What Makes a Man Similar to a Tie? Stimulus Compatibility with Comparison and Integration

Edward J. Wissinowski
University of North Carolina at Greensboro

and

Miriam Bassok
University of Washington

We argue and show that different properties of stimuli are comparable with different types of processing. Specifically, object pairs from the same taxonomic category (e.g., chair:bed) tend to be integratable and thus comparable with comparison, whereas object pairs that play different roles in thematic relations (e.g., chair:carpenter) tend to be nonintegratable and compatible with integration. Using object pairs that varied orthogonally in integratability and thematic relatedness, we demonstrated that stimulus compatibility modulates processing and affects the outcomes of tasks that are currently believed to involve only comparison (similarity ratings, Experiment 1; listing connotativeness and differences, Experiment 2) or only integration (thematic relatedness ratings, Experiment 3). Our findings and others that we have reviewed suggest that: (1) many cognitive tasks involve both comparison and integration, and (2) the relative influence of each process is modulated by an interplay between the task-appropriate and the stimulus-compatible process. We believe that single-process models should be extended to take this interplay into account.

Why is it that “apples and oranges” became an idiom for a bad comparison? (An analogous expression in Hebrew contrasts wheat and barley.) Wouldn’t a pair such as “apples and baskets” be an even better example? As noted by Bassok (1997), “apples and oranges” is an excellent example.

Preliminary results of the first study appeared in Wissinowski and Bassok (1996). We thank Larry Barsalou, Dorell Billman, Danie Greene, Maritza Isac, Gregory Murphy, Jerry Regier, Lyle Rips, Steven Sloman, and especially Doug Medin for stimulating discussions of this work.

Address reprint requests to Edward Wissinowski, University of North Carolina, Department of Psychology, P.O. Box 26164, Greensboro, NC 27402-6164, e-mail: edw@uncg.edu, or to Miriam Bassok, University of Washington, Department of Psychology, P.O. Box 351525, Seattle, WA 98195-1525, e-mail: miriam@uwashington.edu.

Copyright © 1999 by Academic Press. All rights of reproduction in any form reserved.
of a pair of highly alignable entities. That is, apples and oranges share many dimensions on which they can be compared, including those with similar values (e.g., superordinate category, shape, size, parts) and ones with different values (taste, color, thickness of skin). The very ease of this comparison might lead people to commit "near-miss" errors (i.e., confuse "apples" with "oranges" because of their commonalities) when they should have focused on the differences. At the same time, it is very unlikely that people would even attempt to compare (and therefore confuse) such nonalignable entities as apples and baskets, which share considerably fewer common dimensions. Rather, people's knowledge about apples and baskets might lead them to integrate these entities into a joint theme or scenario (e.g., place apples in baskets).

The above example highlights the inherent interdependence between stimulus type and processing type and suggests that different types of stimuli are compatible with different types of processing. In particular, it suggests that taxonomically related concepts such as apples and oranges are compatible with the process of comparison, whereas taxonomically related concepts such as apples and baskets are compatible with the process of integration. This conjecture is consistent with people's everyday experience. The entities that people usually compare tend to be alignable (e.g., "we bought the Saturn because it is safer than the Dodge Neon"), while comparisons of nonalignable entities are often meaningless or useless (e.g., "the Saturn is safer than a car key"). At the same time, the entities people usually integrate into themes or scenarios tend to be nonalignable (e.g., "wiring a car key to start the Saturn"). Because each entity in a thematic relation plays a different role, which typically requires a different set of properties, these roles are unlikely to be played by alignable entities (e.g., using the Dodge Neon to start the Saturn).

People's different interactions with alignable and nonalignable entities are reflected in the organization of their semantic knowledge, which is known to be structured on the basis of taxonomic and thematic relations. Entities that are highly alignable belong to one or more categories from the same taxonomy (e.g., Markman & Wosniowski, 1997) whereas nonalignable entities are clustered in categories that capture thematic relations such as events (e.g., dining, baseball game; Bower, Black, & Turner, 1979), places or scenes

Of course, there are exceptions to these generalizations. First, some thematic relationships can involve highly alignable entities. A Dodge Neon could be used to start a Saturn if the latter had a dead battery that needed charging. Also, cars can be thematically related via a car crash. However, thematic relations between highly alignable entities are rare (see Wosniowski and Love, 1998, for evidence). Second, some thematic relationships are symmetrical (e.g., games with [MacMillan, Milly] in which alignable entities play similar or symmetric roles. In our studies, we only used stimuli with provoking symmetrical relations. It is also likely that the latter are much more common.
(e.g., gasoline station, zoo), and ad hoc categories (e.g., “things to take out of a boning house”; Barrauds, 1983).

The experiments that we will describe in this paper provide evidence that, consistent with the organization of their semantic knowledge, people tend to compare taxonomically related entities and to integrate schematically related entities. In light of the previous discussion, our results should not be very surprising. However, because we show that stimulus compatibility with comparison and integration can override task instructions and the "appropriate" process for a task, our results appear to be utterly counterintuitive. For example, the participants in our studies often integrated taxonomically related entities when judging their similarity (e.g., a man is similar to a tie because he might wear a tie) and compared taxonomically related entities when judging thematic relatedness (e.g., milk is thematically related to lemonade because both are beverages). Such responses strike one as being "judgment errors" rather than being ecologically valid responses to incompatible stimuli. In fact, because the participants in our studies were college undergraduates, their responses seem to contradict the common view that such "judgment errors" are committed only by young children (see Lin, 1996; Markman, 1989, for reviews) and adults from illiterate cultures (Luriú, 1976).

Our results may reflect conditions under which people spontaneously select processes—often in daily life—people are not given task instructions but rather, processing is determined by properties of the stimuli around them. The findings are surprising only in the context of extant single-process models of cognitive tasks. We believe that such models overestimate the extent to which people’s performance is mediated by controlled, task-appropriate processing and neglect the impact of uncontrolled, stimulus-appropriate processing. As detailed in the General Discussion, most models take a stance as to whether integration or comparison is the appropriate process for their task (categorization, induction, conceptual combination, metaphor, etc.). They then evaluate their models by selecting stimuli that tend to be compatible with the hypothesized process, most probably without any explicit awareness of doing so.

Even when researchers do examine stimuli that are incompatible with the hypothesized process, they still assume that this process mediates people’s performance. For example, recent models of similarity assume that it is always determined by a comparison process, which primarily identifies commonalties (same values on common dimensions) and alignable differences (different values on common dimensions) between the compared entities (e.g., Goldstone & Medin, 1994; Markman & Gentner, 1995a, 1995b; Medin, Goldstone, & Gentner, 1993). On this view, and consistent with our stimulus-compatibility argument, the comparison process favors stimuli that have some degree of alignability. However, contrary to our stimulus-compatibility argument, extant models of similarity also hold that people compare non-
alignable entities. Such entities are considered not similar because the comparison process fails to determine any basis for their similarity (i.e., any dimensions on which they can be compared). Much laboratory work has been interpreted as support for this view: subjects tend to compare stimuli along alignable dimensions (Markman & Gentner, 1993a, 1993b; Markman & Wisniewski, 1997) and prefer to compare alignable rather than non-alignable stimuli (Gentner & Markman, 1994).

This process view of similarity notwithstanding, Bassok and Medin (1997) surprisingly found that in making similarity judgments subjects not only compared alignable stimuli but also integrated nonalignable stimuli into a theme or scenario. As other similarity theorists, these researchers began their studies with the premise that similarity is solely determined by a comparison process and were interested in finding out whether people differentially weight relations versus attributes in making similarity judgments. To this end, Bassok and Medin varied the alignability of pairs of noun–verb–noun statements. Undergraduates then rated the similarity of these pairs and explained their ratings. Indeed, when the statements shared a common verb and hence were alignable (e.g., "The carpenter fixed the chair" and "The electrician fixed the radio"), most subjects compared the paired statements (e.g., "Similar because in both statements a professional is doing his job"). However, when the statements differed in their verbs and hence were poorly alignable (e.g., "The carpenter fixed the chair" and "The carpenter sat on the chair"), almost two thirds of the subjects integrated the statements into a joint thematic scenario (e.g., "Similar because the carpenter sat on the chair to see whether he had fixed it well"). That is, contrary to the single-process comparison view of similarity, the subjects also integrated stimuli in making similarity judgments.

Most relevant to this paper, the Bassok and Medin (1997) findings showed a stimulus compatibility effect on processing. Specifically, it appears that subjects construed statements with common verbs as instances of the same kind of action. Such actions are taxonomically related and hence highly alignable, leading subjects to compare them. In contrast, subjects construed statements with different verbs but the same subject and object as different kinds of actions being carried out by the same agent on the same entity. Such actions are taxonomically unrelated and therefore less alignable. At the same time, they readily suggest that the agent is carrying them out in sequence in the service of a goal. These stimulus properties led subjects to schematically integrate the actions. The present work on the respective com-

1 Two-thirds of Bassok and Medin’s statements involved verbs that named actions and subjects that named agents (as in the examples that we present). One third of the stimuli were comparative statements involving subjects that named inanimate objects (e.g., the cathedral is higher than the church). For ease of exposition, we omit discussion of these latter items though they yielded a similar pattern of findings that also reflected a stimulus compatibility effect.
patibility of taxonomic and thematic relations with comparison and integra-
tion was motivated by these surprising findings.

Overview of the Experiments

In the present studies we further investigate the effects of stimulus compat-
ibility on processing. Our first experiment followed up Bassok and Medin’s 
but differed in several important ways. First, the generality of Bassok and 
Medin’s findings is unclear. These researchers examined similarity judg-
ments between statements that primarily described actions. However, actions 
may be especially likely to induce a tendency toward constructing a story 
organized by a theme (Bower, Black, & Turner, 1979; Schank & Abelson, 
1977). To examine the generality of Bassok and Medin’s findings, we exam-
ined similarity judgments for pairs of objects (e.g., milk—lemonade). Second, 
the Bassok and Medin findings are based on justifications provided by partic-
ipants as explanations for their similarity ratings. Verbalizing reasons for 
making such ratings may have changed the usual process by which people 
arrive at similarity judgments (see Wilcoo & Brekke, 1994, for a review of 
how cognitive processes are changed by verbalization). Thus, our first study 
also compared possible effects of stimulus compatibility under conditions in 
which participants did or did not explain their ratings.

Finally and most importantly, Bassok and Medin (1997) did not explicitly 
design their study so as to make predictions about the effects of stimulus com-
patibility on similarity judgments. We intentionally designed our stimu-
lus pairs to vary in terms of both taxonomic and thematic relatedness. Spe-
cifically, our stimuli were chosen so that pairs of objects were either taxo-
nomically related and hence highly alignable (e.g., milk—lemonade) or not 
taxonomically related and hence poorly alignable (e.g., milk—horse). We also 
selected our stimuli so that they varied in terms of thematic relatedness. Our 
measure of thematic relatedness was simply whether a pair of objects is 
linked or not linked by a preexisting thematic relation (e.g., milk—coffee vs 
milk—lemonade; milk—cow vs milk—horse, respectively). This measure 
has been used in the developmental literature to examine how children form 
categories (see Markman, 1989, for a review).

In our first study, we predicted that precuing taxonomic and thematic 
relations would invoke different processing modes in judging similarity be-
 tween objects. In particular, given our previous discussion, we expected sub-
jects to compare highly alignable objects and to integrate poorly alignable 
objects. We also predicted that type of precuing and similarity ratings 
would be affected by preexisting thematic links. Note that Bassok and Medin 
(1997) did not manipulate thematic relatedness. Rather, the participants in 
their study spontaneously constructed novel thematic links between poorly 
alignable stimuli. We predicted that preexisting thematic relations between 
nonalignable stimuli would further increase people’s tendency to con
similarity via integration, rather than via comparison. This prediction is consistent with research which suggests that preexisting thematic relations are readily accessible and hence may affect subsequent processing. For example, in the Palermo and Jenkins (1969) word-association task, frequently produced associates for common objects were either thematic relations (e.g., sleep for bed) or thematically related objects (e.g., pillow for bed; Markman & Wosinska, 1997). As another example, Barsalou (1982) found that reading the name of an artifact (e.g., gun) automatically activates its function (e.g., shoots), independent of context.

Our first study also examined whether alignability and preexisting thematic relations interact. In particular, we expected comparison to be most often invoked when the stimuli are highly alignable and have no preexisting thematic links (e.g., milk–lemonade). On the other hand, we expected thematic integration to be most often invoked when the stimuli are poorly alignable and have preexisting thematic links (e.g., milk–cow). Mixtures of the two processing modes should occur for stimuli that possess other combinations of these factors, i.e., stimuli which are highly alignable but also thematically related (e.g., milk–coffee) and stimuli that are poorly alignable but not thematically related (e.g., milk–horse).

Two additional studies used the same stimuli from Experiment 1 and assessed the degree to which stimulus compatibility interacted with task instructions. The instructions emphasized the appropriateness of either comparison (Experiment 2) or integration (Experiment 3). In Experiment 2, we highlighted the comparison process by instructing participants to list matching and mismatching features as commonalities and differences between entities and showing them clear examples of such features. In Experiment 3, we highlighted integration by instructing participants to rate the degree to which two entities could be thematically related and showing them clear examples of entities that were thematically related. We examined how automatic and thematic relations between the stimuli interacted with these task instructions to influence the relative tendency to use comparison versus integration.

To foreshadow the results, we found systematic interactions between stimulus compatibility and processing in all three experiments. Specifically, both processing mediated subjects’ performance, with stimulus compatibility modulating the relative influence of comparison and integration. Instructions which highlighted the appropriate process for a task did not eliminate the occurrence of the inappropriate process. In the General Discussion, we review findings in the literature which suggest that stimulus compatibility modulates the influence of comparison and integration across a wide range of cognitive tasks typically assumed to involve only one of these processes. All of these tasks may be influenced and shaped by these two modes of processing. Our analysis will highlight the need for researchers to formulate two-process models that are sensitive to stimulus compatibility.
In the first study, we assessed the relative tendency to use comparison versus integration in making similarity judgments by orthogonally varying whether pairs of objects were taxonomically related and functionally related. This procedure results in four kinds of object pairs. For example, milk and coffee are taxonomically related and hence highly alignable. They also share a preexisting thematic relation in that people often put milk in their coffee. (As one would expect from our earlier analysis, it was very difficult to find taxonomically related objects which were also thematically related.) Milk and lemonade are taxonomically related but do not share a preexisting thematic relation. Milk and cow in turn come from different taxonomic categories. On the other hand, they share a preexisting thematic relation (i.e., cows produce milk). Finally, milk and horse are neither taxonomically nor thematically related.

One group of participants rated the degree of similarity between pairs of objects. A second group rated these same pairs but were also instructed to explain why they gave the similarity rating that they did for a pair of objects. We predicted that in making similarity judgments different modes of processing would be invoked by stimulus compatibility with a process. That is, objects which are taxonomically related and hence easy to align will more readily invoke the comparison process whereas those which are not taxonomically related and hence more difficult to align will more readily invoke integration. In addition, we examined whether these predicted effects are influenced by having participants explain their similarity judgments.

The experimental design also allowed us to make more specific predictions about both the similarity ratings and their justifications. In particular, our compatibility hypothesis predicts that there should be a systematic relationship between similarity ratings and processing of the different types of object pairs. We expected participants to rate highly alignable pairs (e.g., milk-lemonade) as more similar than poorly alignable pairs (e.g., milk-horse) because the former pairs have many more dimensions in common. All models of similarity predict this result as well. However, unlike similarity models, we also predicted that participants would give higher similarity ratings to object pairs which they thematically integrated. This prediction is consistent with Bassok and Medin’s (1997) view that people’s sense of similarity is based on the overall relatedness of items, including both common features and thematic links.

Pairs of objects which are easy to align and are not thematically related are the most compatible with comparison and the least compatible with integration. Therefore in their explanations for these pairs, participants should list the most responses consistent with comparison and the fewest responses consistent with integration. Further, pairs which are thematically related and
not easily aligned should be the most compatible with integration and the least compatible with comparison. Therefore, for these pairs participants should find the most responses consistent with integration and the fewest responses consistent with comparison. For the other two types of object pairs, responses consistent with either comparison or integration should fall in between these extremes. Pairs which are highly alignable but also thematically related have a characteristic that is compatible with comparison but also a characteristic that is compatible with integration. Pairs which are poorly alignable but also not thematically related have one characteristic which is compatible with integration but another characteristic which is incompatible with integration.

Method

Participants. One hundred twenty-eight Northwestern University undergraduates participated as part of a course requirement.

Materials. The visual set was 12 quintuplets of objects, each consisting of a base and four targets. A target was either taxonomically related and high in similarity (denoted as A+) or taxonomically unrelated and low on similarity (denoted as A−). Furthermore, a target either shared a thematic relation with the base (*T−); low on thematic relativeness) or did not share a thematic relation with the base (T−; low on thematic relativeness). This set, the four targets in each quintuplet had the following structure: A+ T−, A− T−, A+ T+, and A− T+. The 12 quintuplets appear in Table 1. Within a quintet, we selected the two targets which readily aligned with the base to be approximately equally similar to the base in terms of common features. For example, the targets lieutenant and coffee readily align with the base milk. Each of these pairs shares about the same number of common features with milk. Thus, of the thematically related target which readily aligned with the base (i.e., coffee) was judged more similar than the thematically unrelated target which readily aligned with the base (i.e., lieutenant), then we could be fairly confident that the difference in similarity was due to integration. Similarly, we selected the two targets which poorly aligned with the base in

<table>
<thead>
<tr>
<th>Base</th>
<th>A+T+</th>
<th>A+T−</th>
<th>A−T+</th>
<th>A−T−</th>
</tr>
</thead>
<tbody>
<tr>
<td>milk</td>
<td>coffee</td>
<td>lemonade</td>
<td>cow</td>
<td>hops</td>
</tr>
<tr>
<td>ship</td>
<td>levee</td>
<td>cause</td>
<td>sailor</td>
<td>soldier</td>
</tr>
<tr>
<td>car</td>
<td>row truck</td>
<td>pickup truck</td>
<td>mechanic</td>
<td>plumber</td>
</tr>
<tr>
<td>telephone</td>
<td>table</td>
<td>bed</td>
<td>carpenter</td>
<td>electrician</td>
</tr>
<tr>
<td>tie</td>
<td>am machine</td>
<td>tape recorder</td>
<td>reception</td>
<td>waitress</td>
</tr>
<tr>
<td>closed</td>
<td>suit</td>
<td>dress</td>
<td>mat</td>
<td>woman</td>
</tr>
<tr>
<td>cat</td>
<td>hummer</td>
<td>screwdriver</td>
<td>sculptor</td>
<td>painting</td>
</tr>
<tr>
<td>cap</td>
<td>house</td>
<td>hammer</td>
<td>veterinarian</td>
<td>pediatrician</td>
</tr>
<tr>
<td>fly</td>
<td>kettle</td>
<td>pan</td>
<td>tea</td>
<td>wine</td>
</tr>
<tr>
<td>peanut butter</td>
<td>jelly</td>
<td>beetle</td>
<td>screen</td>
<td>curtain</td>
</tr>
<tr>
<td>apple pie</td>
<td>ice cream</td>
<td>cream cheese</td>
<td>knife</td>
<td>fork</td>
</tr>
<tr>
<td></td>
<td></td>
<td>jello</td>
<td>baker</td>
<td>tailor</td>
</tr>
</tbody>
</table>
TABLE 2
Average Similarity Ratings for the Four Types of Base–Target Correspondence
(Experiment 1)

<table>
<thead>
<tr>
<th>Base–target correspondence</th>
<th>Explanation condition</th>
<th>No explanation condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>A → T+: high alignability + thematic relation</td>
<td>4.88</td>
<td>4.83</td>
</tr>
<tr>
<td>A → T+: High alignability</td>
<td>4.71</td>
<td>4.54</td>
</tr>
<tr>
<td>A → T+: low alignability + thematic relation</td>
<td>3.12</td>
<td>3.29</td>
</tr>
<tr>
<td>A → T+: Low alignability</td>
<td>1.79</td>
<td>1.84</td>
</tr>
</tbody>
</table>

each share approximately the same (low) number of features with the base. For example, the targets cow and horse each share about the same number of features with milk. Thus, if the thematically related target which poorly aligned with the base (i.e., cow) was judged more similar than the thematically unrelated target which poorly aligned was the base (i.e., horse), then we could be fairly confident that the difference in similarity was due to integration.

Four lists of 12 base–target pairs were created by randomly selecting three base-target pairs of each of the four types (A → T+, A → T+, A → T+, A → T+), subject to the constraint that only one base–target pair from each quintuple appeared in a list. The 12 base–target pairs of each list were then typed on sheets of paper, each above a seven-point 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 participant 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. Participants 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 in groups of two to four students, and each group was randomly assigned to either the Explanation (N = 64) or the No Explanation (N = 64) condition. Each of the four base–target lists was rated by 16 participants in each condition (with six to ten participants per each order of the rating booklet). The task took about 35 min to complete.

Results

Similarity ratings. Table 2 shows the average similarity ratings in the Explanation and No Explanation conditions for the four types of base-target pairs (A → T+, A → T+, A → T+, A → T+). As one would expect, similarity ratings for pairs which were highly alignable (A → T+ and A → T+; \( M = 4.74 \)) were reliably higher than for pairs which were poorly alignable (A → T+ and A → T+; \( M = 2.51 \)), \( F(1, 127) = 399.34, \text{MSE} = 642.2, p < .001 \). However, contrary to pre-set accounts of similarity, ratings for pairs

1 The stimulus pair “ship–sailor” was omitted as “ship–tailor.” Thus, all ratings and explanations for this pair were omitted from the analysis.
with preexisting thematic relations \((M = 4.03)\) were higher than for pairs without preexisting thematic relations \((M = 2.54); F(1, 127) = 66.63, \text{MSE} = 81.21, p < .001\). This finding suggests that integration also mediates similarity ratings—increasing the similarity between objects which can be thematically integrated.

The interaction between alignment and thematic relation was also highly reliable \((F(1, 127) = 75.87, \text{MSE} = 43.41, p < .001)\). That is, the effect of thematic relatedness on similarity judgments was larger when the targets poorly aligned with the base than when they readily aligned with the base \(\{(A+T^+)-(A-T^-)\} = 1.39 > \{(A+T^+)-(A+T^-)\} = .23\), although even the difference between the ratings for the \(A+T^+\) and \(A+T^-\) pairs was statistically reliable \((t(127) = 2.38, p < .02)\). These results suggest that factors compatible with integration (i.e., thematic relatedness) trade off with factors compatible with comparison (high alignment) in determining similarity ratings. They also suggest that although high alignment reduced the effect of integration on similarity ratings, it did not eliminate this effect when the stimuli had preexisting thematic relations.

As one can see by comparing the two columns of Table 2, similarity ratings for the Explanation group were nearly identical to those of the NoExplanation group in every condition. Thus, there was no reliable main effect for condition \((F < 1)\) and condition did not interact with either attribute overlap, \(F(1, 127) = 1.64, p < .21\), or thematic relatedness \((F < 1)\). Importantly, 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 reliable and nonreliable findings as in the subject ANOVA. We also compared the rating of each \(A+T^+\) pair (e.g., car-mechanic) with its corresponding \(A-T^-\) pair (e.g., car-plumber) and each \(A+T^-\) pair (e.g., car-tow truck) with its corresponding \(A+T^-\) pair (e.g., car-pickup truck). Every \(A+T^-\) pair had a higher similarity rating than its \(A-T^-\) pair. However, only half of the \(A+T^+\) pairs had higher ratings than their corresponding \(A+T^-\) pairs. Thus, unlike the subject mean ratings, the difference between the item mean ratings for the \(A+T^+\) and \(A-T^-\) pairs failed to reach statistical reliability \((t(11) = 1.35, p < .21)\).

Justifications. Because the similarity ratings were nearly identical in the Explanation and NoExplanation conditions, the justifications generated by participants can shed further light on the effects of stimulus compatibility on similarity judgments. One of the authors and a naive coder divided the justifications into instances of comparison and integration. These response categories accounted for more than 95% of the justifications. The remaining responses included blanks or statements explicitly indicating that two objects had nothing in common (e.g., "no similarity", "peanut butter–fork", "nothing in common" (ship–soldier), "can’t think of any similarities" (fly–
Comparison justifications included both commonalities and differences. Examples of commonalities were: “similar in consistency” (milk–coffee), “similar because they both are used to travel across water” (ship–canoe), and “they are both mammals” (cat–pediatrician). Examples of differences were: “coffee is usually hot while milk is generally served cold” (milk–coffee), “one is much larger (ship)” (ship–canoe), and “peanut butter is food, fork is a utensil” (peanut butter–fork).

Integration justifications included both plausible and implausible thematic relations. Examples of plausible thematic relations were: “the only similarity here would be if someone spread the peanut butter with a fork” (peanut butter–fork), “somewhat similar because a mechanic works on cars” (car–mechanic), “no similarity other than that a fly could sit on a curtain” (fly–curtain), “a chisel can be used to make a painting” (chisel–painting), and “broken cars need tow trucks” (car–tow truck). Some of the plausible thematic relations were novel, rather than preexisting, such as “a pediatrician might own a cat” (cat–pediatrician), “a waitress may use a telephone” (telephone–waitress). Examples of implausible thematic relations were: “you don’t milk a horse” (milk–horse), “an electrician doesn’t repair chairs” (chair–electrician). In some cases, participants explicitly contrasted the implausible relation with its plausible preexisting variant: “men usually wear ties—women do not usually wear ties” (tie–woman), “a chisel is used for sculpture—a painting is entirely different” (chisel–painting).

Table 3 presents the average number of comparison and integration justifications for the four types of base–target correspondences (averaged across participants). As one can see, the distribution of comparison and integration justifications is consistent with the processing modes that we predicted would be triggered by the preexisting taxonomic and thematic relationships. Partici-

<table>
<thead>
<tr>
<th>Base–target Correspondence</th>
<th>Comparison</th>
<th>Integration</th>
</tr>
</thead>
<tbody>
<tr>
<td>A+T+</td>
<td>1.88</td>
<td>0.33</td>
</tr>
<tr>
<td>A+T−</td>
<td>2.29</td>
<td>0.30</td>
</tr>
<tr>
<td>A−T+</td>
<td>54</td>
<td>0.79</td>
</tr>
<tr>
<td>A−T−</td>
<td>71</td>
<td>0.36</td>
</tr>
</tbody>
</table>
punts listed the most comparison justifications and fewest integration justifications for the A+T− pairs. The stimulus properties of these pairs are the most compatible with comparison. Participants also listed the fewest comparison justifications and most integration justifications for the A−T+ pairs. The stimulus properties of these pairs are the most compatible with thematic integration. Also, as predicted, the relative numbers of comparison justifications and integration justifications fell in between these extremes for the A+T+ and A−T− pairs. The stimuli were either compatible with both processes (i.e., A+T+ pairs) or weakly compatible with integration (i.e., A−T− pairs).

Integration justifications were not a peculiarity of a few participants or a few items. Rather, 57 of 64 or 89% of the participants in the Explanation condition constructed at least one integration justification in explaining their similarity ratings. The median number of integration justifications per participant (out of 12 responses in their rating booklets) was 4, and the average was 4.37. For 38 of the 47 object pairs, there were attempts to integrate the pairs.

The analysis of the justification distribution revealed that participants were more likely to integrate pairs of objects which have a preexisting thematic relation (M = .56) than those which do not (M = .20). This difference was highly reliable, F(1, 63) = 131.47, MSE = 8.57, p < .001. They were also more likely to integrate pairs of objects which are poorly alignable (M = .58) than those which are highly alignable (M = .18). This difference was also highly reliable, F(1, 63) = 140.56, MSE = 10.09, p < .001. In addition, the tendency to integrate pairs with prior thematic relations versus those without such relations was greater when the pairs were poorly alignable (i.e., [A − T−] = 4.30) than when they were highly alignable (i.e., [A+T−] = 3.97, MSE = .26, p < .05). These results show that the extent to which stimuli are compatible with integration systematically affects the tendency to integrate those stimuli.

Further analyses of the justification distribution reveal that participants were more likely to compare highly alignable pairs (M = 2.08) than poorly alignable pairs (M = .63). This difference was highly reliable, F(1, 63) = 209.25, MSE = 135.63, p < .001. They were also more likely to compare stimuli without prior thematic relations (M = 1.50) than stimuli with prior thematic relations (M = 1.21). This difference was also highly reliable, F(1, 63) = 21.78, MSE = 5.54, p < .001. Finally, the tendency to compare alignable pairs versus nonalignable pairs was greater when they were not thematically related (i.e., [A+T−] = 1.17) than when they were thematically related (i.e., [A+T−] = 3.49, MSE = .88, p < .001). These results show that the extent to which stimuli are compatible with comparison systematically affects the tendency to compare those stimuli.

Effects of integration on similarity ratings. We also examined more di-
really how integration influenced the similarity ratings. Specifically, we
trolled for spurious effects of differences between items on similarity ratings
by looking at the effects of integration within the same item. We selected
all the pairs for which integration justifications were mentioned (37 of the
47 pairs of items). For 28 of these 37 pairs the average similarity rating
was greater when participants constructed a plausible integration justification
than when they did not construct an integration justification or when they
constructed an implausible integration justification (e.g., ‘an electrician
doesn’t repair chairs’ (chair–electrician)). This effect was statistically reli-
able for the A→T− and A→T+ pairs but not for the A+T− pairs. (We did
not analyze the A+T− pairs as participants rarely constructed integration
justifications for these pairs—see Table 3). This analysis provides stronger
evidence that integration mediates similarity judgments.

Alternative explanations for the findings. The results are consistent with
our claim that stimulus compatibility invoked different processing modes,
which in turn influenced similarity ratings. However, because thematic relat-
edness was a within-subject factor, there are some alternative explanations
for various aspects of the results. First, participants 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 could be linked by a
thematic relation. If so, this demand characteristic should especially manifest
itself in the contrast between the A→T+ pairs (e.g., milk–cow) and A→T−
pairs (e.g., milk–horse). Because these pairs poorly align (and thus share
very few common dimensions) participants should be especially likely to
give higher ratings to the A→T− pairs, which at least share a salient, prex-
isting thematic relation. In order to notice that some items have thematic
relations but others do not, participants must test observe some pairs. Hence,
the rating of an A→T+ pair should be relatively higher than its correspond-
ing A→T− pair when it occurred in the second half of the rating form than
when it occurred in the first half. To test this possibility, we compared perfor-
mance on the first and second halves of the list. There was no statistically
reliable interaction between presentation order (first versus second half) and
thematic relation (absent versus present). In the first half, ratings for A→T−
pairs were higher than their corresponding A→T− pairs for 9 of 11 compari-
sions. In the second half, ratings for all A→T− pairs were higher than their
corresponding A→T− pairs.

Second, the presence of obvious thematic relations linking some stimulus
pairs but not others may have led participants to believe that their task was
to integrate those pairs which lacked preexisting thematic relations. This de-
mand characteristic would lead subjects to list these thematic relations in
their explanations. Three findings argue against this possibility. First, partici-
pants rarely mentioned integration justifications for A+T− pairs (an average
of .03 per participant). Second, this explanation predicts that participants
should be more likely to construct integration justifications for A→T− pairs
when these items occurred in the second half than in the first half. However, the proportion of integration justifications constructed for $A-T-$ stimuli when they occurred in the first half (3.34) was nearly identical to the proportion constructed for $A-T-$ stimuli when they occurred in the second half (3.35). (Across participants, the counterbalancing of stimuli guaranteed that the same $A-T-$ pairs occurred equally often in the first and second half.) Finally, the average position of the $A-T-$ stimuli was 3.4 in the first half of the rating form compared to 9.67 in the second half of the rating form. Thus, the tendency to construct novel integration justifications was present very early in the stimulus sequence and does not appear to be produced by prior exposure of salient, preexisting thematic relations.

Discussion

The results suggest that stimulus compatibility influenced the mode of processing that participants employed in making similarity judgments. When stimuli were taxonomically related (and hence, easily alignable), participants compared the stimuli in order to determine their matching and mismatching features. However, when the stimuli were either poorly alignable or thematically related, participants also attempted to integrate them. In this case, a processing mode which was compatible with the stimuli (i.e., integration) competed with the task-appropriate process (i.e., comparison, as assumed by similarity theorists) in affecting similarity ratings.

Our findings also show that stimulus properties trade off each other in their tendency to invoke one or the other processing mode. The stimuli that were most compatible with integration and least compatible with comparison (i.e., $A-T+$ items) were the most likely to invoke integration and the least likely to invoke comparison. Analogously, stimuli which were most compatible with comparison and least compatible with integration (i.e., $A+T-$ items) were the most likely to invoke comparison and the least likely to invoke integration. The relative tendency to invoke either comparison or integration fell in between these extremes for stimuli which were compatible with both processes (i.e., $A+T+$ pairs) or weakly compatible with integration (i.e., $A-T-$ pairs).

Our results support Bassok and Medin’s (1997) view that similarity judgments are sometimes mediated by both comparison and integration. Our results also replicate Bassok and Medin’s finding that high alignability triggers comparison and poor alignability triggers integration. The present findings extend the Bassok and Medin findings in several ways. First, our study suggests that stimulus compatibility with processing is a broader principle that applies to both alignability and thematic relatedness of the paired stimuli. Second, it shows that alignability trades off with thematic relatedness in determining the relative influence of each process. Third, it shows that the effect of integration on similarity judgments is not unique to ratings of actions but applies to ratings of objects as well. Finally, 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 integration affected their judgments as well.

EXPERIMENT 2

In Experiment 2 we examined the robustness of the compatibility effects observed in Experiment 1. Specifically, we assessed whether a task that emphasizes the appropriateness of the comparison process more explicitly than similarity ratings would eliminate the compatibility effects observed in Ex- periment 1 (i.e., the integration responses to nonalignable and/or themati- cally related stimuli). The task was listing the commonalities and differences between object pairs. It is commonly accepted that this task involves compari- son (Gentner & Markman, 1994; Markman & Gentner, 1993a, 1993b; Mark- man & Winiarski, 1997). The instructions illustrated the task with a list of common and distinctive features for the pair robins and canaries, a pair of highly alignable taxonomically related objects, and did not mention any thematic links between them. We used the same stimulus pairs as in Experi- ment 1, which orthogonally varied in terms of taxonomic and thematic relat- edness.

One could imagine scenarios in which the task instructions are either effective or ineffective in reducing the effects of stimulus compatibility on pro- cessing. On the one hand, instructions which clearly describe an example of similarity based on common and distinctive features should highlight the appropriateness and importance of the comparison process. Thus, partici- pants should primarily list features as commonalities and differences. On the other hand, because the results of Experiment 1 show that processing is strongly influenced by stimulus compatibility, we would expect thematic relat- edness and alignability to influence processing. Thus, the commonality and difference listing task should be influenced by stimulus compatibility (and hence both comparison and integration), despite instructions that em- phasize comparison.

Method

Participants. Thirty-two Northwestern University undergraduates participated as part of a course requirement.

Materials. The stimuli were the 12 quadruples of base-target pairs used in Experiment 1 (shown in Table 1). The same four lists of 12 base-target pairs were also used in this experiment as well as the two orders of the stimuli for each list. Recall that a list consisted of three base-target pairs from each of the four types: A+T-, A+T+, A−T+, A−T−. For each ordering, a form was created containing each target-quadrant by the word COMMONALITIES, 10 blank lines, the word DIFFERENCES, and 10 more blank lines. In a second version of the form the words COMMONALITIES and DIFFERENCES were re-
TABLE 4

Mean Number of Comparison and Integration Responses Listed as Commonalities and Differences for the Four Types of Base-Target Correspondence, Averaged across Participants (Experiment 2)

<table>
<thead>
<tr>
<th>Base-target</th>
<th>Total</th>
<th>Commonalities</th>
<th>Differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>A+T+</td>
<td>7.57</td>
<td>4.24</td>
<td>3.33</td>
</tr>
<tr>
<td>A+T−</td>
<td>8.33</td>
<td>4.63</td>
<td>3.70</td>
</tr>
<tr>
<td>A−T+</td>
<td>6.03</td>
<td>2.25</td>
<td>3.78</td>
</tr>
<tr>
<td>A−T−</td>
<td>6.08</td>
<td>2.22</td>
<td>3.86</td>
</tr>
</tbody>
</table>

Comparison responses

A+T+ .30   .22   .28
A+T− .04   .02   .02
A−T+ .77   .48   .29
A−T− .06   .31   .29

Integration responses

versed for each base-target pair. The procedure yielded 16 forms, which were duplicated to yield a total of 32 forms.

Procedure. Participants read instructions telling them that they would see pairs of everyday things and that they were to look at each pair and to write down those characteristics or features which they have in common as well as the features which differ between them. To illustrate, they were given the pair "robin-catary" and told that, according to the experimenter's intuitions, both robins and canaries were birds, fly, have wings, and have beaks (commonalities). In contrast, robins are generally red whereas canaries are yellow, robins eat worms, and canaries eat seeds, and canaries are often pets but robins are not (differences). Note that the task was illustrated with features describing a pair of things which do not have prior thematic associations. Further, the instructions did not mention any potential thematic links between robins and canaries.

Participants were encouraged to list as many commonalities and differences that came to mind. There were no right or wrong answers and no minimum or maximum number of features that they should list. They were also told to list the commonalities under the place labeled COMMONALITIES and the differences under the place labeled DIFFERENCES. Participants were randomly assigned to a form. They proceeded through the task at their own pace. The task took about 40 min to complete.

Results

One of the authors and a naive coder examined the participants' responses and divided them into instances of comparison and integration, using the same criteria as in Experiment 1. Overall, the author and naive coder agreed on 92% of their classifications. Differences in scoring were resolved through discussion. Table 4 presents the average number of references to each category for the four types of base-target correspondences (averaged across participants).

Despite instructions which explicitly illustrated the task with features that

The stimulus pair "car-plimbo" was mistyped as "car-beetle." Thus, all ratings and explanations for this pair were omitted from the analyses.
were commonalities and differences, participants still integrated the stimulus pairs. As in Experiment 1, stimulus compatibility strongly affected the tendency to integrate or compare items. Again, participants listed the most comparison responses and fewest integration responses for items which were the most compatible with comparison (i.e., the A+T- pairs). They also listed the fewest comparison responses and the most integration responses for items which were most compatible with integration (i.e., the A-T+ pairs, although A-T- pairs had almost the same number of comparison responses as the A-T+ pairs).

Participants were more likely to integrate pairs of objects which have a preexisting thematic relation (M = .64) than those which do not (M = .32). This difference was highly reliable, F(1, 31) = 19.82, MSE = 3.18, p < .001. They were also more likely to integrate pairs of objects which are poorly alignable (M = .68) than those which are highly alignable (M = .27).

This difference was also highly reliable, F(1, 31) = 28.92, MSE = 5.49, p < .001. These findings show that the compatibility of the stimuli with integration systematically affected responding. These factors interacted but in a manner that was opposite to that found in the first study. That is, the tendency to integrate pairs with prior thematic relations versus those without such relations was greater when the pairs were readily alignable (i.e., \(A-T+ \) - \(A+T-\) = .64) than when they were poorly alignable (i.e., \(A-T+ \) - \(A-T-\) = .17); F(1, 31) = 6.12, MSE = .66, p < .02. It is not clear why this interaction occurred. It might be partly explained by noting that the A+T- items are very incompatible with integration and the tendency to integrate these pairs was at floor. In contrast, all of the other stimulus types are at least somewhat compatible with integration.

As in Experiment 1, integration was not a peculiarity of a few participants or a few items. Rather, 31 of the 32 participants generated at least one integration response as a commonality or difference (and 28 of 32 generated three or more). The median number of integration responses per participant was 4, and the average was 6.38. Also, integration responses were listed for 37 of the 47 base-target pairs (all 10 pairs for which integration responses were not listed were A+T- pairs).

Participants listed more comparison responses for highly alignable pairs (M = 7.95) than for poorly alignable pairs (M = 6.06). This difference was highly reliable, F(1, 31) = 38.24, MSE = 114.07, p < .001. They also listed more comparison responses for pairs that lacked prior thematic relations (M = 7.23) than for pairs with prior thematic relations (M = 6.80); F(1, 31) = 4.11, MSE = 5.21, p < .05. Finally, the tendency to list comparison responses for alignable pairs versus nonalignable pairs was greater when they were not thematically related (i.e., \(A+T- \) - \(A-T-\) = 2.23) than when they were thematically related (i.e., \(A+T+ \) - \(A-T+\) = 1.54).

This interaction approached statistical reliability, F(1, 63) = 3.28, MSE = 3.64, p < .08. These results show that the extent to which stimuli are comput-
ible with comparison systematically affects the tendency to compare those stimuli.

Discussion

Overall, comparison was the dominant process in the commonality and difference listing task. Nevertheless, despite instructions which strongly suggested that only comparison was relevant to the task, participants were still affected by the compatibility of stimulus properties with processing. When the stimuli were poorly alignable or had a prior thematic association, participants tended to integrate them. Conversely, when the stimuli were highly alignable or did not have prior thematic associations, participants were more likely to compare them.

In general, the pattern of findings mirrored those of the first study. Alignability and thematic relatedness strongly affected the relative tendency to compare or integrate stimuli. Stimulus properties also traded off each other in their tendency to invoke one or the other processing mode. The stimuli that were most compatible with integration and least compatible with comparison (i.e., A−T+ items) were the most likely to invoke integration and (together with the A−T− items) the least likely to invoke comparison. Analogously, stimuli which were most compatible with comparison and least compatible with integration (i.e., A+T− items) were the most likely to invoke comparison and the least likely to invoke integration. In general, the relative tendency to invoke either comparison or integration fell in between these extremes for stimuli which were compatible with both processes (i.e., A+T+ pairs), or weakly compatible with integration (i.e., A−T− pairs)—although as noted, these pairs showed a tendency to invoke comparison that was just as low as in the A−T+ items).

EXPERIMENT 3

In Experiments 1 and 2 we found that for certain stimuli, integration affected performance on tasks thought only to involve comparison. In Experiment 3 we examined the complementary hypothesis that for certain stimuli comparison would affect performance on a task that should only involve integration. The task was rating the extent to which pairs of objects are thematically related. As in Experiment 2, we illustrated the task by using stimuli pairs that invoked the task appropriate process (e.g., a shovel is used to remove snow) and did not mention any common or distinctive features of the integrated objects. We again used the same stimuli as in the prior experiments, so that the pairs orthogonally varied in terms of alignability and thematic relatedness.

Based on the findings of Experiment 2, we predicted that stimulus compatibility would be sufficiently robust to affect people’s performance. In particular, we predicted that the taxonomically related object pairs would invoke
the stimulus-compatible comparison process, so that the commonalities and
differences between these objects would affect people's integration ratings.

Method

Participants. Thirty-two University of Chicago undergraduates participated as part of a
course requirement.

Materials. The stimuli were the same 12 quintets of base-target pairs used in the first
two experiments (shown in Table 1). The same four lists of 12 base-target pairs were also
used from these studies as well as the two orderings of the stimuli for each list. The rating
forms created from these lists were identical to those of Experiment 1 except that the 7 point
rating scale was anchored with "Thematically unrelated" under 1 (instead of "not at all simi-
lar") and "highly thematically related" under 7 (instead of "very similar").

Procedure. Participants read instructions telling them that many entities (e.g., objects, peo-
ple) are thematically related to each other, i.e., they interact with each other in a systematic
way. Several examples of thematic relations were described: a shovel is used to remove snow,
a bird lives in a nest, a potter makes vases, and a roof is on top of a house. Participants were
told that they would be presented with 12 pairs of familiar entities and that they were to judge
the extent to which the entities in each pair were thematically related, using a 7-point scale.
Participants were to circle a 1 if the paired entities were thematically unrelated and 7 if the
entities were highly thematically related. They were to use the numbers between 1 and 7 to
indicate intermediate degrees of thematic relatedness. After judging the degree of thematic
relatedness, participants were to explain their judgment and to specify in what way the entities
were thematically related. Eight participants rated each of the four base-target lists. The task
took about 15 min to complete.

Results

Thematic relatedness ratings. Table 5 shows the average thematic relat-
edness ratings for the four types of base-target pairs (A+T+, A−T−,
A−T+, A−T−). As one would expect, thematic relatedness ratings for pairs
with preexisting thematic relations (M = 5.68) were reliably higher than for pairs
without preexisting thematic relations (M = 3.32; F(1, 31) = 222.01,
MSE = 177.35, p < .001. Importantly, as predicted by the compatibility

<table>
<thead>
<tr>
<th>Base-target correspondence</th>
<th>Thematic relatedness rating</th>
<th>Similarity rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>A+T±</td>
<td>5.60</td>
<td>4.86</td>
</tr>
<tr>
<td>A−T−</td>
<td>4.21</td>
<td>4.63</td>
</tr>
<tr>
<td>A+T−</td>
<td>5.75</td>
<td>3.21</td>
</tr>
<tr>
<td>A−T±</td>
<td>2.44</td>
<td>1.81</td>
</tr>
</tbody>
</table>

Note. For comparison, the rightmost column shows corresponding average similarity ratings ob-
tained in Experiment 1.
TABLE 6
Mean Number of Comparison and Integration Justifications Listed for the Four Types of Base–Target Correspondence, Averaged across Subjects (Experiment 1)

<table>
<thead>
<tr>
<th>Base–target</th>
<th>Comparison</th>
<th>Integration</th>
</tr>
</thead>
<tbody>
<tr>
<td>A+T+</td>
<td>1.38</td>
<td>3.16</td>
</tr>
<tr>
<td>A+T−</td>
<td>4.40</td>
<td>3.99</td>
</tr>
<tr>
<td>A−T+</td>
<td>1.9</td>
<td>3.72</td>
</tr>
<tr>
<td>A−T−</td>
<td>5.0</td>
<td>2.91</td>
</tr>
</tbody>
</table>

hypothesis, relatedness ratings for highly alignable pairs (M = 4.91) were reliably higher than for poorly alignable pairs (M = 4.09); F(1, 31) = 19.85; MSE = 21.13, p < .001. This finding suggests that comparison also mediates thematic relatedness ratings—increasing the thematic relatedness between objects which can be easily compared.

Mirroring the pattern of results in Experiment 1, the interaction between alignability and thematic relatedness was also highly reliable (F[1, 31] = 31.0, MSE = 29.39, p < .001). High alignability strongly affected relatedness judgments when the pairs did not have prior thematic associations (ΔA+T− + (A−T−) = 1.77), but had basically no effect on relatedness ratings when the items had prior thematic associations (ΔA+T+ + (A−T+) = 1.15). This result suggests that factors compatible with comparison (i.e., high alignability) trade off with factors compatible with integration (i.e., prior thematic associations) in determining thematic relatedness ratings. In this case, prior thematic relations reduced the effects of comparison on thematic relatedness ratings, whereas in Experiment 1, high alignability reduces the effects of integration on similarity ratings.

Justifications: One of the authors and a naive coder classified the justifications as instances of comparison and integration. In addition to these responses, participants sometimes explicitly indicated that two objects were not linked by a thematic relation (less than 0.5% of the responses, with 90% of these responses mentioned for A+T− and A−T+ items). Also, on a few occasions (less than 0.4% of the responses) participants did not write down anything or they made a comment that was irrelevant to the rating (e.g., for the pair milk–cow, one participant remarked, “My uncle has a dairy farm”; for ship–sailor, one participant remarked, “That matters,” and so on). Irrelevant comments and failures to provide justifications were discarded from further analysis. Overall, the author and naive coder agreed on 90% of their classifications. Differences in scoring were resolved through discussion.

Table 6 presents the number of comparison and integration justifications (averaged across participants) for the four types of base–target correspondences. As in the previous studies, the distribution of comparison and inte-
gration justifications is consistent with the influence of stimulus compatibil-
ity on processing. Participants listed the most comparison and fewest inte-
gration justifications for the $A + T+$ pairs and the fewest comparison and
most integration justifications for the $A - T+$ pairs. The relative numbers of
comparison and integration justifications fell between these extremes for
the $A + T+$ and $A - T+$ pairs.

Just as integration responses were not a peculiarity of a few participants
or a few items in the previous experiments, comparison responses were not
uncommon for participants or items. Rather, all 32 participants produced at
least one comparison justification in explaining their relatedness ratings. The
median number of comparison justifications per participant (out of 12 expla-
nations in their rating booklets) was 6, and the average was 6.5. Comparison
justifications were listed for 29 of the 48 items and were listed for almost
every highly-alignable pair (31 of the 24 items), whether or not it had a prior
themetic association.

As predicted by the compatibility hypothesis, participants listed many
more comparison justifications for highly-alignable pairs ($M = 2.91$) than
for poorly alignable pairs ($M = 3.34$), $F(1, 31) = 124.37, MSE = 310.13,$
$p < .001.$ They also listed more comparison justifications for pairs that were
not thematically related ($M = 2.47$) than for pairs which were thematically
related ($M = .78$), $F(1, 31) = 84.64, MSE = 91.13, p < .001.$ Finally, the
tendency to list comparison justifications for alignable pairs versus non-
alignable pairs was greater when they were not thematically related (i.e.,
$[A + T+] - [A - T+] = 3.39$) than when they were thematically related
(i.e., $[A + T+] - [A - T+] = 1.19$). This interaction was highly reliable,
$F(1, 31) = 69.46, MSE = 60.3, p < .001.$ These results show that the extent
to which stimuli are compatible with comparison systematically affects the
tendency to compare those stimuli.

As expected if participants were following instructions to rate the thematic
relatedness of object pairs, participants listed more integration justifications
for pairs with preexisting thematic relations ($M = 3.44$) than for pairs without
preexisting thematic relations ($M = 1.75$). This difference was highly reli-
able, $F(1, 31) = 83.39, MSE = 91.13, p < .001.$ There was also a reliable
main effect for alignability. Participants listed more integration justifications
for pairs which were poorly alignable ($M = 3.31$) than for pairs that were
highly alignable ($M = 1.88$), $F(1, 31) = 20.15, MSE = 66.13, p < .001.$
These results show that the extent to which stimuli are compatible with inte-
gration systematically affects the tendency to integrate those stimuli. Finally,
as in Experiment 2, the tendency to integrate pairs with prior thematic rela-
tions versus those without such relations was greater when the pairs were
highly alignable (i.e., $[A + T+] - [A - T+] = 2.57$) than when they were
poorly alignable (i.e., $[A - T+] - [A + T+] = .81$), $F(1, 31) = 16.69, MSE$
$= 24.5, p < .001.$ It is unclear why we observed this interaction. Again, it
might be partly explained by noting that the $A + T+$ items are very incom-
patible with integration (whereas all the other stimulus types are at least partly
compatible with integration). Even with instructions that highlighted integration, the tendency to integrate A+T− pairs was very low, relative to the tendency for other stimulus types.

Discussion

Overall, integration was the dominant process in a thematic relatedness task. Nevertheless, despite instructions which suggested that this task should only involve integration, participants were strongly affected by stimulus compatibility. In particular, mirroring the results of Experiments 1 and 2, when the paired stimuli were highly alignable (i.e., taxonomic pairs) or did not have prior thematic associations, participants often compared the stimuli instead of integrating them. Conversely, when the paired stimuli were poorly alignable or had prior thematic associations, participants integrated them. Also, as in the previous experiments, stimuli which were the most compatible with comparison were the most likely to invoke comparison and the least likely to invoke integration, whereas stimuli which were the most compatible with integration were the most likely to invoke integration and the least likely to invoke comparison. Again, the relative tendency to invoke either comparison or integration fell in between these extremes for stimuli which were compatible with both processes or weakly compatible with integration. Thus, stimulus compatibility with comparison and integration affected the relative tendency to employ these processes.

The results also suggest that thematic relatedness ratings (like similarity ratings) are sometimes mediated by both comparison and integration. For example, thematic relatedness ratings were much higher for alignable items than for nonalignable items. In corresponding fashion, participants were more likely to provide comparison justifications for their ratings of alignable rather than nonalignable items.

Consistent with the compatibility hypothesis, the thematic relatedness ratings in Experiment 3 show a pattern that is the inverse of the similarity ratings in Experiment 1. Alignability strongly affected thematic relatedness ratings when the stimuli were thematically unrelated; Conversely, in Experiment 1, thematic relatedness strongly affected similarity ratings when the stimuli were difficult to align. Also, alignability did not affect relatedness ratings when the stimuli were thematically related. In corresponding fashion, thematic relatedness had a much smaller effect on similarity ratings in Experiment 1 when the stimuli were highly alignable. Like the pattern of justifications, these findings suggest that alignability and thematic relatedness trade off each other in their tendency to invoke one or the other processing mode.

GENERAL DISCUSSION

Our findings highlight the importance of compatibility between stimuli and processing in affecting task outcomes; certain properties of stimuli are more conducive to one type of processing versus another and tend to trigger
the more compatible process (see Bassok, 1997, for a more general discussion of this point). Specifically, we found that preexisting taxonomic and thematic relations between concepts systematically modulated the likelihood that these concepts will be compared or integrated. Taxonomically related concepts were more likely to be compared and less likely to be integrated than taxonomically unrelated concepts. At the same time, thematically related concepts were more likely to be integrated and less likely to be compared than thematically unrelated concepts. Although the dominant process was always task-appropriate (comparison in Experiments 1 and 2 and integration in Experiment 3), the same modulating pattern held when participants judged similarity (Experiment 1), listed commonalities and differences (Experiment 2), and judged thematic relatedness (Experiment 3).

Our results extend Bassok and Medin’s (1997) surprising finding that similarity judgments are affected by both comparison and integration, with the relative influence of each process determined by the alignability of the stimuli as well as whether the stimuli implied a thematic context in suggesting a series of different actions carried out by the same agent. We showed that Bassok and Medin’s findings are not unique to similarity judgments (they also hold for relatedness judgments), are not unique to actions (they also hold for objects), and are not unique to situations in which people explain their ratings. Importantly, our findings subsume those of Bassok and Medin’s in showing that stimulus compatibility with processing is a more general principle and one in which the effects of alignability and thematic relatedness can trade off each other.

We also found that the modulating effects of stimulus compatibility are sufficiently robust to override instructions that highlight task-appropriate processing and provide clear examples of the output of such processes (Experiments 2 and 3). Other studies have also shown that task instructions fail to override the effects of stimulus properties. Skwarek and Clark (1996) gave subjects preference tasks involving trails of objects consisting of a target and two alternatives—one taxonomically related and one thematically related to the target. In three experiments, they varied instructions in ways which highlighted either the appropriateness of the thematic choice (e.g., “goes with” or the taxonomic choice (e.g., “is similar to”)—even enhancing the critical words by enlarging, underlining, and boldfacing them. Compared to more neutral instructions (e.g., “is related to”), the instructional manipulation only modestly affected the tendency to choose the taxonomic or thematic alternative, suggesting that the particular stimuli primarily affected type of processing.

Ecological Validity of Stimulus-Compatible Processing

The modulating effect of stimulus-compatible processing on task performance ensures that task-appropriate processing is enhanced when it matches the process invoked by the stimuli (i.e., comparing taxonomically related
concepts and integrating thematically related concepts. At the same time, task-appropriate processing is diminished when it mismatches the process invoked by stimuli (i.e., integrating taxonomically related concepts and comparing thematically related concepts). This interaction is ecologically valid because it reflects and supports people's everyday experience with the target stimuli. For example, people often compare taxonomically related objects when deciding between options (Should I have beer or wine with dinner?), and they often integrate thematically related objects when achieving goals (I should call a mechanic to fix my car).

Despite their ecological validity, the task-inappropriate responses generated by our participants may reflect interference that leads to erroneous inferences. For instance, the fact that men wear ties does not make them similar to ties, and just because both soda and lemonade are beverages does not imply that they are thematically related. None, however, that application of the task-appropriate process to stimuli that evoke the mismatching process also leads to inferences with relatively low utility. For example, in our experiments comparison tasks involving thematically related stimuli led to such inferences as: men are similar to ties because “both are longer than they are wider,” a cat and a veterinarian are similar because “both probably eat some type of meat,” a telephone and receptionist are different in that one “has wires and one does not have wires,” and fork and peanut butter differ because “forks don’t stick to the roof of your mouth.” We believe that the very existence of such useless responses highlights the importance and reasonableness of restricting the application of processes to a limited set of stimuli—those that invoke matching compatible processes.

The importance of taking into account the correspondence between stimuli and processing can be understood by analogy to mathematical functions (corresponding to processes), which are always defined with respect to a particular domain and range of values (corresponding to stimuli). For example, rational numbers are closed under the operation of division (i.e., the division of any two rational numbers is itself a rational number). However, natural numbers are not closed under division—the division of some pairs of natural numbers (e.g., 3/5) does not yield a natural number. Furthermore, it is readily apparent that the meaningful application of mathematical functions to real-life situations should be selective. Thus, while it is meaningful to add the number of cars on two floors of a parking lot, it is meaningless to add the number of cars on one floor to the number of miles traveled by the cars on the second floor. In fact, consistent with our present findings, Bower, Chase, and Martin (1980) have shown that college students strongly prefer to add rather than divide object pairs from the same taxonomic category (e.g., tulips and daffodils), while they prefer to divide rather than add object pairs that are related by a thematically asymmetric semantic relation (e.g., tulips and vases). This same strong tendency is seen in the examples used by mathematics textbook writers.
**Stimulus Compatibility across Cognition**

We have shown that stimulus compatibility modulates the influence of comparison and integration in several tasks. We also believe that such effects hold in a great variety of cognitive tasks. However, in sharp contrast to the definition of mathematical functions and their everyday application, researchers who investigate the cognitive processes of comparison and integration typically fail to consider the possibility that these processes may meaningfully operate only over a restricted range of stimuli. Instead, most processing accounts of cognitive tasks assume that (1) each task is mediated by a single process and (2) that this process operates over the entire range of stimuli. Further, this existing conceptual framework overestimates the extent to which people’s performance reflects controlled, task-specific performance and neglects the impact of uncontrolled, stimulus-specific performance. Below, we review evidence which suggests that these assumptions may be incorrect. Many tasks involve both comparison and integration and these processes are modulated by stimulus compatibility.

*Category formation.* Psychological models of category formation have exclusively focused on comparison (e.g., Ahn & Medin, 1992; Anderson, 1990; Busemeyer & Myung, 1988; Estes, Campbell, Hatzopoulos, & Hurwitz, 1989; Fiszer, 1987; Gluck & Bower, 1988; Hantzman, 1986; Medin & Schaffer, 1975). Typically, models are presented with a number of examples of different categories and form representations which capture the commonalities of each category and the differences which distinguish the categories. These models do not learn the thematic relations that hold between entities or that are the central aspect of many categories (e.g., passenger, hunter, shoebox, wine critic, spectator). At present, they have not been applied to the learning of thematic categories such as events, collections, or scenes, and ad hoc categories.

Recent work suggests that stimulus compatibility affects whether category formation occurs by comparison or integration. Lin (1996) constructed triads with one alternative having a highly salient preexisting thematic relation with the target but the second alternative having a less salient taxonomic relation with the target (i.e., the pair was related at the superordinate level). Across a number of studies, subjects selected a majority of the thematically related alternatives as “going best to form a category” with the target, even though the instructions defined a category in terms of shared common features. These results are also consistent with our findings that task instructions have modest effects on overriding stimulus compatibility. Also, to the extent that similarity functions as an explanatory construct for category formation (Goldstone, 1994; Medin, Goldstone, & Gentner, 1993), some of our findings indirectly suggest that stimulus compatibility affects category formation. In particular, recall that similarity judgments were affected by integration and
comparison. The likelihood of their influence was modulated by stimulus compatibility.

Induction. An important issue in induction is understanding how people generalize a feature of one category to another category. Models of induction assume that it is mediated by feature comparison which determines the similarity between categories (Heit & Rubinstein, 1994; Osherson, Sturt, Wilkie, Lopez, & Shrift, 1990; Rips, 1975; Sloman, 1993). For example, people are more likely to extend a novel property to the category birds given knowledge that robins have the property than that penguins have the property. However, recent work suggests that stimulus compatibility modulates the tendency to use comparison versus integration in induction. Lix (1996; Experiment 8; see also Ross & Murphy, 1999) varied the taxonomic and thematic relatedness between triads of categories (e.g., plumbers, carpenters, pipes) in an induction task. Told that one category (e.g., plumber) had a new kind of bacteria, participants overwheimgly believed that the thematically related category (e.g., pipes) rather than the taxonomically related category (e.g., carpenter) was more likely to have the bacteria. Presumably, participants believe that bacteria are transmitted by physical contact, a property that typically characterizes thematically related entities. As further evidence of the role of thematic relations, Lopez, Atran, Coley, Medin, & Smith (1997) found cross-cultural differences between Michigan undergraduates and the Itzaj Maya in reasoning about folkbiological categories. Because of their greater knowledge about how plants and animals interact, the Itzaj Maya based their inductions on thematic relations between entities. In contrast, Michigan undergraduates who lacked such knowledge based their inductions on similarity.

Metaphors, similes, and analogies. A common view of how metaphors, similes, and analogies are understood is by comparing one domain to another (Falkenhainer, Forbus, & Gentner, 1989; Holyoak & Thagard, 1989). At present, it appears that none of these language constructions are understood by integrating the domains (Winston, 1997). For example, people interpret “that vet is an elephant” as a large or clumsy vet (comparison) and not as a vet that treats elephants (integration). Nevertheless, Love and Wissniewski (1998) provided evidence suggesting that integration is involved in the comprehension of some of these constructions. Specifically, they varied whether the topic and vehicle of metaphors and similes could be plausibly linked by a salient thematic relation and found that interpretations took longer when the terms could be thematically integrated. For example, “that river is a turtle” took longer to understand than “that assembly line is a turtle,” even though the same property of turtle (slowness) is involved in both interpretations. This finding suggests that metaphors and similes are interpreted by comparison. However, stimulus properties also trigger integration with the output of this process subsequently suppressed. On the other
hand, as we suggest below, it is plausible that thematic integration may affect the actual interpretation.

Conceptual combination. People frequently combine familiar concepts to produce new ones. For example, the term ‘‘ostrich ranch’’ was recently coined to refer to a ranch that raises ostriches. Psychological accounts of conceptual combination typically assume that people combine concepts by integrating them, as in the example above (Gagne & Shoben, 1997; Cohen & Murphy, 1984; Murphy, 1988). In addition, linguistic analyses emphasize the importance of integration (Downing, 1977; Levi, 1978). These accounts have not addressed the role of stimulus compatibility in processing. Interestingly, of all the cognitive tasks that we have examined, this task is the only one modeled by single-process models of integration. Yet, stimulus compatibility modulates integration and comparison in conceptual combination. Highly alignable combinations are overwhelmingly interpreted by comparison (Wisniewski, 1996), even when their constituents can be plausibly linked by thematic relations (Wisniewski & Love, 1998). The latter result parallels our finding that participants were less likely to integrate stimuli which were thematically related when they were also highly alignable.

The Need for Two-Process Accounts

Our findings and those we have reviewed strongly suggest that researchers need to extend their models to include both integration and comparison processes and the effects of stimulus compatibility. It seems clear that both processes play significant roles in a number of domains such as category formation, conceptual combination, and induction. But even in domains in which one process is considered primary or more appropriate, a two-process account may be necessary because of possible interactions between comparison and integration. Thematically integrating objects often affects the salience of their properties or results in new properties. Thus, integration could influence the properties determined by comparison. For instance, in judging the similarity of ladder box to ladder, a person may interpret a ladder box as a box that contains a ladder (thus thematically integrating ladder and box). As a consequence of integration, a person may infer that a ladder box is shaped like a ladder. Then, in comparing how similar a ladder box is to a ladder, a person might note the common shape between them. Or, as suggested by Bassok and Medin (1997), upon constructing a thematic link between entities people may compare the entities and note that they are similar because they share an abstract commonality (namely, ‘‘participation in the same thematic relation’’). Again, this property arises as a consequence of integration. It seems that a comparison-only model of similarity would not account for these possibilities. Similar reasoning can be applied to the interpretations of metaphors and similes. For example, a person may understand ‘‘boy, this drink is an ice cube’’ as a very cold drink partly because drinks are very cold as a consequence of drinks being thematically linked to ice
cubes. Again, a comparison-only model of metaphor would not account for this possibility.

Of course, effects may occur in the other direction, with comparison influencing thematic integration. For example, interpreting ladder box as "a box that contains a ladder" may involve comparing the shape of a typical box to that of a ladder in order to determine if the ladder could fit into the box. It seems that an integration-only model of conceptual combination would not account for this possibility. Exactly how integration and comparison influence each other is a challenging question and an issue for future research. The present work takes an initial step toward addressing this issue in showing that stimulus properties trade off each other in determining the relative effects of each process.

Taken together, our findings and those we have reviewed suggest that stimulus compatibility is an important factor that modulates the effects of comparison and integration across a wide range of cognitive tasks. However, with some notable exceptions, researchers have ignored stimulus compatibility, which in turn may have led to the dominance of single-process models. Bassok (1997) suggested that this neglect is a case of stimulus-compatibility effect—that researchers generally tend to select stimuli that are compatible with the type of process they are modeling. Thus, these accounts are not necessarily wrong but rather, incomplete. For example, a primary theoretical construct of the category-based induction model (Oehseren et al., 1990) is that people access categories common to the premise and conclusion subcategories of an inductive argument. This stance necessitates the use of taxonomically related stimuli. Thus, these researchers used stimuli with common categories (subcategories of animals), which indeed tend to be fairly alignable. In turn, researchers who have evaluated the category-based induction model used the same kinds of stimuli (Heit & Rubenstein, 1994; Shueman, 1993).

Similarly, early linguistic work on conceptual combination was conducted within the context of generative grammar theories and thus focused on abstract thematic relations between nouns (Levi, 1978). This linguistic work, in turn, heavily influenced several prominent models of conceptual combination (Bose of Gagne & Shoben, 1997, and Coolen, van Jaarsveld, & Schuender, 1991). In evaluating their models, researchers have only examined noun-noun combinations whose constituents are thematically related.

SUMMARY

We have suggested that different properties of stimuli are compatible with different types of processing. Specifically, object pairs from the same taxonomic category tend to be alignable and compatible with comparison, whereas object pairs that play different roles in thematic relations tend to be nonalignable and compatible with integration. We demonstrated how stimulus compatibility modulates processing and thereby affects the out-
comes of tasks that are currently explained by a single process, either com-
parison or integration. Our findings, and others that we have reviewed, sug-
gest that (1) many cognitive tasks involve both comparison and integration,
and (2) the relative influence of each process is modulated by an interplay
between the task-appropriate and the stimulus-compatible process. The exact
nature of this interplay awaits further research. We believe that single-process
models should be extended to take this interplay into account.

REFERENCES

Ahn, W. K., & Medin, D. L. (1994). A two-stage model of category construc-
Memory and Cognition, 10, 82–93.
810–842.
category learning: A comparison of parallel network and memory storage-retrieval mod-
571.
Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989) The structure mapping engine: Algo-
rithms and examples. Artificial Intelligence, 45, 1–63.
Learning, 2, 139–172.
of modifiers and adverbialities. Journal of Experimental Psychology: Learning, Memory,
Science, 5, 132–158.


Received June 8, 1999.
Author Index for Volume 39

A
Aguiar, Andrita, 116

B
Baillargeon, Renée, 116
Hasson, Miriam, 208

H
Hernández-Vasquez, Linda, 3
Homa, Donald, 69
Houston, Derek M., 159

J
Jusczyk, Peter W., 159

K
Katsnelson, Alla S., 3

L
Logan, Gordon D., 1

N
Newsome, Mary, 159

S
Spelke, Elizabeth S., 3

T
Tsa, Peter Ulric, 37

W
Wisniewski, Edward J., 208

Z
Zaki, Safa R., 69