Justification of Analogy by Abstraction

Hiroaki Suzuki

Department of Education,
Aoyama Gakuin University
Shibuya 4-4-25, Tokyo
150-8366, Japan
susan@ri.aoyama.ac.jp

Introduction

How is reasoning by analogy justified? Why can we map non-identical elements in the source and target analogs? Is it valid to transfer some elements in one analog to another? Although the problem of justification is of critical importance for the research on analogy, only a few studies have discussed them seriously (Gentner, 1983; Indurkhya, 1992). The aim of this paper is at developing a new framework that provides an answer to the justification problem.

In the real world, analogical reasoning is widely used and has strong power in various kinds of human activities, such as problem-solving, learning, discovery, natural language, explanation, decision-making, psychotherapy, literature, myth, political and legal argument (Holyoak & Thagard, 1995). It is a powerful tool for providing a solution, creating a new idea, arguing against, and persuading opponents, making ideas more explicit and impressive.

However, some blame analogy and claim not to use it, because analogy is known to be a dangerous mode of reasoning. Analogical reasoning, like induction, does not have logical validity. Actually, there are abundant examples of misuse of analogies in various kinds of human activities, such as education, science, political arguments, commercial advertisement (Gentner & Jeziorski, 1993; Holyoak & Thagard, 1995; Indurkhya, 1992). More depressing findings were obtained by Chi and her colleagues (Chi et al., 1989). Their study found poor learners' excessive reliance on analogy. These learners frequently looked back to previous problems, read them extensively, tried to map them to the current problem, which resulted in poor performance on transfer tasks.

What mentioned above shows two opposing pictures. Analogy enriches human cognition and gives new insights in some cases. In other cases, analogy obscures our rationality and falls into poor learners' desperate heuristics.

The purpose of the paper is to develop a new framework to give explanations of how reasoning by analogy is justified, and in what condition analogies are, at least, psychologically valid.

In the next section, I analyze the conditions for justified analogies. According to the analysis, analogy should be treated as a kind of categorization. This means that analogy is a ternary relation between the base, target, and their superordinate category (abstraction), rather than a binary relation between the base and target. Second, I will show that this formulation greatly reduces the computational complexities in retrieval and mapping. Third, I will try to figure out the characteristics of categories by the findings obtained from an informal observation. Finally, I will reexamine the relationships of analogy to other kinds of cognitive activities, based on the proposed framework.

Justification

Identicality

Although controversies are still continuing about many aspects of analogy, there is one basic assumption that few deny. This assumption is that analogy involves mapping from the base to the target. A set of elements in the base is corresponded to a set of elements in the target. Another set of elements in the base are, then, transferred to the target to create new inferences.

Here, one can ask why some elements in the base can be mapped and transferred to the target. What enables mapping and transfer between the base and target? It is a difficult problem, but Leibnitz gave a partial answer to this problem. According to his principle of the identity of the indiscernibles, if two things are identical, any of their predicates can be transferred. This principle suggests that mapping requires identicality between the base and target.

However, as long as analogy is concerned, this principle is too rigorous to be applied , because a base is not identical to a target, by definition. The base is represented qualitatively different from the target. They are in no way identical.

Categorization

Thus, it is necessary to find a cognitive mechanism that makes two different things identical in some respects. Logically, it is impossible, but there is one psychological mechanism that can do it approximately. It is categorization. If two things belong to the same category, they are properly said to be identical in terms of the category. Suppose, for example, that there are two cats that differ in their size, color, etc. Despite of these differences, they are identical with respect to their "catness." If two cats belong to the same category, attributes and predicates important with respect to the category can be mapped from one cat to the other.

The same argument can be applied to the theory of analogy. If there is a superordinate category whose members are the base and target, they are properly said to be identical, with respect to the category. Thus, properties and relations shared by the base and the category can be transferred to the target.

The discussion so far leads us to change the basic framework of analogy. As I said earlier, analogy has been considered to be a binary relation between the base and target. However, if the argument above is correct, it follows that we should consider analogy as a ternary relation between the base, target and their category. But, the term "category" usually refers to a preexisting taxonomic category, such as animal, plant, dog etc., so I introduce a more neutral term, *abstraction*, here. Note that the term abstraction here refers to an abstracted mental entity, not to the action to abstract.

Attempts have been made to incorporate abstractions to the theory of analogy. In artificial intelligence research, several models of analogy have made explicit use of abstraction (Greiner, 1988; Kedar-Cabelli, 1985; Russell, 1988). Glucksberg and Keysar (1990) and Lakoff (1993) assume abstracted mental entities in understanding metaphorical statements, although there are controversies between them. A number of studies on transfer of learning have shown the importance of abstractions (see, for example, Gick & Holyoak, 1983; Goswami & Brown, 1989). Thus, the framework proposed here is not a new one. Rather, my attempt should be considered as a synthesizing one.

Computational Constraints

By introducing the notion of abstraction, we obtain a couple of constraints that greatly reduce the computational complexities in retrieval and mapping.

Retrieval

An important consequence of introducing the abstraction is that the analog retrieval mechanism can make use of hierarchy. Not a few researchers admit that our long-term memory is represented hierarchically from most concrete to most abstract ones.

If the retrieval mechanism makes use of the information about the hierarchy, the cost of retrieval is obviously reduced. For example, if an abstraction is judged to be irrelevant in the process of categorization, an analogizer needs not consider all of its descendents. Theories ignoring the hierarchical information have to test every subcategory even after its ancestral abstraction is rejected.

There is another benefit. The more ascending a hierarchical tree, the less information is available. Consequently, one may sometimes descend the tree to obtain further information. In this case, the hierarchical structure constrains further search. If you select an abstraction at some level and want to get more information, you need not search the entire space. Instead, it is sufficient to search items that are descendents of the abstraction previously selected.

One of the problems to be considered here is whether concrete base analogs are hierarchically organized. Many agree with the hierarchical organization of the common natural kinds, but how about knowledge structures used in analogy?

Memory organization is found in more complex materials such as stories and episodes. Although there are controversies, some researchers showed that the story grammar type of knowledge structure constrains encoding and retrieval of stories.

The second line of evidence comes from Shank's Mops and TOPs type of knowledge organization. Reflecting Black, Bower, & Turner's experiment, Schank elaborated his theory of scripts to include more abstract knowledge structures. According to him, there are knowledge structures that hierarchically organize concrete representations of specific events. They are called, MOPs, metaMOPs, universal MOPS. In addition, he assumed a different kind of structures that organize thematically similar events, and he called it TOPs (thematic organization packets). He believed that these best explain cross-contextual reminding.

The third line of evidence comes from Fukuda's work. In his experiments, subjects' reminding was greatly improved when they were given cues at the moderately abstract level, compared with when given very similar stories as cues. The superiority of such a cue strongly supports the idea that there exist abstractions and that concrete episodes are organized around the abstraction.

Mapping

In the mapping process, abstractions make two contributions, both of which reduce the computational costs involved in mapping. The first one is concerned with the selection of candidate elements to be mapped. Suppose that a base has n elements. The number of the candidate sets to be mapped amounts to $2^n -1$. This obviously causes combinatorial explosion if n is getting larger.

However, if an abstraction is involved in mapping, one need not suffer from it. It is because what is true for the abstraction must be true for its subordinate target. It follows that every element in the abstraction can be, and should be, mapped. Although one still has to decide which element in the base correspond to which element in the target, the computational gain in the selection of a candidate set is very large.

The second benefit is concerned with the number of elements in the abstraction. The number of

elements in an abstraction is, by definition, smaller than that of its subordinate, concrete base analogs. It is impossible to make a general estimation of how much smaller the elements in the abstraction is, but the reduction in the number of elements produces huge computational gain in many cases.
For example, if one maps *n* elements of the base to the target, the resulting number of possible mappings is the permutation of *n*, shown by the thick line in the graph. As you see, it is approximated by an exponential function. Suppose that an abstraction has a half of the elements. The number of candidate hypotheses is depicted by the broken line (the dotted line shows the number of hypotheses when the number of the base element is reduced to a quarter of *n*). Although the number of possible mapping hypotheses is exponential even assuming the abstraction, the computational gain is huge compared with the cases without abstractions.

Abstractions in analogical reasoning about electrical circuit

Informal observation

This is the stage for the present framework to be more concrete. My favorite example is people's natural reasoning about the electric circuit. As Gentner & Gentner (1983) reported what type of base analog is used affects subjects prediction about the behaviors of the circuit. When a water flow system was introduced, subjects correctly infer the change of the electricity when a battery is added serially. On the other hand, subjects' prediction improved in the case of parallel resistance, when a teaming crowd analogy was taught.
In the experiment, they gave subjects either analog explicitly, and asked them to use it when answering the problems. However, people can draw analogies spontaneously even without such instruction. From my observation, most university students used liquid flow analogies initially, although they were not exactly the same as the water flow, as I will show you later.
Such naturally drawn analogies tell us many things. First of all, although most subjects used a kind of liquid-flow analogy, it is very dubious that their analogies were based on a specific experience about a water flow system. It is hard to imagine that they had seen water flowing in the closed circuit with a pump, even harder to imagine they had seen a parallel circuit with two pumps attached serially! If they did not have any experience with it, how could they make analogies? This shows that in naturally drawn analogies, the possibility of making use of very concrete, episode type of base analog is very low.
Second, the mapping was very immediate, so immediate that they seemed not to be in trouble with candidate mapping hypotheses. From my observation, no single case was found that they made mistake in finding correspondence. Essential parts in the base and target were immediately mapped, while non-essential parts seemed not to be even for a slightest consideration. The protocol shows no statements such as pump's having a lever or a switch, pumps' needs of external forces, although they play causal roles in the actual water flow system.
Third, we observed the on-line construction of a base. When subjects were asked to estimate heating values at resistance, many subjects spontaneously and naturally switched the source analog from the liquid-flow to the particle-flow. That is, they changed the flowing entity from liquid to something solid, such as people, small stones, or particles. The shift seems to be done because water was judged not to be a relevant analog for the generation of heat. These solid objects, instead, enable people to naturally infer the generation of heat by the friction of contacting parts.

Flowing system abstraction

The picture drawn by the observation is quite different from the ones that the current models of analogy do. Despite of the unavailability of concrete base analogs, people had little difficulties in reasoning analogically about the behaviors of the electric circuit. This fact suggests that an abstraction, a flowing system, is responsible for subjects' analogical reasoning. This abstraction is very simple in the sense that it consists of only three components: a flowing entity, path, and force. A typical relation between them is that the force causes the entity to flow through the path.

The simplicity of the abstraction partly explains the immediate mapping. Since there are only three components that are distinct, and every component of the abstraction is applied by definition, there are little possibilities for misunderstanding the mapping relations.

The flowing system abstraction is a higher-order abstraction, in the sense that every component is variable. Thus, it must be supplemented and enriched by contextual information involved in the problem situation, when it is actually used. This enables abstractions to be flexible. Even when people cannot access to a concrete base analog, they can naturally make useful inferences, by instantiating the abstractions under the constraints posed by the problem situation.

Furthermore, these characteristics explain the on-line construction of a new base analog. As I reported earlier, subjects could easily shift from the liquid-flow to the particle flow analog by changing the flowing entity when they dealt with the generation of heat. The ease of the shift cannot be explained, without assuming the flowing system abstraction. If people had used an actual water flow system as a base analog, the shift should not have been done so easily. This is because people have to retrieve a new analog by examining all the candidate analogs again, and they have to replace every component of the analog with new ones: a water pump with a loud speaker, a pipe with a road, etc. However, if one assumes the abstraction, the search for a new analog is constrained. Furthermore, it is enough to change one of the components of the abstraction, because the existence of the pushing force and the path is guaranteed by the abstraction.

In addition, the flowing system abstraction provides the global coherence when changing the analogs. If one uses a completely new analog, there is a possibility of inconsistency between what have been inferred and what will be inferred. On the other hand, inferences based on old and new analogs are consistent if they are descendents of the same abstraction. In the case of the electric circuit analogy, inferences based on the liquid flow analog are guaranteed to be consistent with those based on the particle flow analog.

Some researchers have emphasized the process of adaptation in analogy. Since there are few problems that a base analog can directly be applied, it is often necessary to adapt analogs to the current problem situation. In order for flexible adaptation, it would be better that source analogs are small and simple, like the flowing system abstraction. It is difficult to modify and adapt big, deep, and complex analogs that contains a lot of information.

Contrasting abstraction-based view with current theories of analogy

The framework proposed here contrasts sharply with that of the dominant theories of analogy. According to the dominant view, episodes are represented almost literally in the form of first-order predicate logic. Since no abstraction or summarization is assumed to take place in encoding source episodes, each source episode forms a large, deep, complex structure. In addition, each analog is stored in a relatively isolated fashion. Thus, some assume only surface level matches (Forbus et al., 1995), while others can only make use of word-to-word level relations (Thagard et al., 1990). In mapping, many theories share the assumption that initial mapping is carried out syntactically. Since this type of mapping generates a large number of mapping hypotheses, one or more constraints are called for to reduce them (Falkenhainer et al., 1989; Holyoak & Thagard, 1989). The mapped structure is static and isolated in the sense that it is prestored in the source analog and has few relations to other analogs. Thus, when shifting a source, an analogizer has to reiterate the entire processes.

On the other hand, the abstraction-based view of analogy assumes small, simple abstracted mental entities as source analogs. A small number of variabilized components are involved in abstractions. Each source abstraction is connected to form a hierarchy. In mapping, variable bindings or unification take place, in a deductive fashion. Since an abstraction involves a small number of distinct elements, the number of possible mapping hypotheses is small. The resulting structure is liable to modification under the constraints posed by the target analog and task goal. In this sense, analogy by abstraction is dynamic and constructive.

The findings obtained from the informal observation of people's spontaneous analogical reasoning are not compatible with the dominant view. First, there seem to be no large, complex source analogs available. Second, people retrieved the source analog very rapidly. It seemed that only a limited number of candidate analogs were in consideration. This suggests that subjects may make use of the hierarchical information in retrieval. Third, mapping was rapidly carried out without mistakes. This suggests that they did not suffer from a large number of mapping hypotheses, which in turn leads us to the idea that a source analog actually used did not have a large, complex structure. Finally, subjects shifted from one source to another flexibly and naturally, by changing a part of the source analog. It would have taken relatively long time if they had replaced the original analog with a completely new one. This indicates that they did not use an analog representing the actual water flow system.

These findings are best explained by the abstraction-based view of analogy, which assumes small, simple variabilized mental entities connected hierarchically.

Relations to other kinds of cognition

A number of researchers have explored the processes and structures of analogy for many years. They have revealed what subprocesses are involved in analogical reasoning, what affects human analogy making, as well as how and where analogies are used. These findings lead to computational theories of analogy. By their competition, the levels of analysis have been greatly improved, which in turn leads to greater sophistication of the theories.

However, the relationships of analogy to other kinds of cognition have been missed in the course of the scientific endeavor. Analogy plays a central role in human cognition, but it seems strange that there is a cognitive engine designed specifically for making analogies. It might be that analogy is a special combination of more basic cognitive components. If so, we should explore the relationships of analogy to other kinds of cognition.

The proposed framework opens the door of analogy to other kinds of cognition. In this section, I briefly review the relationship of analogy to categorization and deduction.

Categorization

One important relation is to categorization. Although abstractions accessed in the course of analogical reasoning are different from common categories, the underlying mechanisms are same. Both assume the hierarchical structure and the inheritance of properties.

Certainly, dominant models of categorization seem to be a little bit too simplistic, because they do not have principled methods distinguishing structural and surface information. Thus, as Ramscar and Pain (1996) pointed out, the model of categorization should be modified and enriched by the findings obtained from analogy research.

Deduction

A striking finding provided by the framework is that analogy is similar to deduction. My proposal is follows: given a target is a member of an abstraction, and that abstraction has a property X, then the target has the property X. This form of reasoning is properly said to be a categorical syllogism. We explain why some analogies seem to be psychologically valid. This is because they are deduction.
However, I do not intend to reduce analogy to deduction. My position is opposite. From my viewpoint, deduction is a kind of analogy. Abstractions used in the processes of analogy do not have the same status as premises in deduction. People may induce a wrong abstraction in some cases, while they may access to a wrong abstraction in other cases. Thus, there exists uncertainty in analogical reasoning. On the other hand, categories appeared in deduction are fixed, and proved to be relevant. Thus, no ambiguity is found in deduction. If you admit the discussion above, you will notice that deduction is a special case of analogy, not vice versa.

Acknowledgment

Many thanks to Bipin Indurkhya and Mark Keane for their helpful comments on the talk based on the earlier version of the paper.

References

Chi, M. T. H., Bassok, M., Lewis, M. W., Reiman, P. & Glaser, R. (1989) Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Psychology*, **13**, 145 - 182.

Falkenhainer, B., Forbus, K. D. & Gentner, D. (1989) Structure mapping engine: Algorithm and examples. *Artificial Intelligence*, **41**, 1 – 63.

Forbus, K. D., Gentner, D., & Law, K. (1995) MAC/FAC: A model of similarity-based retrieval. *Cognitive Psychology*, **19**, 144 – 205.

Gick, M. L. & Holyoak, K. J. (1983) Schema induction and analogical transfer. *Cognitive Psychology*, **14,** 1 - 38.

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

Gentner, D. & Gentner, D. R. (1983) Flowing waters or teaming crowds: Mental models of electricity. In D. Gentner & A. L. Stevens (Eds.) *Mental Models*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gentner, D. & Jeziorski, M. (1993) The shift from metaphor to analogy in Western science. In A. Ortony (Ed.) *Metaphor and Thought.* Cambridge, UK: Cambridge University Press.

Glucksberg, S. & Keysar, B. (1990) Understanding metaphorical comparisons: Beyond similarity. *Psychological Review*, **97**, 3-18.

Greiner, R. (1988) Learning by understanding analogy. *Artificial Intelligence*, **35**, 81-125.
Goswami, U. & Brown, A. L. (1989) Melting chocolate and melting snowmen: Analogical reasoning and causal relations. *Cognition*, **35**, 69 – 95.
Holyoak, K. J. & Thagard, P. (1989) Analogical mapping by constraint satisfaction. *Cognitive Science*, **13**, 295 – 355.
Holyoak, K. J. & Thagard, P. (1995) *Mental Leaps: Analogy in Creative Thought*. Cambridge, MA: MIT Press.
Indurkhya, B. (1992) *Metaphor and Cognition: An Interactionist Approach.* Dordrecht, Netherlands: Kluwer Academic Publishers.
Kedar-Cabelli, S. (1985) Purpose-directed analogy. In Proceedings of the Seventh Annual Conference of the Cognitive Science Society, 150 – 159.
Lakoff, G. (1993) The contemporary theory of metaphor. In A. Ortony (Ed.) *Metaphor and Thought.* Cambridge, UK: Cambridge University Press.
Ramscar, M. & Pain, H. (1996) Can a real distinction be made between cognitive theories of analogy and categorisation? In Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society. 346 - 351.
Russell, S. W. (1988) Analogy by similarity. In D. H. Helman (Ed.) *Analogical Reasoning: Perspectives of Artificial Intelligence, Cognitive Science, and Philosophy.* Dordrecht, Netherlands: Kluwer Academic Publishers.
Thagard, P., Holyoak, K. J., Nelson, G., & Gochfeld, D. (1990) Analog retrieval by constraint satisfaction. *Artificial Intelligence*, **46**, 259 – 310.