Goal-Directed Processes in Similarity Judgement

Hiroaki Suzuki
Human Communication Research Centre
University of Edinburgh
2 Buccluech Place
Edinburgh EH8 9LW Scotland
hsuzuki@cogsci.ed.ac.uk

Hitoshi Ohnishi
Department of Systems Science
Tokyo Institute of Technology
4259 Nagatsuda-Cho Midori-Ku
Yokohama, Japan
ohnishi@sys.titech.ac.jp

Kazuo Shigemasu
Department of Education
Tokyo Institute of Technology
Ohokayama Meguro-Ku
Tokyo, Japan
kshigema@nc.titech.ac.jp

Abstract

This study explored the effects c: a goal and subject's knowledge in similarity judgements. We hypothesized that the process of computing similarity consist of two phases: the processes of explanation and feature comparison. When a goal is salient and the knowledge required to achieve it is available, people compute similarity by explaining the goal in terms of a given state by using domain knowledge. Thus, in this case, rated similarity should be a function of the distance between the goal and the state. When the explanation fails, the judgements should instead to be based on the feature comparison. Expert, novice, and naive subjects were asked to solve the Tower of Hanoi puzzle. The subjects were required to judge the similarity between the goal and various states of the puzzle. The results showed that their judgements differed, depending on their expertise. While experts' ratings were best characterized by the number of operators necessary to transform a given state to the goal, those of naive subjects were completely based on the number of shared features. The second experiment revealed that the experts' judgements of similarity are not be due to learned contiguity through practice.

Introduction

People's flexible and intelligent behavior is enabled by their ability to use appropriate and relevant past experiences. Since knowledge required to deal with a situation is likely to have application in similar situations, similarity is often a good measure of the relevance and the appropriateness of such experiences. People retrieve the knowledge of a past experience which is similar to the present situation, and use specific facts or procedures stored in the knowledge to understand, explain, and learn from, the situation.

In this sense, judgements of similarity are ubiquitous in every kind of cognition. Actually, categorization has been considered as processes in which people compute similarities between an instance and its prototype (Smith, 1990). Problem solving may also be characterized as a process by which similar experiences are retrieved and modified (Hammond, 1990). Transfer, one of the central issues in the study of learning, can be thought of as a matter of which types of similarities learners should, or are apt to, attend to — deep ones (school of formal discipline) or shallow ones (Thorndykian). Whichever similarities learners may attend to, it is central to the issue of transfer.

If judgements of similarity are involved in these various kinds of cognition, such judgements cannot be stable. In fact, similarity changes dynamically across situations. For example, studies on novice-expert differences have shown that experts' judgements of similarity of one problem to another are fundamentally different from novices' (Chi, Feltovich, & Glaser, 1981). Smith (1990) suggested that young children's preference for global similarity is gradually replaced by dimensional similarity. Moreover, people's judgements of similarity are highly context sensitive. Tversky and Gati (1982) showed that rated similarity between two countries varied depending on the type of stimulus sets in which they were included. In addition, Goldstone, Medin, & Gentner (1991) revealed that people are sensitive to relational similarities when relations are more salient, and are sensitive to attributional similarities when attributes are salient. In Pylyshyn's terms, similarity judgements are *cognitively penetrable* (Pylyshyn, 1984).

Although the studies described above consistently show that the selection of feature changes across situations, current theories of similarity do not take it into account. For example, the "contrast model" (Tversky, 1977) does not provide any constraints on which features are selected and which features should be considered salient. This problem becomes more serious when one notices that the "frame problem" (McCarthy & Hayes, 1969) also matters in similarity judgement. Since there are potentially infinite number of features, it is impossible to collect all of the features of an object by asking people to list them. In addition, people create idiosyncratic features so that a man who gets drunk sometimes "jumps into a swimming pool with all his clothes on" (Murphy & Medin, 1985).

The second problem is that most of the experiments in the previous studies were conducted in rather "neutral" settings where "disturbing" factors such as subjects' prior knowledge, or goals spontaneously generated by subjects, were carefully removed. Although this approach might be suitable for the investigation of object-level similarity, the results cannot easily be extended to problem-solving and learning. This is because goals and knowledge play critical roles in these activities (Glaser, 1984; Resnick, 1990).

Therefore, a model of similarity should be developed that can deal with the above-mentioned problems. In problem-solving and learning, people have to focus on important features: What determines the importance of each feature? Features contributing to the achievement of the goal are those which should be judged to be important. However, it is not sufficient simply to have the goal, because it is quite often the case that features satisfying the goal are not readily accessible. Thus, the next question is: What determines the degree to which a feature contributes to the goal. It is knowledge of the domain which determines the degree of the contribution to the goal. If features satisfy the triggering condition of knowledge which contributes to achieving the goal, these features should be considered to be contributing to the goal.

The above analysis leads to the idea that the process of computing similarity is *explanation*, because people *explain* the goal, using given features and the domain knowledge (Mitchell, Keller, & Kedar-Cabelli, 1986). For example, suppose that you are looking for an ashtray, and that there is a paper cup, a juice can which has not been opened yet, and a cookie can with some cookies in. The goal, knowledge, and given features of each object in this case are as follows:

Goal:

ashtray(X)	←	heat-proof(X)
	∧	open-concavity(X)
	∧	unimportant-in(X)

Knowledge

heat-proof(X)	←	can(X)
unimportant-in(X)	←	empty(X)
empty(X)	←	mv-content(X,¬ X)
open-concavity	←	can(X)
heat-proof(X)	←	can(X)
heat-proof(X)	←	in(water, X)
⋮	⋮	⋮

Features

can(juice-can)
can(cookie-can)
white(paper-cup)
delicious(cookie)
⋮

The three conditions of the goal are satisfied by transforming the features of the cookie-can, using the domain knowledge of the shape and material. As a result, recognition of similarity between the cookie can and an ashtray is obtained. It is important to note that irrelevant features such as the color and taste of objects are not picked up for the computation of similarity in this process.

It is often the case that there is a difference of complexity between one explanation and another. For example, one can make the paper cup an ashtray by pouring juice into the cup. However, the required explanation is more complex in this case than in the previous case. Generally, an item whose features require more transformation is judged to be less similar. This would be a source of a degree of similarity. Complexity of the explanation may partly be affected by what kinds of knowledge people have. If one has well-organized and readily accessible knowledge, the derived explanation is likely to be much simpler. We may therefore expect to find a difference between experts and novices in similarity judgement.

What happens if the explanation fails? There are two cases when the explanation fails: a case where no explicit goal is concerned, and that where required knowledge is not accessible. In this case, the only available information is features of items to be compared with. As a result, judgement of similarity is carried out on the basis of features.

Thus, we hypothesize that processes of computing similarity consist of two subprocesses. In the first phase, people judge the degree of similarity by explanation. A degree of similarity in this phase is defined as the number of operators that is required to explain one item in terms of the goal. When the goal is not obvious, or relevant operators are not easily accessible, the second subprocess follows. In this phase, people's judgements are based on features shared with or specific to items, as is modeled by the contrast model.

There are several advantages to this model. First, as is obvious, the model can be applied to problem-solving and learning where goal and knowledge play dominant roles. Second, the model can provide an account of feature selection. Since, in theory, one can create an infinite number of attributes which characterize given items, it is crucial for models of similarity to select relevant features. In this model, these are resolved by relevance of features to the goal. Thus, f in the contrast model is no longer defined adhoc in this model.

In order to explore the effects of a goal and knowledge on similarity judgement, we conducted a series of experiments using the Tower of Hanoi puzzle. Subjects were asked to rate similarities of a given state to the goal where all disks were placed at the right peg. The reason why we chose this puzzle is that it is easy to specify the goal, the knowledge (operators), as well as features.

According to our hypothesis, whether subjects know the rule of the puzzle determines whether judgements are based on explanation or feature comparison. If subjects know the rules of the puzzle, judged similarity should be a function of the distance, that is, the number of operators required to transform a given state to the goal. On the other hand, if subjects do not know the rule, the judgements should be carried on the basis of the number of shared features. In addition, the model assumes that the accessibility of the operators also affects similarity judgements. It is likely that novices' judgements are based on the number of shared features, since those who have just been taught the rules would find it more difficult to access an appropriate operator than experts. Therefore, it is predicted that while experts' judgements should be best characterized by the number of operators, novices' ones should be based on both the number of operators and on the number of shared feature.

Experiment 1

Method

Subjects Subjects were 21 Tokyo Institute of Technology graduate and undergraduate students. They were randomly assigned to one of the three conditions: expert, novice and control. None of the subjects in the novice or control conditions had any prior experience with the Tower of Hanoi puzzle.

Procedure Subjects in the expert condition first read instructions that described the goal, available operators, and constraints of the Tower of Hanoi puzzle. Then they proceeded to the training session. In this session, they were given the puzzles with varying initial states and required to solve them within 15 seconds. The goal was fixed so that all disks were placed on the rightmost peg. After subjects could solve them successfully, they proceeded to the next session: pre-judgement session.

In this session, subjects were given a twenty-six page booklet. On each page, one of 26 states of the three-disk Tower of Hanoi puzzle was printed, paired with the goal where all disks were placed on the rightmost peg. Subjects were asked to rate how similar the states were, and to circle "7" if the pictures were very similar, "1" if they were least similar, and other numbers for the intermediary degrees of similarity. Subjects were instructed to respond as quickly as possible. After practice, subjects were given another booklet which consisted of three blocks of 26 pairs. Thus, subjects were required to compare the 78 pairs in the same way as they had done in the previous session.

The procedure for the novice condition was basically same, except that there was no training session. Thus, they read the instructions of the puzzle, then proceeded to the pre-judgement session, and finally rated the similarities of each 78 pair. The control group performed the pre-judgement session and the similarity judgements only. Thus, they had no idea that the presented stimulus was the puzzle. The orders of the stimulus presentation in the pre-judgement and final session were randomized across subjects.

Results and Discussion

In order to examine the effects of the number of shared features and operators, we calculated Spearman's rank order correlation coefficients between the rated similarity and the number of operators and the number of shared features. The number of operators was defined as the distance between a given state and the goal in the problem space of the puzzle. The number of shared features was calculated by adding attributional and relational similarities. The degree of attributional similarity was the number of disks on the target (rightmost) peg. That of relational similarity was the number of *on*-relations. For

Table 1: Spearman's rank order correlation coefficients of rated similarity and the number of features and distance

	No. Shared Feature	No. Distance
Expert	0.328^{**}	-0.534^{**}
Novice	0.425^{**}	-0.150^{*}
Control	0.624^{**}	-0.034

Note: * shows $p < .05$, ** $p < .01$
r_s between the distance and the feature is 0.038.

example, suppose a state where the largest disk is located on the leftmost peg and the other disks are on the rightmost peg. In this case, the degree of attributional similarity is 2 because two disks are on the rightmost peg, and that of relational similarity is 1 because the smallest disk is *on* the medium disk.

The results are shown in Table 1. The similarity ratings in the expert condition are greatly affected by the number of operators, although the number of shared features also affects the ratings. In contrast, ratings in the control condition are based solely on the number of shared features. The more features two states have in common, the more they are judged to be similar. The performance of the novice condition is in between the expert and control conditions. Although the effects of operators are observed, the judgements are oriented mainly by the number of shared features. Additionally, we performed separated ANOVAs, taking the distance and the conditions as the independent variables, and the rated similarity as the dependent variable. All the ANOVAs show significant interaction (for one shared feature, $F(6, 240) = 3.38(p < 0.01)$; for two shared features, $F(2, 120) = 3.80(p < 0.05)$; for three shared features, $F(4, 159) = 6.41(p < 0.01)$).

Our hypothesis is supported by the results that the number of operators affects the rating greatly for the expert condition, moderately for the novice condition, and hardly for the control condition. The differences between the control and the other conditions suggest that the recognition of the goal causes subjects to compute the similarity by explanation. The difference between the expert and novice conditions suggests that the accessibility of the appropriate operators determines whether the judgements are carried out by the explanation or the feature comparison. The reason why the effect of the number of shared features was found even in the expert condition may be due to the fact that the experts carried out the judgements by feature comparison when the required explanation was very complicated.

Although we concluded that differences between the novice and expert are attributed to the difference in the accessibility of operators, there might be an alternative interpretation. Since subjects in the expert condition have a lot of opportunities to observe a sequence of solutions in the training session, they may recognize that states closer to the goal are always contiguous to the goal. Therefore, judgements of the experts might reflect the contiguity rather than the accessibility of operators.

Experiment 2

In order to examine whether the expert's performance reflects the learned contiguity, we conducted another experiment where comparisons were made between two expert groups.

It is well known that there are several different strategies to solve the Tower of Hanoi puzzle (Simon, 1975). One of the strategies, called the "perceptual strategy", can be described as follows: To construct the tower of the disks on the target peg, the largest disk must be placed on the target peg first, then the next largest, and so on. This strategy does not always specify the appropriate operator at each state, because some moves directed by the strategy violate the constraints of the puzzle. However, this strategy gives the better understanding of the subgoal structure for solving the puzzle. Another strategy is called "move-pattern strategy." This strategy can be described as follows: On odd-numbered moves, move the smallest disk; On even-numbered moves, move the next-smallest disk that is exposed; The smallest disk is always moved from the left to the right to the center to the left peg, and so on. As is obvious from the above description, this strategy is quite the opposite of the perceptual strategy. This strategy always specifies the appropriate operators. However, this strategy is rather mechanical or rote, in a sense that people do not have to recognize the subgoal structure at all. What is necessary for the strategy is only to keep track of the parity of the move and the cycling direction for the smallest disk (Simon, 1975).

What happens if the two strategies are used in the similarity judgement? Since a subject using the perceptual strategy understands the subgoal structure, he or she is likely to give good, but not exact, estimates of the number of operators to achieve the goal. For example, when the first subgoal of the strategy has not been achieved yet, he or she may judge the given state to be less similar to the goal. When the second subgoal has been achieved, the state may be judged to be very similar to the goal. On the other hand, a subject who uses the move-pattern strategy may have difficulties in estimating the distance. If a subject tries to explain the given state, he has to move the disks mentally and count the number of operators to be applied. Since it places a substantial burden upon working memory, it is likely that the subject would give up the explanation and shift to feature comparison.

It is important to note that the solution paths are identical between the two conditions. Thus, the "contiguity" hypothesis predicts that there is no difference between the two, because subjects in both conditions observed approximately the same number of the sequence of states in the practice session. On the other hand, our model predicts differences between the two. Subjects who use the perceptual strategy should be more sensitive to the number of operators required to achieve the goal, because they are more likely to recognize the subgoal structure which provides a good basis for the estimation of the distance. By contrast, a subject who uses the move-pattern strategy is likely to judge the similarity on the basis of feature comparison. That is because mentally executing this strategy places a substantial burden upon working memory.

Since the pilot study revealed that the differences were very subtle, we made several changes in order for the experiment to be sensitive to possible differences. First, the 7 points scale of the rating in Experiment 1 was replaced with a 10 points scale. In addition, the five-disk Tower of Hanoi was used for rating, so as to avoid the "ceiling effect."

Method

Subjects Twelve undergraduate students were randomly assigned to the subgoal or rote condition. None of them had experienced with the Tower of Hanoi puzzle prior to the experiment.

Procedure Subjects in both conditions first read instructions which described the rules of the puzzle. Then they were given a description of the strategies to be learned: Subjects in the subgoal condition read the description of the perceptual strategy; Subjects in the rote condition read the description of the move-pattern strategy. Then subjects were asked to understand the procedure. When they did not understand it, an experimenter taught them the strategy according to the instruction. After reading it, they were required to solve the three-disk puzzle, using the taught strategies. The initial state of the practice was fixed so that all the disks were placed on the leftmost peg. If subjects solved the puzzle within ten seconds without mistakes, they were allowed to proceed to the next session. After the practice for rating, subjects were given a nine-page booklet, and asked to judge the similarity of a state to the goal. The goal and one of the nine states were printed on each page. Subjects were asked to judge the similarity of the pairs, as quickly as possible. The nine states that were used for the comparison were selected to approximately balance the number of features and the distance from the goal.

Results and Discussion

It took 116 seconds for subjects in the subgoal condition and 126 for those in the rote condition to solve the five-disk puzzle. This suggests that there is no difference in efficiency of the strategy use in both conditions.

However, Spearman's rank order correlation coefficients between the rated similarity and the number of operators show that there exist differences between the two groups of subjects (for the subgoal condition, $r_s = -0.479 (p < 0.01)$; for the rote condition, $r_s = -0.219 (p = 0.11)$.

These results indicate that the differences observed in the experiment 1 could not be attributed to mere recognition of contiguity. The difference is due to the understanding of the subgoal structure which provides a good basis for estimating the number of operators necessary to achieve the goal.

General Discussion

Similarity must be sensitive to goals, since it is involved in various kinds of human activities in which goals play privileged roles. The experiments presented here clearly show that the judgements of similarity are affected by the goal and knowledge of operators. When the goal is salient, people's judgements of similarity are carried out by *explanation*, sensitive to the number of operators required to transform given states to the goal. On the other hand, judgements come to be based on the number of shared features when there is no explicit goal, or when relevant knowledge to achieve the goal is not readily accessible.

By incorporating goal and knowledge into the model of similarity, we can provide adequate accounts for several phenomena found in people's judgement of similarity. First, our model has direct relevance to the "surface-structural" argument in studies of similarity. Gentner & Landers (1985) found that while people retrieved superficially similar stories in a memory recall task, they tended to choose structurally similar ones in tasks which required the rating of the soundness of analogy between stories. The shift from superficial to structural similarity can be attributed to the fact that there is no explicit goal in the recall task, whereas the goals and the solutions are salient in the rating of analogical soundness. As we suggested before, the recognition of the goal leads people to compute similarity by explanation. In this case, the subjects were sensitive to the goal-subgoal hierarchy which corresponds to the "structure" of the task. This would be the reason why subjects' ratings were based on the structure of the stories in the soundness rating.

More evidence of the "surface-structural" distinction comes from studies on expert-novice differences. It is well known that whereas experts attend to structural aspects of problems, novices attend to superficial ones. These results can be explained by the accessibility of knowledge. In the categorization task in Chi et al's experiments (Chi, Feltovich, & Glaser, 1981), novices seemed to know that they should attend to structural similarity among problems because they had been taught an elementary physics. However, the lack of the appropriate knowledge of physics which related one problem to another caused them to compute the similarity via on the number of shared features.

Our model has much in common with the MAC/FAC model (Gentner & Forbus, 1991). The MAC/FAC model consists of two stages. While in the MAC stage, computationally cheap matchers act on content vectors of items in LTM, structural examinations are made in order to compute "deep" similarity in the FAC stage. The MAC and FAC stages correspond to the feature comparison and the explanation, respectively. This suggests that theories which aim at modeling the processes of similarity judgements in problem-solving contexts should have two subprocesses to compute deep as well as shallow similarities.

However, there are several differences between the two. First, the role of the goal is not explicitly mentioned in the MAC/FAC model. Although the MAC/FAC might be able to explain the effects of goal by modifying content vectors, an initial set of features has to be changed. In contrast, our model explains the effects not by modifying the description of objects, but by adding the goal and the operators. Second, although the accessibility of appropriate operators determines whether the judgements are carried out by explanation or feature comparison, there seem no mechanisms in the MAC/FAC model to explain the effects.

Acknowledgements

This research was partly supported by the British Council. We thank Dedre Gentner, Jon Oberlander, Keith Stenning and Sam Stern for their helpful comments.

References

Chi, M. T. H., Foltovich, P. J., and Glaser, R. 1981 Categorization and representation of physics problems by experts and novices. *Cognitive Science* **5**:121–152.

Gentner, D. and Forbus, K. D. 1991 MAC/FAC: A model of similarity-based retrieval. In Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society, August, 1991.

Gentner, D. and Landers, R. 1985 Analogical reminding: A good match is hard to find. *Proceedings of the International Conference on Systems, Man, and Cybernetics.* Tucson, AZ.

Glaser, R. 1984 Education and thinking: The role of knowledge. *American Psychologist* **39**:93–104.

Goldstone, R. L., Medin, D. L. and Gentner, D. 1991 Relational Similarity and the Nonindependence of Features in Similarity Judgments. *Cognitive Psychology* **23**:222–262.

Hammond, K. J. 1990 Case-based planning: A framework for planning from experience. *Cognitive Science* **14**:385–443.

McCarthy, J. and Hayes, P. J. 1969 Some philosophical problems from the standpoint of artificial intelligence. *Machine Intelligence* **4**:463–502.

Mitchell, T. M., Keller, R., and Kedar-Cabelli, S. 1986 Explanation-Based Generalization: A Unifying View. *Machine Learning* **1**:47–80.

Pylyshyn, Z. W. 1984 *Computation and Cognition: Toward a Foundation for Cognitive Science.* Cambridge, MA.: MIT Press.

Resnick, L. B. ed. 1990 *Knowing, Learning, and Instruction.* Hillsdale, NJ.: Erlbaum.

Smith, E. E. 1990 Categorization and induction, In M. I. Posner ed. *Foundations of Cognitive Science.* Cambridge, MA.:MIT Press.

Smith, L. B. 1989 From global similarities to kinds of similarities: The construction of dimensions in development. In S. Vosniadou. and A. Ortony eds. *Similarity and Analogical Reasoning.* Cambridge, MA.: Cambridge University Press.

Simon, H. A. 1975 The functional equivalence of problem solving skills. *Cognitive Psychology* **7**:268–288.

Tversky, A. 1977 Features of Similarity. *Psychological Review* **84**:327–352.

Tversky, A. and Gati, I. 1978 Studies of similarity. In E. Rosch and B. Lloyd eds. *Cognition and Categorization.* Hilldale, NJ: Lawrence Erlbaum Associates.