Analogical retrieval from everyday experience: Analysis based on the MAC/FAC

Junya Morita (morita@cog.human.nagoya-u.ac.jp)
Graduate School of Human Informatics
Nagoya University, Furo-cho, Chigusa-ku, Nagoya City, Japan

Kazuhisa Miwa (miwa@cog.human.nagoya-u.ac.jp)
Graduate School of Information Science
Nagoya University, Furo-cho, Chigusa-ku, Nagoya City, Japan

Abstract

In order to investigate how analogs are retrieved from everyday experience, we conducted an experiment in which subjects were not presented with analogies by an experimenter, but presented with only one story as a retrieval cue. In our experiment, subjects were divided into four groups varying the cue stories, which were manipulated by their surface and structural features. In all the groups, the subjects were asked to report the cases that came to mind when they read the cue story. After retrieving, the subjects rated the inferential soundness (goodness as analogy) of each retrieved case. We computed similarities between each retrieved case and the cue stories, using a computational model MAC/FAC ("many are called but few are chosen"), which was developed for simulating two stages of analogy making (Forbus, Gentner, & Law, 1995). The results showed that (1) the retrieved cases were similar to the presented story in the surface features rather than in the structural features and (2) the structural similarity between the retrieved cases and the presented story increased with the rated scores of inferential soundness. These results confirmed that, as the results of prior controlled experiments suggested, the surface similarity guides the retrieval of cases and the structural similarity guides the evaluation of the cases.

Introduction

Analogy making is a core component of higher-order cognition, such as problem solving (Gick and Holyoak, 1980), decision-making (Markman and Moreau, 2001), and creative generation (Smith, Word, and Schumacher, 1993). In the past two decades, many researchers have conducted controlled experiments and gained reliable findings on analogy making. However, there have been only a few studies to verify the findings in less controlled environments. Our goal here is to replicate previous findings on analogy making in extended laboratory settings.

Prior to presenting our experiment, we briefly review a framework developed in the area of analogy research. First, in analogy research, a representation of a novel situation is called a target, and a past case that is similar to the target is called a base. The process of analogy making is comprised of two main components: the retrieval of the base and the mapping from the base to the target.

It has been pointed out that the analogy process is guided by similarities between the base and the target. Using propositional representations (predicate-argument formalism), Gentner (1983) distinguished three types of correspondence between the base and the target.

- Correspondence of *attributes*: e.g., The sun is round and yellow → The orange is round and yellow [sun (round) sun (yellow) → orange (round) orange (yellow)]
- Correspondence of *first-order relations*: e.g., The planets revolve around the sun. → The electrons revolve around the atom [revolve-around (planet, solar) → revolve-around (electron, atom)]
- Correspondence of *higher-order relations*: e.g., Because the sun attracts the planets, the planets revolve around the sun. → Because the atom attracts the electrons, the electrons revolve around the atom [cause (attract (solar, planet), revolve-around (planet, solar)) → cause (attract (atom, electron), revolve-around (electron, atom))]

The above discrimination was based on the types of predicates. The attribute is a predicate type that takes only a single argument. On the other hand, the first-order and higher-order relations are predicate types that take multiple arguments. There is no depth in the former, but there is in the later. Based on this discrimination, in analogy research, further discrimination of similarity has been proposed: i.e., surface similarity and structural similarity. The degree of surface similarity is roughly defined as the number of attributes shared between the base and the target. Contrary to the surface similarity, the degree of structural similarity is defined as the depth of structural mapping from the base to the target; so correspondence of higher-order relations is deeper than that of first-order relations (Gentner, 1983).

It has been demonstrated that two types of similarities take different roles in the analogy process (Holyoak, & Koh, 1987; Gentner, Rattermann, & Forbus, 1993; Wharton, Holyoak, Downing, Lange, Wickens, & Melz, 1994). For example, Gentner, Rattermann, & Forbus (1993) conducted experiments in which subjects learned several stories and then retrieved the learned stories when they read new cue stories. The cue stories were manipulated by two factors: surface and structural similarities to the learned stories. As a result, the subjects retrieved more often the surface similar stories than the structurally similar stories. However, once the subjects were presented with the learned stories with the cue stories, they rated the inferential soundness (goodness as analogy) of the structurally similar stories higher than that of the surface similar stories.

To explain these results, Forbus, Gentner, & Law (1995) proposed a computational model, called MAC/FAC ("many are called but few are chosen"), which simulates two stages of the analogy process. In the first stage of MAC/FAC, several potential bases are retrieved from a memory pool, computing the dot product of the target's content vector (CVector), which is a simple list of the predicates contained in the propositional representation, with the CVector of each case in the memory pool. In the next stage, the cases retrieved at the initial stage are further evaluated by using the Structure-mapping Engine (SME), which computes structural alignment and evaluation of the match between each set of cases and the target (Falkenhainer, Forbus, & Gentner, 1989). Finally, MAC/FAC selects the cases that have high structural evaluation scores (SES), which indicate the degree of depth and breadth of the common structure. In brief, MAC/FAC can discriminate the initial stage of retrieval that is guided by surface similarity from the evaluation stage that is guided by structural similarity. In the past, similar discrimination has been employed in many other models of analogy (Thagard, Holyoak, Nelson, & Gochfeld, 1990; Hummel & Holyoak, 1997).

Recently, however, limitations on the above finding have been pointed out. The limitations are derived from the fact that many analogy researchers have only dealt with cases created by the researchers themselves. In other words, the experiments have been conducted in closed laboratories, where the subjects retrieved cases created by researchers in advance. In real-world situations, it is impossible to predict the cases that will be retrieved or used. In real-world situations, the analogy is made from individuals' everyday experience. Therefore to extend the findings to realistic problems, it is necessary to investigate analogy making using cases that the subjects learn in their own everyday life.

From this viewpoint, Blanchette & Dunbar (2000) conducted experiments that examined analogy making in situations where the subjects were not provided analogies guided by an experimenter. In their experiments, the subjects were asked to generate analogies to the zero-deficit problem - the deficit that Canadian governments had to cut. The results showed that the subjects generated few analogies that have surface features in common with the target (the zero-deficit problem), but generated many analogies that shared deep structures with the target. Further, being asked to select the best analogy from the generated analogies, the subjects selected analogies that had deeper structural correspondence than the others.

These results indicate the strong effect of structural similarity on both the retrieval and the evaluation stages, contradicting the previous studies that showed different similarities involved in the two stages. Based on the results, Blanchette & Dunbar claimed that surface similarity has little effect on analogy making in situations where subjects use their own analogies.

Although we agree on the importance of their approach, which aims to combine naturalistic settings and controlled experiments (Dunbar & Blanchette, 2001), we think that further investigation is needed for their claim. Thus, we reexamined the similarity effects on the analogy process in a situation where the subjects retrieved cases that were learned in their own everyday life. The method of our experiment is similar to that of Blanchette & Dunbar, but there are three important differences, as follows.

First, we modified the instruction in which the subjects were asked to "generate analogies". Because many researchers argued that the term "analogy" commonly implies "the cases that have low surface similarity and high structural similarity" (e.g., Gentner, 1983), there exists a possible other account for Blanchette & Dunbar's results: The subjects might actually be reminded of surface similar cases, but would not report those cases. In order to test this possibility, we did not include the term "analogy" in our instruction.

Second, we constructed several controlled experimental conditions. Blanchette & Dunbar conducted the experiments without clear manipulations using surface or structural similarities. In such an experiment, it is difficult to exclude possible conjectured factors, such as the types of cases that subjects hold, or the frequency of using these cases in their everyday life. To control these factors, we divided subjects into several experimental groups and presented the targets whose surface and structural features were systematically changed.

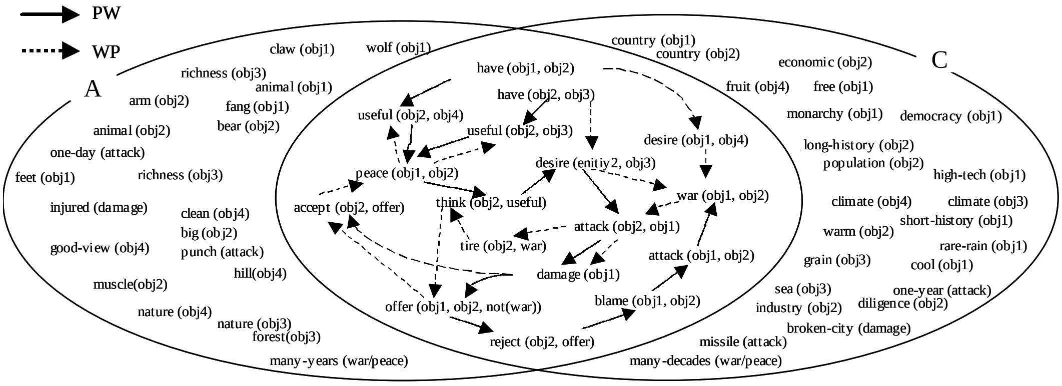
Third, we analyzed the data quantitatively. In Blanchette & Dunbar's study, the generated analogies were analyzed by categorizing their surface/structural features and comparing frequencies of categories. However, they did not show how much the generated analogies shared surface/structural features with the target. For quantitative analysis, we computed similarity scores for each retrieved case based on the algorithm assumed in the MAC/FAC (Forbus, Gentner, & Law, 1995).

Method

Materials

The experiment was conducted to investigate the effects of similarities on the analogy process in a situation where subjects retrieved cases that were learned in their own everyday life. In our experiment, the subjects were presented with a cue story and then were asked to report the cases that came to mind while they read the cue story. The cue story consisted of about 600 Japanese characters. In this paper, we call these stories the target stories.

The texts of the target stories were manipulated with their surface and structural features. The subjects were divided into four experimental groups varying target stories (the between-subjects factor). As the surface features, a set of attributes related to *animals* (A) and a set of attributes related to *countries* (C) were chosen. As the structural features, a story whose plot is a transition from *peace to war* (PW) and a story whose plot is a transition from *war to peace* (WP) were created. Combining the surface and structural features, four types of



CVector between targets

	A	C
A	29	0
C		29

SES between targets

	PW	WP
PW	28.8	12.2
WP		27.9

Figure 1: Propositions contained in the target stories.

target stories were prepared (A/PW, A/WP, C/PW, and C/WP).

Figure 1 shows the propositions converted from the texts in the target stories. Each of them was included in either set A or set C. Two complements $[(A \cap \bar{C})$ and $(\bar{A} \cap C)]$ contain attributes of objects and an intersection $(A \cap C)$ includes first-order relations between two objects. Each of the first-order relations was connected by two types of higher-order relations (PW/WP) represented by two types of arrows (solid/dotted).

Figure 1 also shows similarity scores calculated based on the algorithms of the MAC/FAC. The CVector, indicating surface similarity, was computed as a dot product of each pair of surface features (A vs. A, A vs. C, and C vs. C). In our study, the CVector was represented as a list of attributes, not containing first-order and higher-order relations. The SES, indicating structural similarity, was computed by inputting each pair of relational structures (PW vs. PW, PW vs. WP, and WP vs. WP) into the SME model. From the several matching rules of the SME package, we chose analogy rules that do not compute the match of attributes.[1]

In order to verify the above manipulation, we conducted a preliminary experiment. The subjects (n = 8) were presented with four target stories and then rated the inferential soundness of each pair of the target stories on a 1 ("low") – 5 ("high") scale. Similar to Gentner, Rattermann, & Forbus (1993), soundness was explained as "the degree to which inferences from one story would hold for the other".

[1]Our way of computing similarity was slightly modified from Forbus's method. Forbus treated the CVector as a list of all types of predicates including relations, and computed the SES using the literal-similarity rules that mapped all types of predicates including attributes (Forbus, Gentner, & Law, 1995). In our study, for clear discrimination between surface (attributes) and structural similarity (relations), we chose the above method.

The results showed that the manipulation is consistent with human feelings of soundness. Seven of eight subjects judged the structurally similar pairs (A/PW vs. C/PW and A/WP vs. C/WP) as having higher inferential soundness than the other pairs (A/PW vs. A/WP, CPW vs. C/WP, A/PW vs. C/WP, and A/WP vs. CPW).

Participants

Thirty-three undergraduate and graduate students participated in the experiment. They were divided into four groups: a group presented with A/PW (n = 8), a group with A/WP (n = 9), a group with C/PW (n = 8), and a group with C/WP (n = 8).

Procedure

The subjects participated in the experiment individually or in groups of two to four. The experiment was divided into the following three phases.

Retrieval phase In the first phase, the subjects reported the cases of which the target story reminded them. In explaining the task, we avoided using terms like "analogy" or "analogous". The subjects were simply told that "while reading the presented story, you should write out any cases that come to mind". After the instruction, the subjects were presented with one of the four targets and then they wrote down any reminded cases. This phase continued for twenty minutes.

Evaluation phase Following completion of the retrieval phase, the subjects were given a soundness rating task. The subjects rated the soundness of the match between each retrieved case and the presented target on a 1 – 5 scale.

Subjects' descriptions	Converted propositions
The story about two tigers. In a forest, two tigers lived. Each of them has a turf. And they battled each other for the turf. One day, an animal that lived in the forest persuaded one of the tigers to stop fighting. After this persuasion, the relationship between the two tigers became peaceful. CVector (A) = 8 CVector (C) = 0 SES (PW) = 7.59 SES (WP) = 11.75	((animal tiger1) :name prop3) ((animal tiger1) :name animal2) ((animal animal1) :name animal3) ((have tiger1 turf1) :name have1) ((have tiger2 turf2) :name have2) ((desire tiger1 turf2) :name desire1) ((desire tiger2 turf1) :name desire2) ((war tiger1 tiger2) :name war1) ((and desire1 desire2) :name and1) ((cause and1 war1) :name cause1) ((not war1) :name not1) ((offer animal1 tiger1 not1) :name off) ((accept tiger1 off) :name accept1) ((cause offer1 accept1) :name cause2) ((cause war1 off) :name cause3) ((peace tiger1 tiger2) :name peace1) ((cause accept1 peace1) :name cause4) ((many-tree turf1) :name prop1) ((many-tree turf2) :name prop2)
The gallic war. In order to expand the national land, the Roman Empire kept attacking other countries. CVector (A) = 0 CVector (C) = 5 SES (PW) = 6.18 SES (WP) = 6.10	((country garia) :name country1) ((country other) :name country2) ((monarchy garia) :name prop1) ((have other land) :name have1) ((attack gallia other) :name attack1) ((desire gallia land) :name desire2) ((war gallia other) :name war1) ((cause desire2 attack1) :name cause1) ((cause attack1 war1) :name cause2)

Figure 2: Examples of subjects' descriptions and propositions.

Explanation phase Finally, the subjects were asked to explain the retrieved cases in as much detail as possible.

Coding

The retrieved cases were coded using propositional representations. The subjects' descriptions were segmented by the appearance of a predicate. Then a coder judged whether each segmented sentence could be represented as a proposition by using predicates contained in the targets (the predicates in Figure 1). If possible, a proposition would be constructed by complementing for proper arguments. Examples of the coding are shown in Figure 2.

Results and Discussion

The total number of cases retrieved by the subjects was 266. There was no significant difference on the number of cases among the four experimental groups [$\chi^2(3) = 6.15, ns.$]. Thus, we treated each retrieved case as an individual datum for statistical tests.

In order to examine the relationship between the surface/structural similarities and the retrieval/evaluation stages of analogy making, we tested (1) whether the retrieved cases were similar to the presented target in the surface/structural features, and (2) whether the surface/structural similarity between the retrieved cases and the presented target increased with the degree of

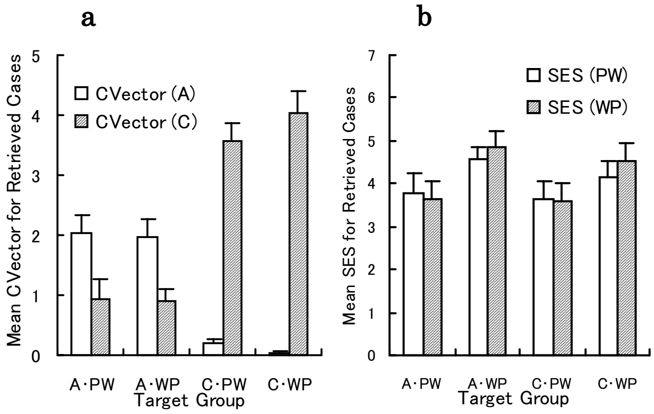


Figure 3: (a) Mean CVector for four groups. (b) Mean SES for four groups. *Note.* Error bars represent one standard error of mean.

soundness rating.

1. Effects of Similarities on Retrieval

For investigation of the effects of similarities on retrieval, four types of similarity scores were computed based on the algorithm assumed in the MAC/FAC.

- *CVector (A)* was computed as the dot product between each retrieved case and the surface feature A ($A \cap \bar{C}$ in Figure 1).
- *CVector (C)* was computed as the dot product between each retrieved case and the surface feature C ($\bar{A} \cap C$) in Figure 1).
- *SES (PW)* was computed by inputting each retrieved case and the structural feature PW (the solid lines in Figure 1) into the SME model.
- *SES (WP)* was computed by inputting each retrieved case and the structural feature WP (the dotted lines in Figure 1) into the SME model.

We conducted two ANOVAs to investigate interaction between the above similarity scores and the experimental groups. If the surface/structure features affected the case retrieval, the retrieved cases would be similar to the presented target rather than the targets that were not presented for each group.

Effects of Surface Similarity on Retrieval Figure 3a shows the mean CVector for each group. A 2 × 2 × 2 surface features of targets (between) × structural features of targets (between) × types of CVector (within) ANOVA revealed significant interaction between the surface features of targets and the types of CVectors [$F(1, 262) = 118.21, p < .05$], indicating CVector (A) was higher than CVector (C) in the group A [$F(1, 262) = 14.22, p < .05$], and CVector (C) was higher than CVector (A) in the group C [$F(1, 262) = 134.67, p < .05$].

Both in group A and group C, the retrieved cases were more similar to the target that was presented for each group than the targets that were not presented. A strong effect of surface similarity on retrieval contradicts the results of Blanchette & Dunbar (2000), but is consistent with the findings of the previous controlled experiments (Holyoak, & Koh, 1987; Gentner, Rattermann, & Forbus, 1993; Wharton et. al, 1994).

Effects of Structure Similarity on Retrieval Figure 3b shows the mean SES for each group. A 2 × 2 × 2 surface features of targets (between) × structural features of targets (between) × types of SES (within) ANOVA revealed significant interaction between the structural features of the target and the types of SES [$F(1, 262) = 8.01, p < .05$]. However, simple main effects were significant only in group WP [$F(1, 262) = 7.50, p < .05$]. There was no significant difference of types of SES in group PW [$F(1, 262) = 1.60, ns.$].

These results suggest that structural similarity has more restricted effects on retrieval than surface similarity. Again, this result contradicts the study by Blanchette & Dunbar, but is consistent with the findings of the previous controlled experiments (Holyoak, & Koh, 1987; Gentner, Rattermann, & Forbus, 1993; Wharton et. al, 1994).

2. Effects of Similarities on Evaluation

In order to investigate the effects of structural similarity on the evaluation stage, we treated the subject groups as counterbalance conditions, and reduced the number of factors for the ANOVA. Therefore, we investigated four types of similarity scores as follows:

- *CVector (presented)* was computed by combining CVector (A) in group A and CVector (C) in group C. This score indicates the degree of surface similarity, meaning how many attributes were shared between each retrieved case and the target that was presented for the subjects.

- *CVector (not presented)* was computed by combining CVector (A) in group C and CVector (C) in group A. This score indicates how many attributes were shared between each retrieved case and the target that was not presented for the subjects. Because two types of surface features (A and C) share no attributes, this score indicates surface dissimilarity.

- *SES (presented)* was computed by combining SES (PW) in group PW and SES (WP) in group WP. This score indicates the depth of structural mapping from each retrieved case to the presented target. Thus, this score indicates structural similarity that reflects higher-order relations.

- *SES (not presented)* was computed by combining SES (PW) in group WP and SES (WP) in group PW. This score indicates the depth of structural mapping from each retrieved case to the target that was not presented for the subjects. Since two structural features (PW and WP) share no higher-order relations but only first-order relations, this score indicates the degree of overlap of the first-order relations.

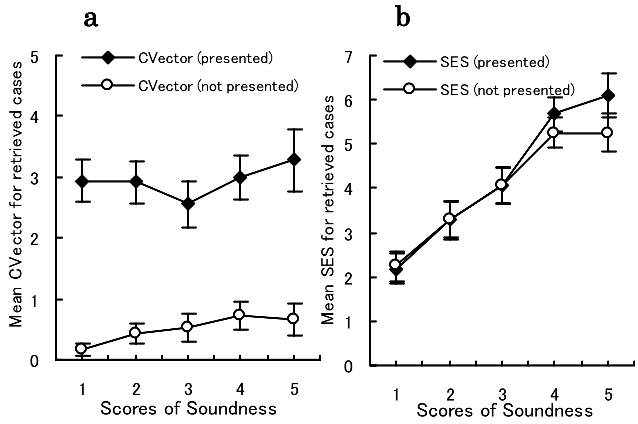


Figure 4: (a) Mean CVector for rating scores of soundness. (b) Mean SES for rating scores of soundness. *Note.* Error bars represent one standard error of mean.

We computed two ANOVAs for the retrieved cases by using the above similarity scores. Each ANOVA tested whether the similarity scores increased with the rated scores of soundness (1 − 5). If the structural similarity had a strong effect and the surface similarity had only a little effect in the evaluation stage, as suggested by the previous studies (Gentner, Rattermann, & Forbus, 1993; Blanchette & Dunbar, 2000), the CVector (presented/not presented) would not increase with the soundness rating, but the SES (presented/not presented) would increase with the soundness rating. Further, the SES (presented), which reflects the higher-order relations, would be related to the soundness ratings more than the SES (not presented), which reflects only the first-order relations.

Effects of Surface Similarities on Evaluation Figure 4a shows the mean CVector for each score of soundness (1 − 5). A 5 × 2 soundness scores (between) × CVector types (within) ANOVA detected a significant main effect of CVector types [$F(1, 261) = 121.17, p < .05.$]. However, a main effect of soundness [$F(4, 261) = 1.91, ns.$] and an interaction between CVector types and soundness scores [$F(4, 261) = 0.40, ns.$]) were not significant. These results indicate an advantage of surface similarity over surface dissimilarity regardless of soundness ratings. Thus, as the previous studies indicated, the results suggest that there is no effect of surface similarity or dissimilarity on the evaluation stage.

Effects of Structural Similarity on Evaluation Figure 4b shows the mean SES for each score of soundness (1 − 5). A 5 × 2 soundness scores (between) × SES types (within) ANOVA revealed a significant interaction between soundness scores and SES types

$[F(4, 261) = 7.52, p < .05]$. Simple main effects of soundness scores were significant for both SES (presented) $[F(4, 261) = 15.46, p < .05]$ and SES (not presented) $[F(4, 261) = 11.09, p < .05]$. Further it was confirmed that SES (presented) was higher than SES (not presented) at the soundness scores 4 $[F(1, 261) = 8.76, p < .05]$, and 5 $[F(1, 261) = 35.73, p < .05]$.

Significant main effects of soundness on both the SES (presented) and the SES (not presented) indicate that the subjects' evaluation was positively correlated to the degree of commonalities in the first-order relations. The fact that the SES (presented) was higher than the SES (not presented) in the cases in which the subjects rated high soundness (rate scores 4 and 5) implies that commonalities in the higher-order relations are more strongly related to soundness than mere first-order relations. In summary, these results are consistent with prior studies (Gentner, Rattermann, & Forbus, 1993; Blanchette & Dunbar, 2000) showing the strong effects of structural similarity on evaluation.

General Discussion

Influence of Similarities on Analogy Process

In this study, we investigated types of similarities influencing the analogy process, conducting an experiment in which the subjects retrieved cases that they had learned in their own everyday life. As with the results of previous controlled experiments, our results also suggest that different types of similarities are responsible for the retrieval and evaluation stages of the analogy process. The results were different from those of Blanchette & Dunbar (2000), which showed little effect of surface similarity on the retrieval phase. The difference between our results and Blanchette & Dunbar's results could be explained by differences in the instructions. The subjects who participated in the experiments of Blanchette & Dunbar may have filtered out the surface similar cases because they were instructed to "generate analogies". Since there exist other differences between our experiment and Blanchette & Dunbar's experiments, such as the reality of the tasks used and the method of analysis, we must conduct further experiments controlling for these factors.

In addtion, our study obtained results indicating a strong effects of structural similarity on the evaluation stage. The results were clearer than Blanchette & Dunbar's results. Blanchette & Dunbar's analysis was based on counting elements shared with the base and the target, without any consideration of relational structures. In contrast, our analysis was based on a computational model that computes structural alignment and structural evaluation. Our analysis showed that the degree of shared attributes did not increase with the soundness rating, whereas the degree of shared relations increased with the soundness rating. Further, sharing higher-order relations made the relation to the soundness rating stronger. These results are important for the extension of the systematicity principle, proposed by Gentner (1983), which predicted that deeper structural mapping would be preferred in an analogical inference.

Investigation based on a Computational Model

The above results imply the benefit of using a computational model for analysis of psychological data. In the past, few studies have used a computational model to analyze data obtained from psychological experiments. However, without a sufficient computational model, it would be impossible to investigate complex cognitive conceptual products such as structural similarity.

Recently, in the community of cognitive science, the connection between theory and experimental data has been stressed. Usage of a computational model for analysis, demonstrated in this paper, could open a new way of directly licensing these two entities that play the most important role in cognitive scientific studies.

References

Blanchette, I., & Dunbar, K., (2000). How analogies are generated: The role of structural and superficial similarity. *Memory & Cognition*, *28*, 108–124.

Dunbar, K., & Blanchette, I, (2001). The in vivo/in vitro approach to cognition: the case of analogy. *Trends in cognitive science*, *5*, 334–339.

Falkenhainer, B., Forbus, K. D., & Gentner, D., (1989). The Structure-Mapping Engine: Algorithm and Example. *Artificial Intelligence*, *41*, 1–63.

Forbus, K. D., Gentner, D., & Law, K. (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science*, *19*, 141–205.

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

Gentner, D., Rattermann, M., & Forbus, K. D. (1993). The role of similarity in transfer: Separating retrievability for inferential soundness. *Cognitive Psychology*, *25*, 524–575.

Holyoak, K. J., & Koh, K. (1987). Surface and Structural Similarity in Analogical Transfer. *Memory & Cognition*, *15*, 332–340.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, *104*, 427–466.

Markman, A. B., & Moreau, C. P. (2001). Analogy and analogical comparison in choice. In D. Gentner, K. Holyoak & B. N. Kokinov (Eds.), *Analogical Mind: Perspectives from cognitive science*. Cambridge, MA: The MIT press

Smith, S. M., Ward, T. B., & Schumacher, J. S. (1993). Constraining effects of examples in a creative generation task. *Memory & Cognition*, *21*, 837–845.

Thagard, P., Holyoak, K. J., Nelson, G., & Gochfeld, D. (1990). Analog retrieval by constraint satisfaction. *Artificial Intelligence*, *46*, 259-310.

Wharton, C. M., Holyoak, K. J., Downing, P. E., Lange, T. E., Wickens, T. E., & Melz, E. R., (1994). Below the surface: Analogical similarity and retrieval competition in reminding *Cognitive Psychology*, *26*, 64–101.