended.

References

Bassok, M., & Medin, D. L. (1996). Birds of a feather flock together: Similarity judgments with semantically-rich stimuli. Manuscript submitted for publication.

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

Goldstone, R.L. (1994). Similarity, interactive activation and mapping. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20(1), 3-28.

Luria, A. R. (1976). Cognitive development: Its cultural and social foundations. Cambridge, MA: Harvard University Press.

Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. Cognitive Psychology, 23, 431-467.

Markman, E. M. (1989). Categorization and naming in children. Cambridge, Mass.: MIT Press.

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

Nisbett, R. E., & Wilson, T.D. (1977). Telling more than what we can know: Verbal reports on mental processes. Psychological Review, 84, 231-259.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. Psychological Review, 90, 293-315.

Wisniewski, E. J. (in press) Construal and similarity in conceptual combination. Journal of Memory and Language.