Memory & Cognition
1999, 27 (5), 844-855

Justification effects on the judgment of analogy

WINSTON R. SIECK
University of Michigan, Ann Arbor, Michigan

CLARK N. QUINN
University of New South Wales, Kensington, New South Wales, Australia

and

JONATHAN W. SCHOOLER
University of Pittsburgh, Pittsburgh, Pennsylvania

Many of us share a strong intuition that justification forces us to better understand the situations we face. And there is substantial evidence indicating that this is often the case. However, there is a growing body of research showing that, under certain circumstances, explanation and justification can impair performance on a variety of cognitive tasks. In the present research, the effects of justification on judgment of the soundness of analogies were examined. Subjects judged the quality of the match between pairs of stories with varying degrees of superficial and analogical similarity. Experimental subjects either provided reasons for their judgments or wrote recollections of the target stimuli. These subjects rated the match between stimulus pairs as more sound than did control subjects. Also, providing reasons led to poorer discrimination between superficially similar aspects of the stimuli and analogous aspects. Explanations of these findings are proposed, and implications for problem solving and confidence judgment are discussed.

We are often called upon to explain or justify our proposed actions (Koehler, 1991), and many of us believe that this process aids our understanding of the situations we face. This belief appears to be well founded. Studies of learning and instruction indicate that students learn more when explaining examples to themselves (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; VanLehn & Jones, 1993), and justifying one's reasoning has been shown to facilitate problem-solving performance, at least under some conditions (Ahlum-Heath & Di Vesta, 1986; Gagné & Smith, 1962; McGeorge & Burton, 1989; Stinessen, 1985). There is also evidence that a request to provide a rationale or to write a justification for a prospective choice can aid decision making (Miller & Fagley, 1991; Sieck & Yates, 1997). A possible explanation of these findings is that the act of translating thoughts into words increases the degree to which people search for and process information relevant to their situations. This extra processing increases activation of the relevant aspects of the situations in people's mental representations. Performance is facilitated because the relevant aspects of the situations are more heavily weighted in determining subsequent actions.

However, there is also a large body of research illustrating that explaining and analyzing reasons can also be detrimental to performance on cognitive tasks (Levine, Halberstadt, & Goldstone, 1996; Schooler & Engstler-Schooler, 1990; Schooler, Ohlsson, & Brooks, 1993; Wilson, Dunn, Kraft, & Lisle, 1989; Wilson, Hodges, & LaFleur, 1995; Wilson & LaFleur, 1995; Wilson & Schooler, 1991). For example, Schooler and Engstler-Schooler discovered that verbalization of nonverbal memories impairs performance in recognition tasks. In their study, subjects who viewed a picture of a face and then described it performed worse on a recognition test than subjects who did not describe it. This effect was postulated to occur because faces are extremely difficult to describe. Schooler and Engstler-Schooler concluded that the new verbal representation *overshadowed* the original memory and, hence, referred to the phenomenon as the *verbal overshadowing* effect. Such verbal overshadowing effects have been found to generalize to a wide variety of difficult-to-describe memories, including colors (Schooler & Engstler-Schooler, 1990), visual forms (Brandimonte, Schooler, & Gabbino, 1997), maps (Fiore, 1994), tastes (Melcher & Schooler, 1996), and music (Houser, Fiore, & Schooler, 1997).

Further studies have suggested that verbal overshadowing may influence performance on other cognitive

This research was supported by an NIMH grant to the third author. These experiments were reported in W.R.S.'s undergraduate honors thesis submitted to the Department of Psychology, University of Pittsburgh. We greatly appreciate the many useful comments on earlier versions of this research provided by Ed Smith, Colleen Seifert, and Miriam Bassok, as well as by the members of the Michigan Judgment and Decision Lab: Frank Yates, Beth Veinott, Paul Price, Andrea Patalano, Joseph Magee, and Paul Estin. We also thank Tom Ward, Bobbie Spellman, and two anonymous referees for their helpful remarks on prior drafts of this paper. Correspondence concerning this article should be addressed to W. R. Sieck, Department of Psychology, University of Michigan, 525 East University Avenue, Ann Arbor, MI 48109-1109 (e-mail: sieck@umich.edu).

tasks in addition to memory tasks. Schooler et al. (1993) found that verbalization can interfere with the solving of insight problems, and Wilson and Schooler (1991) have shown that analyzing reasons can impair the quality of affective decisions. More recently, Wilson and LaFleur (1995) found that analyzing reasons for acting (or not acting) in a certain way led to both a decrease in predictive accuracy and an increase in confidence that the predicted behavior would occur. These studies suggest that justification may lead to an increased reliance on irrelevant aspects of situations when the relevant information is inaccessible or difficult to verbalize.

In conclusion, we propose that justification induces people to search for and manipulate information in the situations they face. The verbalized elements become more active in memory via rehearsal and, hence, are more influential on subsequent actions. Thus, if the relevant aspects of a situation are accessible and readily verbalized, justification should be beneficial. However, if it is very difficult to express the relevant aspects verbally or if they are inaccessible, less relevant aspects are substituted in the justification, and performance is hindered.

Justification and Analogical Match Evaluation

Analogy is a powerful tool that is used in a wide array of cognitive tasks, including learning, reasoning, problem solving, and argument (see, e.g., Bassok, 1990; Gentner & Holyoak, 1997; Marchant, Robinson, Anderson, & Schadewald, 1993; Spellman & Holyoak, 1992; Vosniadou & Ortony, 1989). Analogical learning can be partitioned into the subprocesses of (1) accessing the previously learned (base) system, (2) mapping a target system onto the base, (3) evaluating the match, (4) storing target inferences, and (5) extracting the common principle (Gentner, 1989). In this study, we explore the effects of justification on the subprocess of match evaluation. Evaluations of match quality have been found to be quite good (Gentner & Landers, 1985; Gentner, Rattermann, & Forbus, 1993; Rattermann & Gentner, 1987). However, justification may disrupt this process, especially if important aspects of similarity are difficult to articulate. We discuss two potential effects below.

Increased Match Judgments

One possible effect of justification would be to increase the perceived quality of match. The reasoning is that people may have natural inclinations to *seek out* ways in which situations match, rather than to look for differences. Verbalized aspects become more activated in people's internal representations, which increases match judgments.

The idea that people look for the ways in which situations match is derived from the notion that they generally use a positive test strategy in considering evidence for hypothesis testing (Klayman & Ha, 1987). In analogical learning, the natural focal hypothesis is that the base situation one has been reminded of will provide useful inferences in the novel, target situation (i.e. that the situations match in important ways). Specific evidence supporting this position also exists in the similarity literature. For example, Sjöberg (1972) wrote that "rating similarity may be construed sometimes as a problem solving task. There may be a cognitive set towards finding the similarity or, in other words, a tendency to justify as high a rating of similarity as possible" (p. 20). Markman and Gentner (1993a) instructed subjects to list either commonalities or differences for pairs of words that varied in similarity and then regressed these on similarity ratings. They found that the ratings increased with the number of listed commonalities and decreased with the number of differences, but that much more weight was given to commonalities than to differences. Medin, Goldstone, and Gentner (1993) asked subjects to type in the similarities and differences between an ambiguous base stimulus and one of two possible target stimuli. They found that the subjects interpreted the ambiguous stimulus in a way that increased its similarity to the target. These two studies suggest that commonalities are more important than are distinctions to rated similarity, even when the task explicitly requests that differences be brought to mind. Medin, Goldstone, and Markman (1995) provided other arguments that people have a tendency to search for similarity. Taken together, these studies provide substantial evidence that people attend more to positive correspondences than to differences when judging similarity. Although we are concerned with evaluations of the match between potentially analogical situations, there is evidence that the process of judging similarity is much like that of drawing an analogy (Gentner & Markman, 1995, 1997), so the same effects should apply to the present situation.

It seems quite plausible that people will naturally focus on commonalities rather than on distinctions in judging a match and, hence, that requests to justify these judgments will induce subjects to verbalize existing similarities, but not differences. We have hypothesized that verbalized aspects receive increased activation via rehearsal, so that justification may lead to increased ratings of match between situations.

Decreased Aspect Importance Discrimination

In addition to increased judgments of the quality of match, justification may also lead to a more subtle and, perhaps, more interesting effect. In particular, we suspect that justification may reduce people's ability to discriminate between good and bad analogies. However, we first consider arguments consistent with the more intuitive prediction that justification should improve discrimination.

Gagné and Smith (1962) explained their finding that people performed better on the Tower of Hanoi problem when asked to provide reasons for each move, by suggesting that justification induced people to think harder about the problem. And this probably best captures the intuition regarding justification that most of us share.

There is some evidence that thinking more about similarity tasks encourages people to respond to more important aspects of the stimuli. Markman and Gentner (1993b) asked subjects to examine pairs of scenes and, for each pair, to select an object in one scene that went with an object in the other. The scenes were designed so that one possible target object had characteristics that were very similar to those of the original, whereas a different object played a role within the scene similar to that of the original. They found that subjects who rated the similarity of the scenes prior to choosing were much more likely to pick the object that played the similar role. This provides some evidence that forcing people to think more about potential analogues induces them to attend to more analogically relevant aspects of those situations. Also, Goldstone and Medin (1994) found, in their studies of the time course of comparison, that global correspondences were more influential than local consistencies when subjects had more time to make comparisons. Since justification requires people to spend more time thinking about comparisons to be made, it could influence them to focus more on the global commonalities on which good analogies depend.

The intuition that justification should improve analogical discrimination is quite compelling. However, it is also possible that people will shift weight from the important shared aspects of stimuli to unimportant shared aspects as a result of justifying their judgments. We briefly introduce Gentner's (1983) structure-mapping theory to clarify what we mean by *important* and *unimportant* commonalities and to provide theoretical support for our claim.

Within structure-mapping theory (Gentner, 1983), situations or domains are characterized by systems composed of individual objects and their attributes, first-order relations (FORs) between these objects, and higher order relations (HORs) between FORs (or other HORs). Object attributes and relationships are represented as syntactically distinguishable predicates. Attributes are defined as predicates that take only one argument, such as "x is blue" or BLUE (x). Relations are predicates taking at least two arguments, such as "x attacks y" or ATTACKS (x, y). The nature of the arguments further distinguishes FORs (arguments are objects) from HORs (arguments are other relations). For example, if ATTACKS (x, y) and FLEES (y, x) are each first-order predicates, CAUSE[ATTACKS (x, y), FLEES (y, x)] is a second-order predicate. Notice that the number of elements in HORs increases in an essentially multiplicative fashion, and so the system can quickly become quite complex. For instance, the FOR ATTACKS (x, y) is composed of only three elements (two objects and the predicate), whereas the second-order relation (2OR) is composed of seven (two arguments times the number of elements per argument, plus the CAUSE [x, y] predicate). This complexity may make commonalities among HORs especially difficult to describe.

Within this framework, an analogy is defined as a mapping from a *base* domain into a *target* domain that conveys that the domains share a system of relations, independent of any object attribute correspondences. The *systematicity* principle is a crucial element of the mapping process. This principle states that people tend to map systems of predicates containing HORs that constrain the common lower order relations, rather than mapping individual predicates. Gentner and her colleagues (Gentner & Landers, 1985; Gentner, Rattermann, & Forbus, 1993; Rattermann & Gentner, 1987) found support for this principle in the fact that people intuitively gave more weight to analogical than to object attribute or first-order relational commonalities in their judgments of match. However, as was mentioned above, the HORs can be quite complex, and so the mapping between two sets of these relations should be much more difficult to discuss than should mappings between individual predicates. The prior research by Schooler and his colleagues suggests that people's verbalizations may not adequately represent the higher order relational commonalities existing between two situations, owing to the complexity of the stimuli. The simpler FORs and object attributes would be verbalized and, hence, made more active in the people's internal representations. The relative increase in activation of the superficial over the analogical commonalities would be reflected in judgments that discriminated less between these types of shared aspect. In fact, there is considerable evidence that deep-structural similarities are more difficult to describe than superficial commonalities.

In the first of their experiments on the role of analogy in problem solving, Gick and Holyoak (1980) collected think-aloud protocols from their subjects. In their discussion of these protocols, Gick and Holyoak (1980) noted that some of their subjects explicitly claimed that "talking aloud hindered their thinking" (p. 325). They further mention several examples of subjects' failing to verbalize specific correspondences between the base story and the target problem used. In later studies of analogical transfer, Gick and Holyoak (1983) provided people with two base stories, some of which shared analogical relations. The subjects were asked to provide written descriptions of how the two stories were similar. They found that the majority of the subjects (60%) produced poor descriptions, in that they did not include any part of the analogical relationship. The few people that did mention the analogy were much more likely to transfer their knowledge to a similarly structured problem. These results give credence to the notion that verbalization of analogical relations is difficult. Further support comes from Quinn (1990), who measured performance on various analogical processes, including what he called *confirmation of analogy*. Although the idea of confirmation corresponds directly with Gentner's idea of soundness of match (Gentner & Landers, 1985; Rattermann & Gentner, 1987), Quinn used a different technique for mea-

surement. In addition to judging the quality of the match, the subjects in Quinn's study justified their judgments and were evaluated on these explanations. The subjects were found to provide inadequate justifications, where adequacy meant writing about a common deep structure of the two domains. Quinn explains these results by saying that, "justification may be a more complex and difficult task [than judgment]" (p. 20).

In sum, deep-structural commonalities may be much more difficult to describe than superficial similarities, so that the shared surface aspects are more readily verbalized during justification. Since verbalized aspects of internal representations become more active in those representations, it seems plausible that justification will lead to a decrease in the weight given to analogical commonalities and to a corresponding increase in the weight given to superficial similarities for judgments of match. Or, said differently, justification may lead to a decrease in judgment discrimination between important and unimportant shared aspects of the stimuli. This, of course, sharply contrasts with the more intuitive notion that justification should improve discrimination. In two experiments, we tested the hypotheses that justification leads to increased judged match and to decreased aspect importance discrimination.

EXPERIMENT 1

To test these predictions, we asked subjects to read pairs of stories and to evaluate the match between each pair. The subjects engaged in one of three possible tasks, after reading a story pair but prior to evaluating the match. The primary tasks requested subjects either to justify their prospective match evaluations or to work on a filler activity during the same length of time. According to our hypotheses, the subjects engaging in the justification task should exhibit a main effect of higher match ratings and interaction effects such that less weight is given to analogical commonalities and more is given to superficial similarities, relative to the filler task. In addition to the justification and filler tasks, we also included a recall task, wherein subjects were asked to write down their recollection of the first (base) story. This task provides more direct evidence regarding the specific mechanism proposed for higher match ratings. We suggested that this effect would result from rehearsal-driven increases in activation of the commonalities between the story pairs. The subjects performing the recall task were also engaging in rehearsal that activates elements of the first story and common elements from the second story, via spreading activation. In this task, each pair of stories is presented together on the same page, and the subjects know that they will need to rate the match between story pairs. Under these conditions, analogical and superficial commonalities should be connected in memory (Seifert, McKoon, Abelson, & Ratcliff, 1986), and so, both types should receive increased activation from the rehearsal. Thus, we should find increased ratings in the recall condition, as well as in the justification condition. The recall task should also help to dissociate the mechanisms underlying our first and second predictions. Our second prediction, regarding a reduction in aspect importance discrimination, is rooted in the notion that analogical commonalities are difficult to describe and, thus, are not rehearsed. Since no corresponding difficulties exist in recalling a single story, the subjects in the recall task should weight analogical and superficial relations much as the control subjects do.

Method

Subjects. The subjects were 72 undergraduate students enrolled in an introductory psychology class at the University of Pittsburgh.

Materials. The materials were a subset of those used by Gentner et al. (1993) and Rattermann and Gentner (1987). They consisted of 12 story sets with five stories per set. Each set contained an original story plus four matching target stories, which differed in the level of similarity they shared with the original. An example of a story set is provided in Table 1.

Table 1
Example of Story Set

Base Story

Karla, an old hawk, lived at the top of a tall oak tree. One afternoon, she saw a hunter on the ground with a bow and some crude arrows that had no feathers. The hunter took aim and shot at the hawk but missed. Karla knew the hunter wanted her feathers so she glided down to the hunter and offered to give him a few. The hunter was so grateful that he pledged never to shoot at a hawk again. He went off and shot deer instead.

Literal Similarity

Once there was an eagle named Zerdia who nested on a rocky cliff. One day she saw a sportsman coming with a crossbow and some bolts that had no feathers. The sportsman attacked but the bolts missed. Zerdia realized that the sportsman wanted her tailfeathers so she flew down and donated a few of her tailfeathers to the sportsman. The sportsman was pleased. He promised never to attack eagles again.

Analogy

Once there was a small country called Zerdia that learned to make the world's smartest computer. One day Zerdia was attacked by its warlike neighbor, Gagrach. But the missiles were badly aimed and the attack failed. The Zerdian government realized that Gagrach wanted Zerdian computers so it offered to sell some of its computers to the country. The government of Gagrach was very pleased. It promised never to attack Zerdia again.

Surface Similarity

Once there was an eagle named Zerdia who donated a few of her tailfeathers to a sportsman so he would promise never to attack eagles. One day Zerdia was nesting high on a rocky cliff when she saw the sportsman coming with a crossbow. Zerdia flew down to meet the man, but he attacked and felled her with a single bolt. As she fluttered to the ground Zerdia realized that the bolt had her own tailfeathers on it.

First-Order Relational Commonalities Only

Once there was a small country called Zerdia that learned to make the world's smartest computer. Zerdia sold one of its supercomputers to its neighbor, Gagrach, so Gagrach promised never to attack Zerdia. But one day Zerdia was overwhelmed by a surprise attack from Gagrach. As it capitulated the crippled government of Zerdia realized that the attacker's missiles had been guided by Zerdian supercomputers.

From "The Roles of Similarity in Transfer: Separating Retrievability From Inferential Soundness," by D. Gentner, M. J. Ratterman, and K. D. Forbus, 1993, *Cognitive Psychology*, **25**, p. 533. Copyright 1993 by Academic Press. Reprinted with permisssion.

As is shown in Table 1, all the targets shared FORs (e.g., *x* attacks *y*) with the original. They varied by the other levels of shared similarity. In *literal similarity* (LS) targets, common higher order relational structure (i.e., the plot structure) and object matches (i.e., characters, physical objects, and locations) were added to the FOR matches. In *analogy* targets, common higher order relational structure was added, but objects differed. In *surface similarity* (SS) targets, object matches were added to the FOR matches, but the higher order relational structure differed. In *FOR-only match* targets, only the FORs matched. Table 2 illustrates the shared relations between the base stories and each of the targets.

All the subjects received the same original 12 stories. Each subject received only 1 matching target story for each original. The 12 pairs were divided into four groups in order to counterbalance the assignment of targets to originals. See Gentner et al. (1993) for further details on the story sets.

Experimental booklets were used to run the experiment. The first page of the booklet contained the instructions. The rest of the booklet was divided into 12 sets. Each set consisted of three pages. The first page of a set had a base story, which was always labeled *Story A*, and one of its targets, labeled *Story B*. The second page differed by condition. The page for the justification task had the instructions, "Please write down the reasons upon which you base your evaluation of the match between the stories," written at the top, leaving the rest of the page blank for writing. The recall task page had these instructions: "Please try to recall the first story (story A). Write down your recollection to the best of your ability." The page for the filler task had a crossword puzzle. The third page in each set instructed the subject to rate the match between the two stories on a 5-point scale (1, *extremely spurious*; 5, *extremely sound*).

Design. The experiment was a 3 (verbalization task) × 4 (similarity type) mixed design. Verbalization task was varied between subjects. The justification task required the subjects to write reasons for their prospective judgments of the match between pairs of stories. The recall task required the subjects to write their recollections of the base story for each pair of stories. The filler task required the subjects to solve crossword puzzles. Type of similarity was varied within subjects, with each subject receiving three LSs, three analogies, three SSs, and three FOR matches. The dependent measure was the subjects' judgments of the soundness of the matches between the story pairs.

Procedure. The subjects were given individual experimental booklets that contained all the relevant materials. The general instructions on the cover page of the booklet explained that the experiment was about judging the match between pairs of stories. It also informed the subjects that they should read through each page in order, without looking back, and stop at the end of each page until told to continue. This was done so that the experimenter could allot 2-min reading periods for the first page of each set and 3-min writing periods for the second page of each set. After the subjects finished reading the directions, the experimenter instructed them to turn the page and begin reading the first pair of stories. The subjects were comfortable doing this task from the start, so there were no practice trials. When all the subjects were finished, they were given written feedback explaining the experiment and then dismissed.

Table 2
Type of Story by Aspect Commonality Type

	Aspects Shared by Base and Target Stories		
	Superficial Similarity		Analogical Similarity
Similarity Type	Objects	First-Order Relations	Higher Order Relations
Literal	yes	yes	yes
Analogy	no	yes	yes
Surface	yes	yes	no
First-order relation	no	yes	no

Coding. The written responses of the subjects in the justification condition were coded for additional analysis. To accomplish this, the first author initially read the complete set of justifications and divided each into units corresponding to distinct comparisons being made by the subjects. These units were then coded on three dimensions: comparison type, level of detail, and polarity. There were five possible values for the comparison type:

1. *Object comparison.* The unit described correspondences between people, places, or objects in the two stories (e.g., *Both stories were about birds*).
2. *First-order relation comparison.* The unit compared a single action or proposition of the stories (e.g., *A person attacked a bird in each story*).
3. *Second-order relation comparison.* A 2OR involves a linkage (usually causal) between two FORs, and this code was used when comparisons of 2ORs were made (e.g., *In both stories, a person attacked an animal, causing that animal to fall down*).
4. *Higher order relation comparison.* Comparisons between relations of third or higher order were given this code (e.g., *In the first story, a person attacked an animal. In return, that animal appeased the person so that the person promised not to attack the animal again. And in the second story, one country attacked another. The second country appeased the first so that the first promised not to attack the second again*).
5. *Other.* The subject discussed something that was unrelated to comparing the stories. This code was rarely used and was treated as a nonutterance in subsequent analyses.

As can be seen in the FOR comparison example, at least some of the FOR comparisons may have also involved implicit object comparisons, and similarly for the higher level comparisons. This was handled by assigning the highest possible level of comparison to each unit. This coding scheme underestimates lower level comparisons but, hence, provides conservative tests of our hypotheses. Also, we did not originally anticipate that the subjects would make many 2OR comparisons, and they do not fit into our framework as clearly as do either superficial or analogical comparisons. This is because they are not technically FOR comparisons, but they are insufficient for determining whether the stories are truly analogous (i.e., as can be seen in Table 1, SS and FOR stories also share a number of 2ORs). Nevertheless, they did appear in the data sufficiently often to warrant their own category, so we report them in the hope that they may inform future theory but do not use them in tests of our specific hypotheses here.

The second dimension, level of detail, referred to whether the comparison was fully described (as in the examples above) or whether there was only a vague description. Two examples of vague comparisons are (1) the themes in the 2 stories were the same, and (2) the characters in the 2 stories were the same. The third dimension, polarity, pertained to whether the comparison described a similarity or a dissimilarity between the 2 stories. The first author coded the full set of justifications, and a second rater, naive as to the study hypotheses, then scored a random 17% subsample (48 stories) of the data. The raters agreed on 70%, 94%, and 97% of their codings for each dimension, respectively.

Results and Discussion

Match judgments and aspect weights. The subjects' match judgments from the first experiment are summarized in Table 3 by task and target story type. These data were analyzed with a split-plot analysis of variance (ANOVA), with task and aspect commonality type as the predictors. One-tailed p values are provided, since specific predictions are being tested. The associated model

Table 3
Mean Match Ratings (With Standard Errors), Experiment 1

Similarity Type	Verbalization Task Condition: Filler (n = 24)		Recall (n = 24)		Justification (n = 24)		Overall	
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Literal	4.2	0.12	4.4	0.10	4.6	0.08	4.4	0.06
Analogy	3.5	0.14	4.0	0.12	3.9	0.13	3.8	0.08
Surface	3.0	0.11	3.5	0.13	3.6	0.14	3.3	0.08
First-order relation	2.7	0.13	3.0	0.14	3.2	0.13	3.0	0.08
Overall	3.4	0.07	3.7	0.07	3.8	0.07	3.6	0.04

Note—Match ratings ranged from 1 to 5, with 1 = *extremely spurious* and 5 = *extremely sound.*

weights given to each type of shared aspect for each task group are presented in Table 4. The intercept terms in Table 4 are the model estimates of the average judgments given on FOR items. Note that these do not correspond exactly to the FOR means presented in Table 3, because the model estimates depend on the values in the other conditions. The b_1 and b_2 weights are the adjustments made to that baseline value when object and/or HOR commonalities are present (see Table 2), respectively.

As can be seen in Tables 3 and 4, there was a main effect for average match judgments, so that subjects in the justification and recall conditions tended to give higher ratings than did subjects in the filler condition [$F(1,45) = 10.73, p = .001$, and $F(1,45) = 4.57, p = .019$, respectively]. Although the differences look quite small in those tables, the standard deviations are also small. We thus divided the differences by the standard deviation, to get a better sense of the actual size of the effects (Cohen, 1988). The resulting effect size estimates were .85 and .64, which are large and moderate effects for the justification and recall conditions, respectively (cf. Cohen, 1988).

Also, the average judgments for the justification condition were slightly larger than those for the recall condition, although this was not statistically significant. As is shown in Table 4 and consistent with our second hypothesis, the weight given to object commonalities appears to be larger for the justification condition than for the other conditions, and the weight given to HOR commonalities appears to be smaller for the justification condition. However, the verbalization task × aspect commonality type interactions were not statistically significant.

Table 4
Linear Model Coefficients, Experiment 1

Verbalization Task	Weight Given to Each Type of Commonality: b_0 (Intercept)	b_1 (OB)	b_2 (HOR)
Filler	2.64	0.46	0.97
Recall	3.00	0.43	0.94
Justification	3.10	0.52	0.88

Note—OB, object match; HOR, higher order relation match.

Justification content. Analyses were also performed on the codes derived from the statements written by subjects in the justification condition. These analyses were collapsed across type of target story. As was predicted, the subjects made far more similarity ($M = 45.7$) than difference comparisons [$M = 7.1; t(23) = 11.96, p \approx 0$]. The first two columns in Table 5 present the mean number of times each comparison type was used and the percentage of times each comparison was only vaguely specified. As is shown, the majority of subjects' comparisons were between pairs of FORs, whereas fewer comparisons were made between HORs [$t(23) = 8.22$, $p \approx 0$]. And even when HOR comparisons were made, only 24% were specified in any detail. Although it appears that few object comparisons were made, this may well be a by-product of the coding procedures, as was described above.

Relation between content and judgment. A linear model related the number of times each content category was used for a story pair to the corresponding match judgments. A within-subjects ANOVA tested the statistical significance of the model coefficients. The coefficients and test results are presented in the third and fourth columns of Table 5. As can be seen, only the number of object-difference comparisons was not statistically related to match judgment. As was expected, judgments generally increased with the number of similarity and decreased with the number of difference comparisons, as is indicated by the signs on the coefficients in Table 5. Also, 2OR and HOR comparisons are weighted more heavily than either FOR or object comparisons. Unexpectedly, the magnitude of the weight given to differences appears to be larger than that given to commonalities, suggesting that the relative uncommonness of the comparisons being made was responsible for the match judgment effect, rather than lack of impact once they were considered.

Our hypotheses suggest two specific individual-difference relations between justification content and match judgments. The first is that people who cite a higher percentage of similarity comparisons should have higher average match judgments. Supporting this notion, there was a sizable positive correlation ($r = .40, p = .03$) between mean judgment and proportion of similarity

Table 5
Summary Statistics of Justification Content, Experiment 1

Type of Comparison	Number Used/ Story	Proportion of Vague Comparisons	Similarity Comparison Coefficients	Difference Comparison Coefficients
Intercept			3.29†	
Object	0.42	.10	0.21*	0.08
First-order relation	2.64	.04	0.11†	−0.26†
Second-order relation	0.83	.02	0.39†	−0.69†
Higher order relation	0.51	.76	0.29†	−0.59†

*$p < .025$. †$p < .005$.

comparisons. The second is that people who are more inclined to discuss higher order relational commonalities should better discriminate analogical relations from superficial similarities than do those who predominately discuss lower order relational and object commonalities. In order to test this, we developed a subject level measure of *analogical discrimination* from subjects' match judgments. Hence, prior to discussing the pertinent result, we digress to explain how the measure was created and provide estimates of it for each condition.

The analogical discrimination measure was created by first regressing each subject's match soundness (MS) judgments on a pair of dummy variables indicating whether object (OB) and/or higher order relational similarities existed for each pair of stories. That is, the following model was fitted for each subject:

$$MS = \beta_1 \cdot \text{HOR} + \beta_2 \cdot \text{OB}. \quad (1)$$

The analogical discrimination (AD) measure was then formed by subtracting each individual's beta weight for object commonalities from the beta weight for higher order relational commonalities (i.e., $\text{AD} = \beta_1 - \beta_2$). Positive values of AD indicate that analogical similarity is given more weight than superficial similarity, and negative values indicate the opposite. The mean values of AD for the filler, recall, and justification conditions in order are .22, .25, and .16, which parallels the findings presented in Table 4. It is also convenient to compute effect sizes for the reduction in discrimination, using these values. They are .14 and .21, which are small reductions from the filler and recall conditions to the justification group, respectively.

AD was developed to determine whether justification subjects who were more inclined to discuss higher order relational commonalities were better at discriminating analogical relations from superficial similarities than were those who predominately discussed lower order relational and object commonalities. However, the proportion of higher order relational commonalities mentioned by individuals in the justification group was not well correlated with AD ($r = .09, p = .34$). This may be due to many inappropriate claims of HOR similarity. That is, 42% of the HOR similarity claims made were for SS and FOR story pairs.

Summary. We found clear evidence in this first experiment that justification increases match judgments. In addition to finding higher baseline match judgments in the justification than in the filler conditions, we also found that justifiers wrote down far more commonalities than differences and that, within the justification group, the proportion of commonalities written down was positively correlated with average match judgments. We explain these findings by suggesting that justification encouraged people to search for ways in which the story pairs matched, to the neglect of ways in which they differed. The rehearsal involved in actually writing down the shared aspects increased the activation of those commonalities in memory, thereby increasing perceived MS. Support for the involvement of rehearsal comes from the fact that merely recalling the base story also increased baseline match ratings, in comparison with the filler condition.

We also found some evidence that justification led to diminished aspect importance discrimination, although these findings were less clear. In these data, we did observe that people who justified their prospective match evaluations gave slightly less weight to HOR commonalities and somewhat more weight to object commonalities than did people in the other two conditions. However, the effects were not statistically significant. Examination of the actual justifications revealed that people made far more superficial than analogical comparisons. Also, those analogical comparisons that were made were typically only vaguely expressed.

EXPERIMENT 2

The main purpose of Experiment 2 was to attempt, with a larger sample size, to replicate the effects of analyzing reasons on judgment. This should aid us in determining whether writing reasons reliably reduces the ability to discriminate analogical from superficial similarities. Also, there was some concern that the filler task added error to the match judgments, since the subjects may have forgotten the evaluations at which they had arrived, owing to interference from working the crossword puzzles. Hence, a no-delay condition was added, to determine whether such effects existed. Two predictions

Table 6
Mean Match Ratings (With Standard Errors), Experiment 2

Similarity Type	Verbalization Task Condition: No Delay (n = 32)		Filler (n = 32)		Recall (n = 32)		Justification (n = 32)		Overall	
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Literal	4.2	0.13	4.3	0.13	4.4	0.12	4.4	0.12	4.3	0.07
Analogy	3.8	0.14	3.9	0.15	4.0	0.13	3.9	0.13	3.9	0.06
Surface	3.0	0.13	2.8	0.13	3.2	0.12	3.5	0.12	3.1	0.07
First-order relation	2.5	0.13	2.7	0.13	2.6	0.14	2.9	0.13	2.7	0.06
Overall	3.4	0.08	3.4	0.08	3.5	0.08	3.7	0.07	3.5	0.04

Note—Match ratings ranged from 1 to 5, with 1 = *extremely spurious* and 5 = *extremely sound*.

were made. One was that replication of the effect of the justification and recall groups' giving higher ratings of MS would occur. Second, it was expected that a decrease in discrimination between analogical and superficial similarities would occur in the justification group alone.

Method

Subjects. There were 128 subjects, 13 of which responded to posted advertisements and were paid $5 for their participation. The rest were undergraduate students enrolled in an introductory psychology class at the University of Pittsburgh. Preliminary analyses indicated no apparent differences between the two groups.

Materials. The experimental booklets in Experiment 2 were organized in nearly the same way as those in Experiment 1. One exception was that the story-processing instructions came at the end of each set for a new, no-delay condition. Also, only 8 story sets were used, rather than 12. Finally, a separate page was added to each set with a yes/no question about whether the lessons of the first (base) story were applicable to the second (target) story. These ratings proved to be noninformative and so will not be discussed further.

Design. The design was essentially the same as that in Experiment 1, with the exception that a fourth story-processing task was added. This was a no-delay control group, which involved no intervening manipulation. Thus, there was a 4 (verbalization task; between) × 4 (similarity type; within) design. The dependent variable was the judgment of match between pairs of stories.

Procedure and Coding. The procedure and coding for Experiment 2 were identical to those in Experiment 1. The first author coded the full set of justifications, and a naive rater then scored a random 16% subsample (40 stories) of the data. The raters agreed on 84%, 98%, and 99% of their codings for the comparison type, level of detail, and polarity dimensions, respectively.

Results and Discussion

Match judgments/aspect weights. The subjects' match judgments from the second experiment are summarized in Table 6 by task and target story type. As in the first experiment, these data were analyzed with a split-plot ANOVA, with task and aspect commonality type as the predictors. The associated model weights given to each aspect for each task group are presented in Table 7. As can be seen in Table 6, the standard errors in the filler condition did not appear to be any higher than those in the no-delay group. Also, the judgments for these two conditions were not statistically or conceptually different from one another and so are treated as coming from a single control condition in subsequent analyses. As can be seen in Tables 6 and 7, there was a main effect for match judgments, so that the subjects in the justification and recall conditions tended to give larger judgments than did subjects in the control condition [$F(1,93) = 7.19$, $p = .004$, and $F(1,93) = 2.48$, $p = .06$, respectively]. The corresponding effect sizes were .60 and .28, which are medium and small effects. The average judgments for the recall condition were again slightly, but not statistically, smaller than those for the justification condition. As is shown in Table 7 and consistent with our second hypothesis, less weight is given to HOR commonalities in the justification condition than in the recall and control conditions [$F(1,45) = 2.68$, $p = .05$, and $F(1,93) = 3.55$, $p = .03$, respectively]. Also, the weight given to object commonalities again appears to be slightly larger for the justification condition than for the other conditions, but the differences are not statistically significant.

Justification content. As was expected, the subjects made far more similarity ($M = 27.0$) than difference comparisons [$M = 3.5$; $t(31) = 12.23$, $p \approx 0$]. The first two columns in Table 8 present the mean number of times each comparison type was used and the percentage of times each comparison was only vaguely specified. As in Experiment 1, the majority of the subjects' comparisons were between pairs of FORs, and far fewer comparisons were made between HORs [$t(31) = 8.05$, $p \approx 0$]. Again, relatively few (19%) of the HOR comparisons were specified in any detail.

Relation between content and judgment. As in the first experiment, a linear model related the number of times each content category was used for a story pair to the corresponding match judgments. A within-subjects

Table 7
Linear Model Coefficients, Experiment 2

Verbalization Task	Weight Given to Each Type of Commonality: b_0 (Intercept)	b_1 (OB)	b_2 (HOR)
No delay	2.51	0.45	1.25
Filler	2.62	0.23	1.35
Recall	2.65	0.51	1.27
Justification	2.90	0.54	1.02

Note—OB, object match; HOR, higher order relation match.

Table 8
Summary Statistics of Justification Content, Experiment 2

Type of Comparison	Number Used/ Story	Proportion of Vague Comparisons	Similarity Comparison Coefficients	Difference Comparison Coefficients
Intercept			2.61†	
Object	0.35	.06	0.18†	0.54*
First-order relation	2.30	.02	0.23†	−0.09
Second-order relation	0.71	.01	0.48†	−0.15
Higher order relation	0.46	.81	0.71†	−0.52†

*$p < .025$. †$p < .005$.

ANOVA tested the statistical significance of the model coefficients. The coefficients and test results are presented in the third and fourth columns of Table 8. In this experiment, neither the number of FOR nor that of 2OR differences were statistically related to match judgment. However, judgments tended to increase with the number of similarity comparisons and to decrease with the number of difference comparisons. Also, 2OR and HOR similarity comparisons were weighted more heavily than either FOR or object similarity comparisons. Contrary to the results of the first experiment, but consistent with past research (e.g., Markman & Gentner, 1993a), the magnitude of the weight given to differences was smaller than that given to commonalities. We interpret the lack of replication of the impact of differences but the successful replication of the importance of commonalities as evidence that, even when listed, commonalities are indeed more influential than differences.

As in Experiment 1, mean judgments were positively correlated ($r = .28, p = .06$) with proportion of similarity comparisons. The same AD measure as that described in Experiment 1 was computed. The mean values of AD for the control, recall, and justification conditions were .40, .35, and .23, respectively (reduction in discrimination effect sizes for the justification condition, as compared with the control and recall conditions, were .37 and .11, which are small). Also, there was some evidence that the proportion of higher order relational commonalities mentioned by individuals in the justification group was positively correlated with AD ($r = .21, p = .12$). As in the first experiment, this correlation was likely dampened by inappropriate claims of HOR similarity (33% of the HOR similarity claims made were for SS and FOR story pairs).

Summary. In this experiment, we replicated the findings that justification and recall increase match judgments over control tasks, that justifiers wrote down far more commonalities than differences, and that the proportion of commonalities written was positively correlated with average match judgments. We also found much stronger evidence that justification led to diminished aspect importance discrimination. The effect of justification's leading to a reduced weighting of HOR commonalities, as compared with the other conditions, was replicated, and the reduction achieved statistical significance in this experiment. The justification subjects again appeared to give more weight to object matches than did the control subjects, although this finding was not statistically significant. Examination of the actual justifications revealed that the subjects made far more superficial than analogical comparisons, and the analogical comparisons that were made were typically only vaguely described.

GENERAL DISCUSSION

The present experiments explored two potential influences of justification in the context of analogy: (1) increases in judgments of match soundness and (2) decreases in discrimination between analogical and superficial similarity. We discuss the relevant findings and implications of these effects in turn.

Increased Match Judgments

We have suggested two distinct influences from justification that might lead to increases in judgments of analogical MS. First, people look for positive matches between situations. Thus, when justifying, people tend to explain ways in which situations correspond, neglecting those aspects that differ. Second, activation is increased for explained features via rehearsal, so the explained, similar aspects are weighted more heavily than the unexplained, dissimilar aspects in judgments of match. Thus, overall judged MS is increased.

This proposal received fairly strong support from our two experiments. Justification led to increased judgments of match in both experiments, and computed effect sizes were medium to large. Content analyses of the justifications in both experiments revealed that the subjects made far more similarity than difference comparisons and that the proportion of similarity comparisons made was positively correlated with match judgment, both results being in accord with our first assumption. However, that assumption might also imply that similarity comparisons in the justifications would be weighted more heavily than differences, and the results on that point were mixed. In particular, that was found to be the case in the second experiment, but not in the first. Finally, our second premise, that rehearsal is a primary driver of the effects, was supported by the finding that the subjects who were asked merely to recall the base story also provided higher ratings of match than did control

subjects (the effect sizes were small to medium). This was anticipated by our proposal, since recall subjects were also engaging in rehearsal, which activated elements in the first story and common elements in the second story.

These ideas can be extended to another line of research that suggests that explanation influences likelihood judgments in a similar manner. A substantial number of studies support the claim that explanation of a possibility increases the confidence in or the likelihood of the truth of that possibility (Koehler, 1991; Koonce, 1992; Phillips, Koonce, & Hopkins, 1994; Sieck & Yates, 1997). In his review of earlier studies, Koehler suggests that confidence in a hypothesis is increased because people approach the task by assuming that the hypothesis is true and assessing how plausibly it accounts for the available evidence. He also notes that basing confidence on plausibility is an example of likelihood judgment by representativeness (Kahneman & Tversky, 1972). That is, Koehler suggests that confidence estimates are determined by the degree to which ideal evidence obtained under the condition that the hypothesis is true is similar to the actual evidence at hand. Thus, our present findings give some suggestions about why confidence is increased in these situations. Our findings suggest that, when writing reasons for a prospective confidence judgment, people will mention how the actual evidence is similar to ideal evidence generated under the assumption that the hypothesis is true, neglecting dissimilarities between ideal and actual evidence. Aspects included in the justifications become more active, and, hence, people who justify or explain their confidence judgments will implicitly judge the similarity between actual and ideal evidence to be greater than will those who do not explain themselves, which leads to explicit increases in the judged likelihood that the hypothesis is true.

Decreased Aspect Importance Discrimination

The effect of justification's reducing discrimination between analogical and surface similarities is predicted by the verbal overshadowing principle, as applied to match evaluations. That is, since deep-structural commonalities are difficult to describe, people will essentially ignore these in their explanations and will focus instead on describing superficial commonalities. The more easily described, superficial similarities are thus made more active and, hence, given more weight, relative to analogical relations, in the resulting judgments.

These ideas were fairly well supported by the experiments. In both experiments, justification reduced discrimination between analogical and surface similarities, although the effects were not statistically significant in the first study. In order to clarify the existence of this important result, we combined the data on our AD measure for the two experiments, using a Bayesian version of meta-analysis (see Lee, 1989). The resulting posterior probability of reduced AD for the justification group, as compared with the other groups, was .969, implying that the AD population mean is most likely smaller for people that justify their judgments. However, the effects were admittedly small. Also supporting our proposal, content analyses of the justifications revealed that the subjects' comparisons focused much more on superficial than on analogical aspects of the stories in both experiments. Finally, in both experiments, there were slightly positive correlations between the proportion of higher order commonalities mentioned in the justifications and AD, although those correlations were not statistically significant in either experiment. In short, although the subjects in these studies do not exhibit overwhelming reductions in AD from justifying their judgments, they are clearly not aided by that demand, as might well have been expected on the basis of more intuitive arguments.

In addition to the justification effects found here, there is also evidence that verbalization can impede the retrieval of relevant analogies. Lane et al. (1998) asked subjects to read a set of short stories. Later, these subjects were told to retrieve an analogous story from the set for each of several test stories. There was one analogous and one superficially related story in the original set for each of the test stories. Some subjects completed the test while talking aloud, and others did not. Lane et al. found that verbalization subjects were less likely to retrieve the analogical stories, relative to the control subjects, hence corroborating the present evidence that verbalization can be disruptive to analogy.

Justification Improvement

Part of the present research has been somewhat negative, in that we discussed and found some detrimental effects of justification on judgment, as people naturally approach the task. Although we feel this is important to acknowledge, it should not be taken as an indication that our outlook on explanation and justification processes is necessarily bleak. On the contrary, it is our view that deepening our understanding of these processes can only help us to improve them.

One possible approach to improving justification might be to increase the precision with which reasons are stated. One study suggesting that this might be helpful was done by Stein et al. (1982). This study examined the effects of the precision of self-generated elaborations on retention in fifth graders. It suggested that more precise elaborations were associated with greater retention and that training could help students become more precise in their elaborations. In the analogy literature, Mandler and Orlich (1993) found that giving precise, abstract descriptions of the analogical relations of a base problem situation at the time of encoding greatly increases the chances that the solution given in the base will be used to solve a later, target problem. These researchers also suggested that this level of description of the base could be trained for.

Although the research above dealt with memory and transfer, and judgment is being discussed here, it might not be overly presumptuous to suggest that preciseness

of justification may alleviate the detrimental effects suggested by this study. It would certainly be worthwhile to examine this possibility, especially if training could improve the justification process.

REFERENCES

AHLUM-HEATH, M. E., & DI VESTA, F. J. (1986). The effect of conscious controlled verbalization of a cognitive strategy on transfer in problem solving. *Memory & Cognition*, **14**, 281-285.

BASSOK, M. (1990). Transfer of domain-specific problem-solving procedures. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 522-533.

BRANDIMONTE, M. A., SCHOOLER, J. W., & GABBINO, P. (1997). Attenuating verbal overshadowing through visual retrieval cues. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 915-931.

CHI, M. T. H., BASSOK, M., LEWIS, M. W., REIMANN, P., & GLASER, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, **13**, 145-182.

COHEN, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

FIORE, S. M. (1994). *Verbal overshadowing of macro-spatial memory*. Unpublished masters thesis, University of Pittsburgh, Pittsburgh, PA.

GAGNÉ, R. M., & SMITH, E. C. (1962). A study of the effects of verbalization on problem solving. *Journal of Experimental Psychology*, **63**, 12-18.

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

GENTNER, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 199-241). Cambridge: Cambridge University Press.

GENTNER, D., & HOLYOAK, K. (1997). Reasoning and learning by analogy: Introduction. *American Psychologist*, **52**, 32-34.

GENTNER, D., & LANDERS, R. (1985). Analogical reminding: A good match is hard to find. In R. Ferrell (Ed.), *Proceedings of the International Conference on Cybernetics and Society*. (pp. 607-613). Tucson/ New York: IEEE.

GENTNER, D., & MARKMAN, A. B. (1995). Similarity is like analogy. In C. Cacciari (Ed.), *Similarity* (pp. 111-148). Brussels: BREPOLS.

GENTNER, D., & MARKMAN, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist*, **52**, 45-56.

GENTNER, D., RATTERMANN, M. J., & FORBUS, K. D. (1993). The roles of similarity in transfer: Separating retrievability from inferential soundness. *Cognitive Psychology*, **25**, 524-575.

GICK, M. L., & HOLYOAK, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, **12**, 306-355.

GICK, M. L., & HOLYOAK, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, **15**, 1-38.

GOLDSTONE, R. L., & MEDIN, D. L. (1994). The time course of comparison. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 29-50.

HOUSER, T., FIORE, S. M., & SCHOOLER, J. W. (1997). *Verbal overshadowing of music memory: What happens when you describe that tune?* Unpublished manuscript.

KAHNEMAN, D., & TVERSKY, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, **3**, 430-454.

KLAYMAN, J., & HA, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, **94**, 211-228.

KOEHLER, D. J. (1991). Explanation, imagination, and confidence in judgment. *Psychological Bulletin*, **110**, 499-519.

KOONCE, L. (1992). Explanation and counterexplanation during audit analytical review. *Accounting Review*, **67**, 59-76.

LANE, S., SCHOOLER, J., VILLA, D., HAYS, D., TRACHTMAN, D., DAVIS, C., & SHEREMETA, D. (1998, May). *Retrieval shifts in the verbal overshadowing of analogies*. Paper presented at the Annual Meeting of the Western Psychological Association, Albuquerque, NM.

LEE, P. M. (1989). *Bayesian statistics: An introduction*. New York: Halsted.

LEVINE, G. M., HALBERSTADT, J. B., & GOLDSTONE, R. L. (1996). Reasoning and the weighting of attributes in attitude judgments. *Journal of Personality & Social Psychology*, **70**, 230-240.

MANDLER, J. M., & ORLICH, F. (1993). Analogical transfer: The roles of schema abstraction and awareness. *Bulletin of the Psychonomic Society*, **31**, 485-487.

MARCHANT, G., ROBINSON, J., ANDERSON, U., & SCHADEWALD, M. (1993). The use of analogy in legal argument: Problem similarity, precedent, and expertise. *Organizational Behavior & Human Decision Processes*, **55**, 95-119.

MARKMAN, A. B., & GENTNER, D. (1993a). Splitting the differences: A structural alignment view of similarity. *Journal of Memory & Language*, **32**, 517-535.

MARKMAN, A. B., & GENTNER, D. (1993b). Structural alignment during similarity comparisons. *Cognitive Psychology*, **25**, 431-467.

MCGEORGE, P., & BURTON, A. M. (1989). The effects of concurrent verbalization on performance in a dynamic systems task. *British Journal of Psychology*, **80**, 455-465.

MEDIN, D. L., GOLDSTONE, R. L., & GENTNER, D. (1993). Respects for similarity. *Psychological Review*, **100**, 254-278.

MEDIN, D. L., GOLDSTONE, R. L., & MARKMAN, A. B. (1995). Comparison and choice: Relations between similarity processes and decision processes. *Psychonomic Bulletin & Review*, **2**, 1-19.

MELCHER, J., & SCHOOLER, J. W. (1996). The misremembrance of wines past: Verbal and perceptual expertise differentially mediate verbal overshadowing of taste. *Journal of Memory & Language*, **35**, 231-245.

MILLER, P. M., & FAGLEY, N. S. (1991). The effects of framing, problem variations, and providing rationale on choice. *Personality & Social Psychology Bulletin*, **17**, 517-522.

PHILLIPS, F., KOONCE, L., & HOPKINS, P. (1994). Development of explanations and repeated judgments in an audit. *Perceptual & Motor Skills*, **79**, 75-81.

QUINN, C. N. (1990, July). Analogical process performance. In M. Piattelli-Palmarini (Ed.), *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society* (pp. 16-21). Hillsdale, NJ: Erlbaum.

RATTERMANN, M. J., & GENTNER, D. (1987). Analogy and similarity: Determinants of accessibility and inferential soundness. In E. E. Smith (Ed.), *Proceedings of the Ninth Annual Conference of the Cognitive Science Society* (pp. 23-35). Hillsdale, NJ: Erlbaum.

SCHOOLER, J. W., & ENGSTLER-SCHOOLER, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology*, **22**, 36-71.

SCHOOLER, J. W., OHLSSON, S., & BROOKS, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology: General*, **122**, 166-183.

SEIFERT, C. M., MCKOON, G., ABELSON, R. P., & RATCLIFF, R. (1986). Memory connections between thematically similar episodes. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 220-231.

SIECK, W., & YATES, J. F. (1997). Exposition effects on decision making: Choice and confidence in choice. *Organizational Behavior & Human Decision Processes*, **70**, 207-219.

SJÖBERG, L. (1972). A cognitive theory of similarity. *Göteberg Psychological Reports*, **2**, 1-23.

SPELLMAN, B. A., & HOLYOAK, K. J. (1992). If Saddam is Hitler then who is George Bush? Analogical mapping between systems of social roles. *Journal of Personality & Social Psychology*, **62**, 913-933.

STEIN, B. S., BRANSFORD, J. D., FRANKS, J. J., OWINGS, R. A., VYE, N. J., & MCGRAW, W. (1982). Differences in the precision of self-generated elaborations. *Journal of Experimental Psychology: General*, **111**, 399-405.

STINESSEN, L. (1985). The influence of verbalization on problem solving. *Scandinavian Journal of Psychology*, **26**, 342-347.

VANLEHN, K., & JONES, R. M. (1993). Learning by explaining examples to oneself: A computational model. In S. Chipman & A. L. Megrowitz (Eds.), *Foundations of knowledge acquisition: Cognitive models of complex learning* (pp. 25-82). Boston: Kluwer.

VOSNIADOU, S. & ORTONY, A. (Eds.) (1989). *Similarity and analogical reasoning*. Cambridge: Cambridge University Press.

WILSON, T. D., DUNN, D., KRAFT, D., & LISLE, D. J. (1989). Introspection, attitude change, and attitude behavior consistency: The disruptive effects of explaining why we feel the way we do. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 19, pp. 123-205). New York: Academic Press.

WILSON, T. D., HODGES, S. D., & LAFLEUR, S. J. (1995). Effects of introspecting about reasons: Inferring attitudes from accessible thoughts. *Journal of Personality & Social Psychology*, **69**, 16-28.

WILSON, T. D., & LAFLEUR, S. J. (1995). Knowing what you'll do: Effects of analyzing reasons on self-prediction. *Journal of Personality & Social Psychology*, **68**, 21-35.

WILSON, T. D., & SCHOOLER, J. W. (1991). Thinking too much: Introspection can reduce the quality of preferences and decisions. *Journal of Personality & Social Psychology*, **60**, 181-192.

(Manuscript received February 20, 1997;
revision accepted for publication November 2, 1998.)