

Structural Constraints and Real-World Plausibility in Analogical Inference

Linsey A. Smith (linsey@u.northwestern.edu)

Dedre Gentner (gentner@northwestern.edu)

Department of Psychology, Northwestern University

2029 Sheridan Road, Evanston, IL 60208 USA

Abstract

Theoretical accounts of analogy have largely agreed that structural constraints play a substantial role in the mapping process. Less is known, however, about the robustness of these constraints in the inference process and the way in which particular content influences the use of structural constraints in analogical inference. We conducted two studies testing whether the plausibility (or implausibility) of an inference influences adherence to general structural principles in analogical reasoning. We found substantial reliance on the predicted structural constraints, but also an influence of the plausibility of the inference.

Introduction

Our goal in this research is to explore the stability of analogical inference under different conditions: specifically, whether analogical inference is a domain-general reasoning process, governed by structural constraints inherent to the analogical process, or whether it is a loosely constrained process whose outcome is strongly influenced by the plausibility of the potential inferences in particular domains. This question is important not only for what it can tell us about basic analogy processes, but also because the use of analogy in scientific discovery (and even in science learning) sometimes requires making initially implausible inferences. We first review research on this issue in the arena of analogical mapping and alignment, which has been extensively studied, and then turn to analogical inference.

Structural Constraints on Analogical Mapping

Reasoning by analogy involves identifying a common system of relations between two domains and generating further inferences driven by these commonalities (Gentner, 1983; Holyoak & Thagard, 1989; Hummel & Holyoak, 1997; Kokinov & French, 2003). According to structure-mapping theory, the comparison process involves aligning a pair in such a way as to achieve a consistent structural alignment between two domains (Falkenhainer, Forbus & Gentner, 1989; Gentner, 1983; Gentner & Markman, 1997). The structural alignment process is guided by a set of tacit constraints that lead to structural consistency and inferential clarity: *parallel connectivity*, which requires that arguments of matching predicates must also be placed into correspondence; and *one-to-one correspondence*, which requires that each element of a representation match, at most, one element of the other representation. Importantly,

deep matching systems are preferred over shallow matches (the *systematicity principle*), which reflects a preference for coherence and inductive power in analogical processing (Clement & Gentner, 1991; Falkenhainer, Forbus & Gentner, 1989). Candidate inferences are generated by completing the pattern in the (initially) less-structured member of the pair, based on the common structure.

Models of analogy have largely converged on a set of assumptions like those outlined above (Falkenhainer, Forbus & Gentner, 1989; Gentner, Holyoak & Kokinov, 2001; Holyoak and Thagard, 1989; Hummel & Holyoak, 1997; Kokinov & French, 2003; Larkey & Love, 2003). Further, there is substantial empirical evidence in support of the idea that analogical reasoning obeys these constraints. A variety of studies have provided evidence that analogical matching is constrained by both structural consistency (including one-to-one mapping) (e.g., Krawczyk, Holyoak, & Hummel, 2005; Markman, 1997; Markman & Gentner, 1993; Spellman & Holyoak, 1992) and systematicity (e.g., Clement & Gentner, 1991). For example, Clement and Gentner (1991) showed participants analogous scenarios and asked them to judge which of two lower-order assertions shared by the base and target was most important to the match. Participants chose the assertion that was connected to matching causal antecedents – their choice was based not only on the goodness of the local match, but also on whether it was connected to the larger matching system. Thus, matching lower-order relations that are interconnected by higher-order relations were considered more important to the analogy. In sum, people demonstrate considerable structural sensitivity in analogical mapping.

Analogical Inference

There is some research on the degree to which structural constraints hold in analogical inference. In the Clement and Gentner (1991) research just described, a second study found evidence for systematicity in inference projection. People generated inferences that were part of a shared system, rather than equally applicable inferences that were not. Markman (1997) also found evidence for systematicity in inference generation. In addition, he found that people based their inferences on one-to-one mappings. When given analogies with two possible sets of correspondences, people noticed both possibilities, but drew inferences from only one of them. These findings suggest a role for structural consistency in inference, as in alignment.

However, one question that is largely unexplored is the degree to which the analogical inference process is influenced by the factual plausibility of the inference in the target. That is, are people able to track structural consistency despite implausibility in making inferences? The studies described above did not involve wide variations in plausibility, so they do not answer this question. Work by Keane (1996) does bear on this issue. He found that people readily accepted inferences that were both highly plausible [had high “entity utility”] and easy to place into correspondence with the target [“entity parallelism”]—that is, highly *adaptable*—compared to those inferences that were less adaptable. These findings suggest that plausibility in the target is important in analogical inference. However, the question remains open as to what people will do if structural consistency directly conflicts with target plausibility.

Another way to put this question is, are there *content effects* in analogical inference? The issue of content effects has been investigated extensively in the research on deductive reasoning. Deductive reasoning has traditionally been considered a relatively rigorous, principle-governed process, although empirical support for this claim (e.g., Marcus & Rips, 1979) is punctuated by many observations that show that people’s judgments about the logical validity of deductive arguments is influenced by the 1) specific content that is being reasoned about (e.g., Cheng & Holyoak, 1985; Cummins, Lubart, Alksnis, & Rist, 1991; Rips, 2001; Thompson, 1994), and 2) whether the reasoner agrees with the premises and conclusions of the argument (e.g., Markovits, 1995; Newstead, Pollard, Evans, & Allen, 1992). Thus, there is evidence that logical reasoning is swayed by particular content.

a. *Logically valid, real-world plausible:*

If Fred sprinkles water on wood shavings, the shavings get wet.

Fred sprinkles water on wood shavings.

The shavings get wet.

b. *Logically invalid, real-world plausible:*

Fred sprinkles water on wood shavings.

The shavings get wet.

For example, Rips (2001) asked participants to evaluate arguments like (a) and (b) in which the plausible conclusion was either logically valid or invalid. The question was whether people could track deductive logic regardless of the plausibility of the conclusion. A substantial number of participants (mistakenly) identified invalid arguments as logically correct when they were plausible. Overall, Rips’s (2001) findings suggest that people were largely able to maintain logical rigor under the strain of real-world implausibility, but that logical rigor was sometimes compromised by the content of the arguments: people could not wholly divorce logical form from content in this task.

A parallel question can be asked about analogical inference: can people maintain structural consistency despite

real-world implausibility in making analogical inferences (which we will refer to as *analogical rigor*)? Our question in this paper is what happens when the structural alignment process leads to inferences that the reasoner considers implausible. On the one hand, some prior research shows reliable effects of structural consistency on inference (Clement & Gentner, 1991; Markman, 1997). On the other hand, these studies (and Keane’s (1996) study) did not directly pit structural consistency against plausibility. And unlike deductive reasoning, analogical reasoning is generally not explicitly taught. Thus we might expect people to be less committed to maintaining analogical rigor than they are to maintaining logical rigor.

The Current Experiments

In this set of studies, we asked participants to evaluate analogies where the inferences derived from the structure-mapping process are at odds with the real-world plausibility of the inferences. This method allowed us to identify how much people rely on domain-specific content over general mapping principles in analogical inference.

For the task, we adapted the deductive reasoning task from Rips (2001). As discussed above, in that experiment, participants evaluated the validity of conclusions from arguments that orthogonally varied in logical validity and real-world plausibility. His study assessed whether people would follow deductive logic in drawing conclusions even when these conclusions conflicted with plausibility. In this research, we posed the parallel question for analogical inference, that is, would people respect the structural constraints of analogy in drawing inferences even when these inferences conflicted with real-world plausibility. To put it another way, are people able to maintain analogical rigor in the face of real-world implausibility? We asked participants to assess whether a particular inference followed from an analogy. We created materials whose inferences varied in *structural consistency*, that is, we varied whether the inference was a structurally consistent completion of the analogy. Table 1 shows an example set. The inferences in (a) and (b) are structurally consistent and those in (c) and (d) are structurally inconsistent. The pairs also varied orthogonally in *real-world plausibility*, with (a) and (c) having plausible inferences and (b) and (d) having implausible inferences. Participants might find analogies (b) and (d) (both implausible inferences) to be odd or downright wrong, but this is precisely the point: when an analogical inference conflicts with reasoners’ knowledge, the question is whether they can identify inferences that the analogy must structurally yield, without being swayed by the plausibility of those inferences.

Of course, the ultimate evaluation of an analogical inference is not solely contingent on structural consistency, but also involves checking the factual validity of the inference (and in a real problem solving situation, the contextual relevance) (Gentner & Clement, 1988; Holyoak & Thagard, 1989). To this end, we also asked participants to provide ratings of the *overall goodness* of each analogy. We

Table 1: Sample materials from Experiment 1.

| |
|---|
| Base (constant) |
| Mary has built a sandcastle. Her younger brother comes by and kicks the base of the castle. The sandcastle crumbles. |
| Target (four versions) |
| <i>a. Structurally consistent, factually plausible</i> A wrecking ball knocks into a building's foundation. Conclusion: The building comes crashing to the ground. |
| <i>b. Structurally consistent, factually implausible</i> A tennis ball knocks into a building's foundation. Conclusion: The building comes crashing to the ground. |
| <i>c. Structurally inconsistent, factually plausible</i> A tennis ball knocks into a building's foundation. Conclusion: The building stays standing. |
| <i>d. Structurally inconsistent, factually implausible</i> A wrecking ball knocks into a building's foundation. Conclusion: The building stays standing. |

had two goals with this question. First, for implausible inferences, this question would give participants a way to indicate that they considered some analogies to be quite poor. We hoped that this would leave them more free to judge structural consistency on its own. Second, a more direct goal was to discover whether participants would incorporate both structural consistency and real-world plausibility into their judgments, as we expected they would. If so, we would expect only analogies that yield structurally consistent and plausible inferences to receive high overall goodness ratings.

Experiment 1

Method

Participants 19 Northwestern University undergraduates took part in the study individually or in small groups of up to four people. Participants completed the task in 10-15 minutes and for their time they received credit towards a course requirement or monetary compensation.

Procedure and Materials The experimenter gave one task booklet to the participant, and upon completion they returned the booklet to the experimenter. The booklet contained a page of instructions, followed by eight analogies (one per page). The analogies came from quartets of items, as in Table 1, that varied in structural consistency and real-world plausibility. We assigned each participant eight analogies, two of each type (structurally consistent and real-world plausible, structurally consistent and implausible, structurally inconsistent and plausible, structurally

inconsistent and implausible), as in Table 1. For an individual participant, however, different content instantiated each of these arguments. Thus, for example, no participant received more than one pair from the Table 1 quartet. The order of the problems in the test booklet was pseudo-randomized into four orders.

Measures Participants rated their agreement with the statement "The conclusion follows directly from the analogy." Responses were measured on a 7-point Likert scale, ranging from 1 (strongly disagree) to 7 (strongly agree). To facilitate analysis, responses were recoded into a dichotomous variable (with responses > 4 recoded as "Yes, the conclusion follows" and ≤ 4 recoded as "No, the conclusion does not follow"). The proportion of "Yes" responses for each type of stimuli was the measure of interest, and these were aggregated within conditions to form a measure of *inference acceptance rates*, which we'll simply refer to as *acceptance rates*. To the extent that participants strongly differentiate structurally consistent from inconsistent inferences, such that structurally consistent inferences have high acceptance rates and structurally inconsistent inferences have low acceptance rates, this measure will approximate analogical rigor.

In addition participants were asked to judge the *overall goodness* of each analogy. Participants rated their agreement with the statement "Overall, this is a good analogy." Responses were measured on a 7-point Likert scale, ranging from 1 (strongly disagree) to 7 (strongly agree).

Results

Figure 1 presents the inference acceptance rates for each of the four types of stimuli. The data were analyzed with a two-way ANOVA, with structural consistency and real-world plausibility as within-subjects factors.

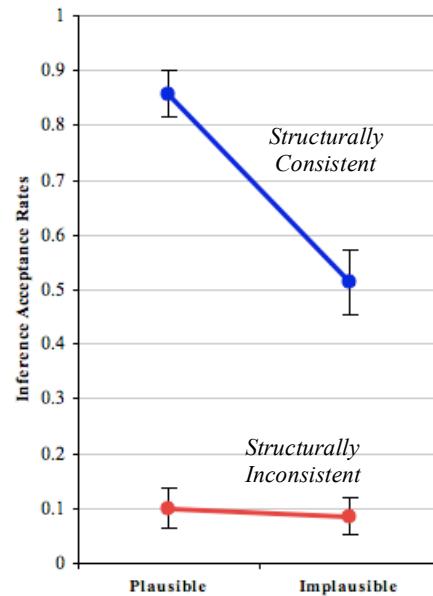


Figure 1: Inference acceptance ratings for Experiment 1.
Error bars reflect the standard error.

Overall, there was a strong effect of structural consistency on acceptance rates, $F(1,37) = 110.87, p < .001, \eta^2 = .38$; people were far more likely to accept structurally consistent inferences ($M=.63, SD=.49$) than structurally inconsistent inferences ($M=.09, SD=.29$). There was also a main effect of real-world plausibility on acceptance ratings, $F(1,37) = 8.74, p < .01, \eta^2 = .05$; a greater proportion of plausible inferences was judged as following from the analogy ($M=.45, SD=.50$) than implausible inferences ($M=.30, SD=.46$). The effect size for real-world plausibility was considerably smaller ($\eta^2 = .05$) than that for structural consistency ($\eta^2 = .38$).

There was also a significant interaction between structural consistency and plausibility, $F(1,37) = 27.89, p < .001, \eta^2 = .10$. For structurally consistent analogies, participants were less likely to judge implausible inferences as following from the analogy (implausible: $M=.50, SD=.51$; plausible: $M=.89, SD=.31$), $t(37) = 4.09, p < .001$. No such difference was obtained for structurally inconsistent analogies.

We reserve the analysis of overall goodness judgments until after we present Experiment 2.

Discussion

Our primary question is whether people can maintain analogical rigor in the face of real-world implausibility. We found fairly good support for this possibility. Acceptance ratings were higher overall for structurally consistent analogies, indicating that people are able to track the structural consistency of an inference regardless of the plausibility of that inference. Additional support for this claim comes from the observed effect sizes: structural consistency explains 38% of the overall variance on inference acceptance rates, whereas real-world plausibility only accounts for 5% of the variance. However, analogical rigor is also influenced by particular content. Specifically, participants were more likely to reject structurally consistent inferences when they were implausible. If individuals had been entirely rigorous, we would not have expected to see this difference between plausible and implausible conditions. Interestingly, this effect of plausibility did not appear for structurally inconsistent inferences, which were uniformly rejected.

In short, the results so far suggest that people are able to abide by structural constraints when making inferences; however, conflicting content can influence whether people maintain these constraints. In the next study, we sought to identify whether clarifying the instructions would attenuate these content effects.

Experiment 2

This study tested whether more explicit instructions would lead participants to more strictly observe analogical constraints. We used the same basic method as Experiment 1, with one important modification: we re-wrote the question to clarify that the focus should be on what follows from the analogy.

Method

Participants 19 Northwestern University undergraduates took part in the study individually or in small groups of up to four people. Participants completed the task in 10-15 minutes and for their time they received credit towards a course requirement or monetary compensation.

Materials and Measures The materials for the analogy task were the same, except that the question used to elicit inference acceptance ratings was modified from rating agreement with the statement “The conclusion follows from the analogy?” to instead read “The conclusion in Situation 2 would necessarily follow if Situations 1 and 2 were truly analogous, regardless of whether the conclusion *could* be true or not.” Participants were then asked to circle “Yes” or “No.” The proportion of “Yes” responses for each type of stimuli was the dependent measure, and these were aggregated within conditions to form a measure of *inference acceptance rates*. The *overall goodness* question remained the same. The procedure was as in Experiment 1.

Results

The results showed a strong effect of structural consistency; structurally consistent inferences had higher acceptance rates ($M=.91, SD=.29$) than did structurally inconsistent inferences ($M=.12, SD=.33$). Figure 2 shows the inference acceptance rates for each of the four types of stimuli. For ease of comparison, the results from Experiment 1 (dotted lines) have also been included. Analysis entailed a two-way within-subjects ANOVA, with structural consistency and real-world plausibility as within-subjects factors.

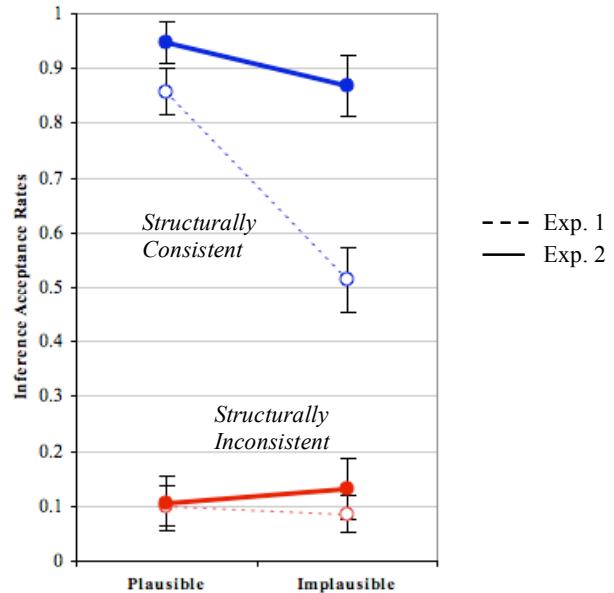


Figure 2: Inference acceptance ratings for Exp. 1 (dotted line) and Exp. 2 (solid), divided into structurally consistent and inconsistent. Error bars reflect the standard error.

As in Experiment 1, there was a main effect of structural consistency, $F(1,37) = 311.22, p < .001, \eta^2 = .71$. Real-world plausibility no longer influenced inference acceptance: there was no main effect of real-world plausibility nor an interaction between the factors (real-world plausible: $M=.53, SD=.50$; implausible: $M=.50, SD=.50$).

Cross-Experiment Analysis To further test whether more explicit instructions to focus solely on whether an inference follows from the analogy bolstered participants' focus on structural constraints, we entered Experiments 1 and 2 into a three-way mixed ANOVA, adding in instruction type (i.e., Experiment 1 or 2) as a between-subjects factor. In addition to the main effects of structural consistency and real-world plausibility, there was also a main effect of instruction type, $F(1,74) = 6.26, p < .05$. These main effects were qualified by a significant three-way interaction between all three variables, $F(1,74) = 5.31, p < .05$. This significant interaction is due to different patterns of performance on structurally consistent inferences: in the explicit instructions condition (Experiment 2), there was no difference in acceptance rates between plausible and implausible inferences, but in the implicit instructions condition (Experiment 1), acceptance rates were higher for plausible inferences, $t(37) = 4.09, p < .001$.

Judgments of overall goodness We elicited judgments of overall goodness for the analogies to identify participants' overall impression of the analogy, which may not have been captured in the acceptance rates, especially in the case of implausible inferences. To identify whether judgments of overall goodness for the analogies varied by instruction type, we entered both experiments into a three-way mixed ANOVA, with *overall goodness* as the dependent measure. There was only a marginally nonsignificant effect of instruction type, $F(1,74) = 3.33, p = .07$; participants rated overall goodness similarly across both instruction conditions. There were main effects of both structural consistency ($F(1,74) = 97.35, p < .001, \eta^2 = .27$) and real-world plausibility ($F(1,74) = 28.43, p < .001, \eta^2 = .06$), which were qualified by a significant interaction between the two, $F(1,74) = 43.02, p < .001, \eta^2 = .11$. Structurally inconsistent pairs were given low overall ratings that did not vary by real-world plausibility (max = 7, plausible: $M=1.92, SD=1.16$; implausible: $M=2.20, SD=1.77$); structurally consistent pairs that were plausible were given higher ratings than implausible pairs (plausible: $M=5.05, SD=1.52$; implausible, $M=2.91, SD=1.86$), $t(75) = 8.25, p < .001$. This pattern of goodness ratings partly mirrors the pattern of inference acceptance ratings in Experiment 1: there was an effect of both structural consistency and plausibility, with a stronger effect of structural consistency; and structurally consistent analogies were rated lower when their inferences were implausible. Thus, with the exception of the Experiment 2 acceptance ratings, the deviation from analogically rigorous behavior occurs only for structurally consistent but implausible analogies.

Discussion

Our primary question in Experiment 2 was whether people are capable of separating structural consistency from real-world plausibility when explicitly told to do so. The results indicate that the answer is yes: people were able to ignore the real-world plausibility of analogical inferences in making their judgments.

General Discussion

Two studies probed the robustness of structural constraints on analogical inference when challenged by the particular content of the inferences. In Experiment 1, we investigated whether people would follow the structural constraints of analogy in drawing inferences even when they conflicted with plausibility. Acceptance rates were higher for structurally consistent inferences than inconsistent inferences; overall, people can reliably follow structural consistency in inference. Plausibility did influence inference acceptance rates, but only for structurally consistent analogies. Structurally inconsistent inferences were noticed as such, regardless of their real-world plausibility. However, when people encountered potentially analogous (i.e., structurally consistent) inferences, their judgments were influenced by target plausibility.

Experiment 2 tested whether more explicit instructions would lead participants to make a clearer separation between analogical rigor and plausibility. The results indicate that this is indeed the case: participants no longer demonstrated content effects, but instead recognized inferences that followed from completing the common system, as predicted by structure-mapping and other current models of analogy (Falkenhainer, Forbus & Gentner, 1989; Holyoak and Thagard, 1989; Hummel & Holyoak, 1997; Kokinov & French, 2003). Understanding the conditions under which people will put aside their knowledge to work through an analogy has implications for educational contexts, where analogies are used extensively to promote knowledge acquisition and conceptual change (e.g., Richland, Holyoak, & Stigler, 2004). Importantly, the analogies used by instructors may require learners to make ostensibly implausible inferences (e.g., Clement, 1993).

In both experiments, we elicited judgments of overall goodness of the analogies. We found, as expected, that people considered both structural consistency and real-world plausibility in judging the analogies. Ratings for overall goodness did not vary as a function of instructions. In both experiments, people reliably indicated that only those analogies that were both structurally consistent and real-world plausible were good analogies. This pattern of judgments is in accord with the general assumption that while analogy may involve a mapping process guided by structural constraints, ultimate evaluation of the analogy involves checking the factual validity of projected inferences.

Although Experiment 1 demonstrates that analogical rigor is influenced by content, for both experiments, participants showed a general tendency to identify structurally consistent inferences as following from the analogy. Furthermore,

effect sizes were moderate for structural consistency, whereas they were extremely small for plausibility. Perhaps more tellingly, in judgments of overall goodness, the effect of structural consistency was much larger ($\eta^2=.27$) than that of plausibility ($\eta^2=.06$). Taken together, these observations suggest that people are relying heavily on structural principles to guide their evaluations of overall analogical goodness. The results of these experiments are consistent with the claim that analogical processing involves a structure-mapping process of alignment and inference largely governed by structural constraints.

One concern here is that the materials were too simple to engage serious content-based reasoning. It will be necessary to investigate a wider range of material to determine the whether the effects identified in these studies will generalize to more natural materials. However, the results so far suggest that analogical inference is to a large extent guided by a tacit set of structural constraints that may function something like the principles that guide deductive reasoning. In future studies it would be of interest to contrast these two reasoning tasks to see whether similar patterns emerge. Another future direction would be to obtain online measures, such as reading times, to investigate the time course of content effects in analogy and further explicate the interaction between mapping processes and target content in analogical inference.

Acknowledgments

This research was supported by ONR Grant N00014-08-1-0040. We thank Doug Medin, in addition to Julie Colhoun and the rest of the Cognition and Language Lab, for many helpful comments and suggestions on this work. We also thank Katherine James, Ted Bedell, Noel Dwyer, and Lindsey Zamarripa who helped with recruitment, scheduling, and data collection.

References

Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, 17, 391–416.

Clement, J. (1993). Using bridging analogies and anchoring intuitions to deal with students' preconceptions in physics. *Journal of Research in Science Teaching*, 30, 1241-1257.

Clement, C. A., & Gentner, D. (1991). Systematicity as a selection constraint in analogical mapping. *Cognitive Science*, 15, 89-132.

Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991) Conditional reasoning and causation. *Memory & Cognition*, 19(3), 274-282.

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

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

Gentner, D., Holyoak, K. J., & Kokinov, B. (Eds.). (2001). *The analogical mind: Perspectives from cognitive science*. Cambridge, MA: MIT Press.

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist*, 52, 45–56.

Gentner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive Science*, 10, 277-300.

Holyoak, K. J. & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science*, 13, 295–355.

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

Keane, M.T. (1996). On adaptation in analogy: Tests of pragmatic importance and adaptability in analogical problem solving. *Quarterly Journal of Experimental Psychology*, 46, 1062-1085.

Kokinov, B., & French, R. M. (2003). Computational models of analogy making. In L. Nadel (Ed.), *Encyclopedia of Cognitive Science*. London: MacMillan.

Krawczyk, D. C., Holyoak, K. J., & Hummel, J. E. (2005). The one-to-one constraint in analogical mapping and inference. *Cognitive Science*, 29, 797 – 806.

Larkey, L. B., & Love, B. C. (2003). CAB: Connectionist Analogy Builder. *Cognitive Science*, 27 (5), 781-794.

Marcus, S. L., & Rips, L. J. (1979). Conditional Reasoning. *Journal of Verbal Learning and Verbal Behavior*, 18(22), 199-223.

Markman, A. B. (1997). Constraints on analogical inference. *Cognitive Science*, 21, 373–418.

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

Markovits, H. (1995). Conditional reasoning with false premises: Fantasy and information retrieval. *British Journal of Developmental Psychology*, 13, 1-11.

Newstead, S. E., Pollard, P., Evans, J. S., & Allen, J. L. (1992) The source of belief bias effects in syllogistic reasoning. *Cognition*, 46(1), 87-92.

Reeves, L. M., & Weisberg, R. W. (1994). The role of content and abstract information in analogical transfer. *Psychological Bulletin*, 115(3), 381-400.

Richland, L. E., Holyoak, K. J., & Stigler, J. W. (2004). The role of analogy in teaching middleschool mathematics. *Cognition and Instruction*, 22, 37-60.

Rips, L. J. (2001). Two kinds of reasoning. *Psychological Science*, 12, 129-134.

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 and Social Psychology*, 62(6), 913-933.

Thompson, V. A. (1994). Interpretational factors in conditional reasoning. *Memory and Cognition*, 22(6), 742–758.