A Computational Level Theory of Similarity

Bradley C. Love
Department of Psychology
The University of Texas at Austin
MEZ 330
Austin, TX 78712 USA
love@psy.utexas.edu

Abstract

Why are some pairs of objects (or events) perceived to be more similar to each other than other pairs? A computational level theory of perceived similarity is presented that extends previous geometric and set-theoretic formulations. Like previous approaches, the current account posits that the similarity of two objects is a function of the common and distinctive features of the two objects. Unlike previous approaches, similarity is also a function of higher-order compatibility relations among features (as it is in models of analogy). Objects (or concepts) are represented as directed feature graphs as opposed to feature vectors or sets. Like current accounts of human analogical processing, the approach presented here holds that representational elements are put into correspondence during the comparison processes. Correspondences are chosen in order to maximize an objective function. The function contains four terms that are motivated by theories of human comparison. The maximum of the function is monotonically related to perceived similarity. Thus, similarity is characterized as the byproduct of comparison and structural alignment. The objective function serves as a quantitative computational level theory of human comparison.

Introduction

Since William James (1890/1950), psychologists have held that detecting the "sameness" or similarity of objects is at the backbone of cognition. Clearly, detecting similarities between novel events and previous experiences is crucial in reasoning, analogy, and object recognition. Many theories of category learning hold that similarity is the basis for categorization (see [7]). A fundamental question then is what makes two objects similar?

Almost all accounts of perceived similarity hold that similarity increases as the number of feature matches increases and decreases as the number of feature mismatches increases. In geometric models of similarity, such as multidimensional scaling (MDS) models of similarity, concepts or objects are represented as points in a multidimensional space and similarity is inversely related to the distance between points in the space [20]. Objects that match on many features will be closer together in the space than objects that mismatch on a number of features. Unfortunately, the axioms of metric spaces (e.g., minimality, symmetry, and the triangle inequality) appear to be violated by human similarity judgments (see [22]). More recently, Medin, Goldstone, and Gentner (1993) have demonstrated that an object can be rated as both more similar and more dissimilar to the same object in an object pair, which seems problematic for distance models.

Tversky's (1977) contrast model is a non-metric set-theoretic account of perceived similarity that aims to address some of the shortcoming of the distance models. Tversky's model is based on evaluating sets of matching and mismatching features:

$$\mathbf{sim}(x, y) = \gamma_1 \mathbf{F}(X \cap Y) - \gamma_2 \mathbf{F}(X - Y) - \gamma_3 \mathbf{F}(Y - X) \quad (1)$$
$$\text{where } \gamma_1, \gamma_2, \gamma_3 \geq 0$$

where $\mathbf{sim}(x, y)$ reads "the similarity of x to y," X is the set of features that represents x, Y is the set of features that represents y, $X \cap Y$ is the set of features common to x and y, $X - Y$ is the set of features uniquely possessed by x, $Y - X$ is the set of features uniquely possessed by y, γ_1, γ_2, and γ_3 are free parameters, and **F** is a function over sets of features related to the features' saliency. For simplicity and without loss of generalization, we assume here that all features are equally salient:

$$\mathbf{F}(X) = |X| \quad (2)$$

where $|X|$ denotes the cardinality of the set X. Of course, in many situations certain features will weigh more heavily on the evaluation of similarity than other features.

Tversky's contrast model can account for asymmetries that occur in similarity judgments. For example, "North Korea" is rated as being more similar to "China" than vice versa. The contrast model can explain such asymmetries by setting $\gamma_2 > \gamma_3$. Ostensibly, when comparing x and y, the focus is on first term x, which I will refer to as the *target*, and not on the second term, which I will refer to as the *base*. Both x and y will be referred to as *analogs*. In the example above, most people know more about China than North Korea. Accordingly, when evaluating how similar China is to North Korea $|X - Y|$ will be larger than $|Y - X|$. Another comparison related explanation for asymmetries is that subjects prefer the base to be the object out of an object pair that allows for more analogical inferences to be projected [1]. Such asymmetries may be attributable to the similarity predicate in particular. Another alternative is that asymmetries arise from general principles related to sentence interpretation such as the figure/ground relationship between the target and base [21] or from general syntactic properties [6].

Although the contrast model can address a wide range of data, it cannot account for judgments of similarity that are relational or analogical in nature. Two analogs can be similar

even when the analogs do not have many features in common. To use Gentner's example, one reason people judge the solar system and an atom as being similar is that our representations of these two systems share a number of higher-order relational matches (as opposed to simple feature matches). For example, electrons revolving around a larger nucleus can be put in correspondence with planets revolving around a larger sun. Although the elements of the two systems are not inherently similar (e.g., a nucleus and the sun differ in size and composition), the two analogs are judged to be similar because a mapping between the two systems exists that preserves higher-order commonalities (e.g., the sun maps to the nucleus and the planets map to the electrons). Of course, there are simpler cases of different dimensions being put in correspondence. For example, people equate high-pitched sounds with bright lights and when asked, "Which is brighter, a sneeze or a cough?" people readily answer that a sneeze is brighter [16]. The contrast model assumes that only identical features can match and does not envisage a matching process that attempts to preserve higher-order compatibilities.

More current models of comparison and analogy (e.g., [2, 13, 11]) do establish relational correspondences when comparing objects. These models tend to prefer mappings between analogs when 1) identical features can be mapped to one another, 2) there are higher-order compatibilities and structures replicate in both analogs, 3) the mapping between the two analogs is or almost is one-to-one. Although these constraints are common to all successful models of human comparison, it is not always clear how these constraints are weighted and manifested in models. In other words, different models may adopt widely different matching algorithms (e.g., [2, 11]), but can be quite similar at the computational level of analysis (in the sense of Marr, 1982). It is important to know what the commonalities and differences of competing models are in order to identify the critical issues that deserve empirical investigation.

The goal of this paper is to specify a computational level theory of comparison and similarity that is quantitative, easily understood, and falsifiable. The computational level theory takes the form of a similarity equation consisting of four terms that combine linearly (weighted by parameters). Unlike algorithmic level models where principles are often obscured within the details of the processing mechanisms, principles in the similarity equation appear as separate terms and it is clear how different principles are weighted. Best fitting parameters for a data set are interpretable and clear predictions can be made about how the best fitting parameters should change as task demands change.

Such a theory might make the common ground between models more obvious to the extent that process models conform to the same underlying computational level theory. A successful computational level theory would also make each algorithmic model's contribution to the field clearer. For instance, a model that simply conformed to the computational level theory and made no new predictions would have no

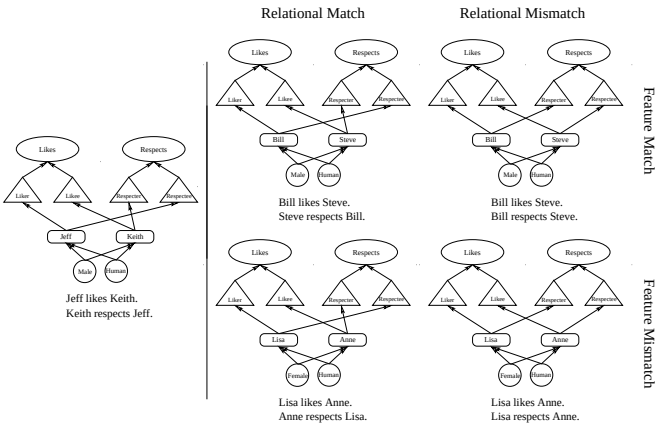

Figure 1: The target analog and its corresponding representation S^x are shown to the left of the dividing line (subjects were not shown S^x, just the text). To the right of the dividing line, an example base analog is shown for each of the four conditions (along with its S^y).

"added value." Many current models do have "added value." For example LISA [11] makes predictions about how working memory limitations and discourse setting affects comparison, MAC/MAC [3] explicitly models retrieval processes in comparison, and IAM [13] focuses on the incremental nature of analogical mapping.

Although theories of comparison hold that similarity and analogical processing are deeply related [5], the linkage between similarity and comparison is even more direct in the similarity equation. In the similarity equation, the correspondences (i.e., mappings) between the two analogs are chosen so as to maximize the similarity equation, much like how an energy function is minimized in a Hopfield network when performing a computation [10]. The maximal value of the similarity equation is monotonically related to the perceived similarity of the two analogs. Thus, similarity both drives the comparison process and is an outcome of it. In the remainder of the paper, the details of the similarity equation are presented as well as some data fits.

Mathematical Formulation

Analogs x and y are represented as directed graphs. Analog x has m different representational elements or nodes, while analog y has n. S^x is an m by m matrix that capture the connectivity of analog x. Each entry in S^x is either 0 or 1. S^x_{24} set to 1 signifies that node 2 binds to node 4. Notice that this relationship is not symmetrical — part 4 is a parent of part 2, but part 2 is not necessarily a parent of part 4. Analog y is represented in an identical fashion by S^y. Figure 1 illustrates some examples of analogs and the graphs that represent them.

In evaluating the similarity of x to y, correspondences are established between the representational elements of x and y (i.e., the nodes in S^x and S^y) . These correspondences or mappings are recorded in the m by n matrix A. Each entry in A is either 0 or 1. A_{ij} equal to 1 indicates that element x_i

(of x) maps to element y_j (of y). The mappings are selected so as to maximize the value of the similarity equation. The idea is that perceived similarity arises out of a comparison process that establishes mappings between the two analogs. Such mappings would prove useful in analogical reasoning and inference.

For real world problems, it is impractical to try all $(\frac{mn}{2})^2$ combinations in search of the best mapping. The theory presented here is a computational level theory of comparison and similarity and does not address this issue. The solutions to difficult real world problems can be approximated using combinatoric optimization procedures such as simulated annealing [14]. In essence, every algorithmic model of analogy solves this combinatoric optimization problem by heuristically combining mapping constraints.

The similarity equation consists of four terms that combine linearly:

$$\mathbf{E}(x,y) = \alpha_1\mathbf{\Theta}(x,y) + \alpha_2\mathbf{\Upsilon}(x,y) + \alpha_3\mathbf{\Omega}(x,y) + \alpha_4\mathbf{\Phi}(x,y) \quad (3)$$
$$\text{where } \alpha_1, \alpha_2, \alpha_3, \alpha_4 \geq 0.$$

The four terms are defined below. The terms are organized from most semantic in nature to most structurally focused.

The $\mathbf{\Theta}$ term is analogous to Tversky's (1977) contrast model and captures raw semantic similarity:

$$\mathbf{\Theta}(x,y) = \varphi_1\mathbf{f}(X \cap Y) - \quad (4)$$
$$(1-\varphi_1)\Big(\delta_1\mathbf{f}(X-Y) + (1-\delta_1)\mathbf{f}(Y-X)\Big)$$
$$\text{where } 1 \geq \varphi_1 \geq 0, \text{ and } 1 \geq \delta_1 \geq 0$$

where φ_1 determines the relative importance of commonalities and differences in determining the similarity of two analogs. The parameter δ_1 determines how asymmetric the similarity judgment is. Given that the focus of a comparison is usually on the target, δ_1 should be greater than $.5$.

The second term captures semantic similarity arising from correspondences:

$$\mathbf{\Upsilon}(x,y) = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}\mathbf{C}(x_i,y_j) \quad (5)$$

where $\mathbf{C}(x_i,y_j)$ is:

$$\mathbf{C}(x_i,y_j) = \varphi_1\mathbf{F}(x_i) \text{ if } x_i \text{ is identical to } y_j, \quad (6)$$
$$\text{else } (\varphi_1 - 1)\mathbf{F}(x_i).$$

The $\mathbf{\Theta}$ term and Tversky's contrast model do not distinguish between commonalities and differences that arise from relational elements that are in correspondence and those that are not in correspondence. The Υ term specifically addresses commonalities and differences that are in correspondence (i.e., elements linked in A). Commonalities arising from correspondences are processed differently (i.e., have different time courses and differentially affect perceived similarity) than commonalities (or matches) that are not in correspondence [8]. Likewise, differences that can be put into correspondence are psychologically distinct from differences that cannot be put into correspondence [15].

Humans are also sensitive to higher-level matches (i.e., compatibility relations), as in the solar system/atom example. Analogs are perceived as similar when they have a common relational structure. The Ω term captures this type of similarity and it is high when elements from one analog map to elements in the other analog and their parents are also in correspondence.

$$\mathbf{\Omega}(x,y) = \quad (7)$$
$$\sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{n}\sum_{l=1}^{n} S^x_{ij}S^y_{kl}A_{ik}A_{jl}\mathbf{F}(x_i)\mathbf{F}(x_j)\mathbf{F}(y_k)\mathbf{F}(y_l).$$

The Φ term is purely structural and ranges from 0 to 1. The Φ term is 1 when the mapping between x and y is a bijection (one-to-one and onto). In general people prefer analogies or mappings that are one-to-one [4]. Here, we assume that complete mappings are also preferred. The Φ term is defined as:

$$\mathbf{\Phi}(x,y) = \delta_1\left(\frac{1}{1+\sum_{i=1}^{m}\mathbf{F}(x_i)\left(1-\sum_{j=1}^{n}A_{ij}\right)^2}\right) + \quad (8)$$
$$(1-\delta_1)\left(\frac{1}{1+\sum_{i=1}^{n}\mathbf{F}(y_i)\left(1-\sum_{j=1}^{m}A_{ij}\right)^2}\right)$$

The data sets considered in the next section are all from controlled experiments and mappings exist that lead to maximal values of Φ. Under these conditions, only these solutions are considered by subjects (i.e., the parameter α_4, weighting Φ, is set to a value large enough to prohibit consideration of incomplete or conflicting mappings).

The Similarity Equation and Human Data

In this section, data from Goldstone (1994) and data from two new experiments is fit by the similarity equation. The fits are intended to illustrate how the similarity equation can be applied to human data and should not be taken as a definitive test of the equation's form. The equation's form will certainly be refined as it is applied to more data sets.

Goldstone (1994) Experiment 2

The similarity equation was applied to data from Goldstone's (1994) Experiment 2. Subjects rated the similarity of two displays. Each display consisted of a pair of schematic butterflies. Each butterfly could be represented by four features (type of tail, type of body, type of wings, type of head). The number of matches in place (correspondences as in Equation 5) and the number of matches that were not in correspondence was manipulated, yielding fifteen conditions. Table 1 illustrates the fifteen within subject conditions. Only the Θ and the Υ terms from the similarity equation are relevant

Table 1: Various feature transformation from Goldstone's (1994) Experiment 2. In the Target column, each of the four positions in each letter string represents a part of the target stimulus (i.e., the second position in each string could denote the head). Each letter (A, B, C, D, W, X, Y, or Z) represents a particular feature value. The base stimulus is always represented as ABCD. The next two columns list the number of Θ and Υ matches. The Rated Similarity column denotes human subjects' rated similarity of the base and target stimuli.

Target	Θ Matches	Υ Matches	Rated Similarity
WXYZ	0	0	1.91
XYDZ	1	0	1.48
YZDD	1	1	3.09
XYCZ	1	1	3.12
BAYZ	2	0	3.11
YZCD	2	2	4.65
BADZ	3	0	3.60
BACZ	3	1	4.52
BADD	3	1	4.57
ABDZ	3	2	4.96
ABDD	3	3	6.56
ABCZ	3	3	6.78
BADC	4	0	4.82
BACD	4	2	6.38
ABCD	4	4	8.79

for fitting this dataset, along with the α_1 and α_2 parameters, because 1) asymmetries were not a concern ($|X|$ is equal to $|Y|$), 2) differences and commonalities in Equations 4 and 5 are perfectly correlated allowing all difference terms to be dropped (i.e., φ_1 is set to 1), 3) the value of Ω and Φ do not vary across conditions.

For simplicity, a linear relationship was assumed between the maximal value generated by the similarity equation and subjects' similarity ratings. Of course, similarity is probably a more complex function of $\mathbf{E}(x, y)$ for a number of reasons, including the presence of scale effects [19]. Nevertheless, with this simplifying assumption, the similarity equation accounted for 97.0 % of the variance in the data. The similarity equation states that different sources of similarity combine additively. A modified version of the equation was fit to the data that contained a term capturing the interaction between Θ and Υ. This augmented equation did not capture significantly more variance (97.3%), supporting the stance that the terms combine additively. The fit for SIAM, a special purpose interactive connectionist model developed by Goldstone (1994), was equivalent (it accounted for 97.7 % of the variance). The similarity equation offers a simpler account of the data — perceived similarity arises from a linear weighting of the Θ and Υ terms. To SIAM's credit, it can account for aspects of the time course data, like that subjects tend to weight matches in correspondence (i.e., Υ matches) more later in processing than Θ matches (this is SIAM's added value). Such data is outside the province of a computational level theory.

Experiment 1

In Goldstone's (1994) Experiment 2, the Ω term was not relevant to the data fit. To further test the predictions of the similarity equation, I collected data from subjects in a task in which higher-order relations could impact the similarity equation's predictions (i.e., the Ω term's value varies across conditions). In Experiment 1, subjects rated the similarity of two situations. The number of higher-order relations shared by the two situations was manipulated (as well as the number of Θ and/or Υ matches). The main prediction the similarity equation makes in Experiment 1 is that feature and relation matches will affect rated similarity additively.

Subjects Twenty-one Northwestern University undergraduate students participated in the experiment for course credit.

Design and Overview of the Experiment The two variables (Feature Match/Mismatch and Relation Match/Mismatch) were crossed for a 2 X 2 within subjects factorial design. The design is illustrated in Figure 1. The value of each term in the similarity equation for each condition is shown in Table 2. On each trial, subjects rated the similarity of two situations. Subjects completed 20 trials in each condition for a total of 80 trials. The order of trials was randomized for each subject.

Stimuli and Counterbalancing Each stimulus contained the descriptions of two situations. Each situation description consisted of two sentences (see Figure 1). One situation description was displayed on the left side of the screen. The other situation description was displayed on the right side of the screen. Above the description on the left side of the screen was the label "Situation A." Above the description on the right side of the screen was the label "Situation B." Underneath the descriptions was a rating scale (1 indicated low similarity and 9 indicated high similarity).

Each situation contained two characters that were either both male or both female. Character names were randomly chosen (subject to constraints imposed by the trial's condition) without replacement from the following list of names: Anne, Jennifer, Linda, Susan, Wendy, Bill, Jeff, John, Keith, and Steve. On Feature Match trials, the gender of the characters in both situations matched (i.e., all characters were male or female). Whether the common gender was male or female was randomly determined for each Feature Match trial. On Feature Mismatch trials, the genders of the characters in the two situations were different such that the two characters from one situation were both male and the two characters from the other situation were both female. On each Feature Mismatch trial, it was randomly determined whether situation A contained the male or female characters.

Both characters from a situation appeared in both sentences. Each sentence contained a predicate that linked the two characters. The same predicates appear in both situations. Two predicates were randomly chosen without

Table 2: The values of the four terms for the four conditions in Experiments 1 and 2. Notice that Θ and Υ are perfectly correlated.

	Θ	Υ	Ω	Φ
Feature Match/Relational Match	8	8	12	1
Feature Mismatch/Relational Match	7	7	12	1
Feature Match/Relational Mismatch	8	8	10	1
Feature Mismatch/Relational Mismatch	7	7	10	1

replacement for each trial from the following list: is taller than, respects, and likes. Which predicate appeared in the first or second sentence within a situation was random. It was also random whether or not the same character appeared first in both sentences in situation A (the character order is fixed for situation B given the character order in situation A and the trial's condition). Again, Figure 1 illustrates an example situation pair for each condition.

Procedure Text was displayed in black on a white background. Trials began with a message displayed in the upper left corner of the screen alerting the subject to prepare for the next trial. After 1000 ms, this message was removed and the stimulus was displayed (i.e., the two situations along with the rating scale). Subjects then pressed a key (1 through 9) to indicate how similar the two situations were (1 indicated low similarity and 9 indicated high similarity). After subjects responded, there was a 1500 ms pause and then the next trial began.

Results Table 2 shows the values of the four terms for each condition. As in the previous fit, the number of relevant parameters required can be reduced to 2 (the α_1 parameter for the Θ term and the α_3 parameter for the Ω term). The mean similarity ratings (averaged across subjects) for each condition are shown in Table 3. The similarity equation fit 99.9% of the variance in the data. To provide a stronger test, individual subjects' data was fit. Of course, the fit for this data will not be as good because the data for individual subjects is not as stable and each subject uses a slightly different rating scale (i.e., high similarity for one subject may result in a rating of 8, while another subject may give a rating of 7). Nevertheless, the equation fit 71.9% of the variance (df=81). A modified version of the equation was fit to the data that contained a term capturing the interaction between Θ and Ω. This augmented equation did not capture significantly more variance (72.1%, df=80), supporting the stance that the terms combine additively.

Experiment 2

A second experiment explored how task demands affect comparison. The materials and procedure were identical to the previous experiment except that after making a similarity judgment subjects were asked to state how the people in the target and base analogs corresponded to one another. This judgment should force subjects to focus more on high-order relational matches and should lead to a higher weighting of the Ω term relative to the Θ term in the similarity equation.

Table 3: Similarity ratings for each condition (averaged over subjects) in Experiment 1.

	Relational Match	Relational Mismatch
Feature Match	8.23	4.50
Feature Mismatch	7.51	3.39

Subjects Seventy-one Northwestern University undergraduate students participated in the experiment for course credit. The subjects were from the same population as the subjects in Experiment 1. Experiments 1 and 2 were run concurrently (though no subjects participated in both experiments).

Design and Overview of the Experiment The design was very close to that of Experiment 1. The main difference was that subjects made a correspondence judgment after making a similarity judgment. Another difference was that subjects performed sixty trials (fifteen in each condition) as opposed to the eighty trials performed in Experiment 1.

Stimuli and Counterbalancing The stimuli and counterbalancing were identical to Experiment 1 with the following addition — after making a similarity judgment, one character was randomly chosen from situation A and another character was randomly chosen from situation B and subjects were asked if they corresponded to one another.

Procedure The procedure was identical to Experiment 1 except that subjects made a correspondence judgment immediately after making a similarity judgment. After making the similarity judgment, a text message appeared below the rating scale. The message asked if a particular character from situation A corresponded to a particular character from situation B. Subjects were instructed to press the 'Y' key if they thought the two characters corresponded and to press the 'N' key if they thought the two characters did not correspond. The Yes/No question was displayed along with both situation descriptions and the rating scale. After making the correspondence judgment, there was a 1500 ms pause and then the next trial began.

Results The main predictions held. Table 4 shows the mean ratings for each condition. Feature matches had a small effect on rated similarity while relational matches had a large effect on rated similarity. The ratio α_3/α_1 was three times larger in Experiment 2 than it was in Experiment 1. Subjects also made the correspondence judgments in the manner predicted by the similarity equation's mapping matrix A. In terms of fit, 99.9% of the variance in the averaged data was accounted for. For individual subject fits, 66.1% (df=281) of the variance was accounted for and adding an interaction term did not improve the fit (66.1%, df=280).

Table 4: Similarity ratings for each condition (averaged over subjects) in Experiment 2

	Relational Match	Relational Mismatch
Feature Match	8.32	4.47
Feature Mismatch	8.02	4.08

While the fits from Experiments 1 and 2 (as well as from Goldstone's Experiment 2) suggest different sources of similarity combine additively, I predict that after consideration of more diverse data sets (e.g., [9]) the similarity equation will be revised to make allowances for interactions between terms under certain conditions. The equation and data presented here are simply intended to motivate a new framework for approaching comparison and similarity.

Discussion

The similarity equation presented here is a computational level theory of human comparison and perceived similarity that can account for basic findings in the similarity and analogy literatures. The equation provides clarity to the discussion of similarity because it distinguishes between a number of different factors that can affect perceived similarity. An accurate characterization of similarity is critical given its central role in theories of categorization, decision making, analogy, problem solving, and object recognition.

Twenty years after Tversky's (1977) classic paper, many of the same questions remain. How are the representations of analogs determined, how do they change as an outcome of comparison, and how is feature saliency modulated? One possibility is that instead of static representations being compared, retrieval and comparison are interleaved such that the current mappings between the analogs direct which other information is retrieved and represented in the base and target. Analogical inference may also direct the construction of the target analog's representation. In other words, properties or features can emerge as a result of the comparison process [18]. Hopefully this work will demarcate what is known and what is common to competing models so that researchers can wisely focus their efforts.

Acknowledgments

I would like to thank John Hummel and Keith Holyoak for illuminating discussions. I would like to thank Dedre Gentner and Arthur Markman for their helpful comments on a previous draft. Finally, I would like to thank Rob Goldstone for providing me with his data.

References

[1] BOWDLE, B. F., AND GENTNER, D. Informativity and asymmetry in comparisons. *Cognitive Psychology 34* (1997), 244–286.

[2] FALKENHAINER, B., FORBUS, K. D., AND GENTNER, D. The structure mapping engine: Algorithm and examples. *Artificial Intelligence 41* (1989), 1–63.

[3] FORBUS, K. D., GENTNER, D., AND LAW, K. MAC/FAC: a model of similarity-based retrieval. *Cogntive Science 19* (1994), 141–205.

[4] GENTNER, D. Structure-mapping: A theoretical framework for analogy. *Cogntive Science 7* (1983), 155–170.

[5] GENTNER, D., AND MARKMAN, A. B. Structure mapping in analogy and similarity. *American Psychologist 52* (1997), 45–56.

[6] GLEITMAN, L. R., GLEITMAN, H., MILLER, C., AND OSTRIN, R. Similar, and similar concepts. *Cognition 58* (1996), 321–376.

[7] GOLDSTONE, R. L. Similarity, interactive activation, and mapping. *Journal of Experimental Psychology: Learning, Memory, & Cognition 20* (1994), 3–28.

[8] GOLDSTONE, R. L., AND MEDIN, D. Time course of comparison. *Journal of Experimental Psychology: Learning, Memory, & Cognition 20* (1994), 29–50.

[9] GOLDSTONE, R. L., MEDIN, D., AND GENTNER, D. Relational similarity and the nonindependence of features in similarity judgments. *Cognitive Psychology 23* (1991), 222–262.

[10] HOPFIELD, J. J., AND TANK, D. W. Neural computation of decision in optimation problems. *Biological Cybernetics 52* (1985), 141–152.

[11] HUMMEL, J. E., AND HOLYOAK, K. J. Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review 104* (1997), 427–466.

[12] JAMES, W. *The Principles of Psychology: Volume 1*. Dover, New York, 1890/1950.

[13] KEANE, M. T., LEDGEWAY, T., AND DUFF, S. Constraints on analogical mapping: A comparison of three models. *Cogntive Science 18* (1994), 387–438.

[14] KIRKPARTICK, S., GELATT, C. D., AND VECCHI, M. P. Optimization by simulated annealing. *Science 220* (1983), 671–680.

[15] MARKMAN, A. B., AND GENTNER, D. Splitting the differences: A structural alignment view of similarity. *Journal of Memory and Language 32* (1993), 517–535.

[16] MARKS, L. E. Bright sneezes and dark coughs, loud sunlight and soft moonlight. *Journal of Experimental Psychology: Human Perception and Performance 8* (1982), 177–193.

[17] MARR, D. *Vision*. W. H. Freeman, San Francisco, 1982.

[18] MEDIN, D. L., GOLDSTONE, R. L., AND GENTNER, D. Respects for similarity. *Psychological Review 100*, 2 (1993), 254–278.

[19] PARDUCCI, A. Category judgment: A range-frequency model. *Psychological Review 72* (1965), 407–418.

[20] SHEPARD, R. N. The analysis of proximities: Multidimensional scaling with an unkown distance function. Part 1. *Psychometrika 1* (1962), 125–140.

[21] TALMY, L. Figure and ground in complex sentences. In *Universals of human language*, J. Greenburg, C. Ferguson, and M. Moravcsik, Eds. Stanford University Press, Stanford, 1978, pp. 625–649.

[22] TVERSKY, A. Features of similarity. *Psychological Review 84* (1977), 327–352.