When a bad metaphor may not be a victimless crime:
The role of metaphor in social policy

Paul Thibodeau (pthibod1@stanford.edu)
James L. McClelland (mcclelland@stanford.edu)
Lera Boroditsky (lera@stanford.edu)
Stanford University, Department of Psychology
Jordan Hall (Bldg 420), 450 Serra Mall, Stanford, CA 94305 USA

Abstract

Metaphors are pervasive in our discussions of abstract and complex ideas (Lakoff & Johnson, 1980), and have been shown to be instrumental in problem solving and building new conceptual structure (e.g., Gentner & Gentner, 1983; Nersessian, 1992; Boroditsky, 2000). In this paper we look at the role of metaphor in framing social issues. Our language for discussing war, crime, politics, healthcare, and the economy is suffused with metaphor (Schön, 1993; Lakoff, 2002). Does the way we reason about such important issues as crime, war or the economy depend on the metaphors we use to talk about these topics? Might changing metaphors lead us to different conceptions and in turn different social policies? In this paper we focused on the domain of crime and asked whether two different metaphorical systems we have for talking about crime can lead people to different ways of approaching and reasoning about it. We find that framing the issue of crime metaphorically as a predator yielded systematically different suggestions for solving the crime problem than when crime was described as a virus. We then present a connectionist model that explores the mechanistic underpinnings of the role of metaphor.

Keywords: Metaphor, analogy, connectionism, social policy

Introduction

We use a variety of metaphors when discussing crime. In some cases, crime and criminals are described as predators, preying on the public, lurking in the shadows, stalking their victims. As William James put it, "Man… is simply the most formidable of all beasts of prey, and, indeed, the only one that preys systematically on its own species" (p. 846, 1904). The police in this set of metaphors are meant to hunt down, lay traps and attempt to catch or capture the criminals, so as to lock them away.

In other cases, crime is described as a disease or epidemic. It infects cities and plagues neighborhoods. On this framing, the job of police is centered on diagnosing and treating the root cause of the problem, stopping the spread of the infection, identifying risk factors to prevent future outbreaks, and restoring the health of the community. Public health researchers have explicitly proposed that treating crime as a disease will help us find the cure (Guerrero & Concha-Eastman, 2001). A violence prevention program operated by an epidemiologist in Chicago takes this metaphor to heart, treating crime according to the same regimen used for contagious diseases like AIDS and tuberculoses (Kotlowitz, 2008).

In some cases, scholars have even cast bad metaphors as a societal danger. George Kelling, a criminal justice scholar, describes a case in which a serial rapist attacked 11 girls over a 15-month period before being captured by the police (Kelling, 1991). The police later revealed that they had deliberately withheld information from the public that could have prevented at least 8 of the attacks, because it might have compromised the traps they had laid for the suspect. The girls, Kelling argues "were victims… not only of a rapist, but of a metaphor" (p. 1, 1991). The police in this view were too entrenched in the role of hunting down and catching the criminal, and neglected their responsibility to inoculate the community against further harm.

In this paper we empirically test whether metaphors indeed structure how people reason about social issues like crime. We also present a computational model that explores the mechanisms through which metaphors may shape people's thinking.

The Current Study

We focus on two common frameworks for talking about crime: "crime as a *predator*" and "crime as a *virus*." Both are used frequently and productively in discourse about crime. However, if we take these metaphors seriously, they offer very different implications for how societies should deal with crime. For example, to deal with a dangerous predator on the loose, one might try to hunt, trap or cage the animal. If crime is like a predator, then the best way to deal with crime is to catch and imprison as many criminals as possible. Solutions to the crime problem might include increasing the police force, harsher enforcement of laws, longer prison terms, and so on. If crime is a disease, the set of implications is rather different. To treat a disease, one might attempt to diagnose and treat the root cause of the problem, and one would also aim to restore the organism's immune system so that it is not susceptible to future infections[1]. If crime is a disease, then really dealing with crime involves not only treating the symptoms, but getting

[1] There are two somewhat different metaphorical frameworks that treat crime as an illness. In one, the community or population is seen as an organism, and crime is a disease that is developing inside that organism (e.g., "'Violent crime is a cancer that eats away at the very heart of society."). In another, the community is seen as individual agents and crime is a contagious disease that can be passed on from one person to another forming an epidemic. In this paper the stimuli did not strongly distinguish between these two metaphors, but doing so would be an interesting extension of this work, as the two metaphors suggest somewhat different implications for treating crime.

to the root cause of the problem, and restoring the health of the community so that it is no longer susceptible to future crime flare-ups.

While these analyses of the metaphors seem plausible, what we don't know is whether such metaphors in fact have any psychological weight. Does casually encountering one metaphor or another in discourse about crime actually lead regular English speakers to come to different conceptions of the crime problem? Would people unwittingly come up with different types of solutions for the crime problem when exposed to one metaphor versus another?

Experiment

The experiment was designed to explore whether simply embedding a common metaphor in an otherwise neutral report about crime can systematically influence people's approach to solving the crime problem. In the task, participants read a report about crime in a fictional city and then answered questions about the city. The report contained mostly crime-relevant statistics, and also two brief instances of either the crime as predator metaphor or the crime as virus metaphor. After reading the report, participants answered questions relating to crime in the city. Critically, in one of these questions, participants were asked to propose a solution to the crime problem.

If metaphors in fact have psychological weight, then being exposed to different metaphors for crime may lead people to propose different solutions to the city's crime problem. For example, people exposed to the crime as a predator metaphor might propose toughening law enforcement, while people exposed to the crime as disease metaphor might think about dealing with problems in the community and improving the social environment to prevent future crime. Of course, it is also possible that such metaphors are simply ornamental flourishes of language, and do not influence how people conceive of important social issues like crime.

Method

Participants Four-hundred sixty-three students participated in the study as part of a course requirement – 104 from Stanford University and 359 from the University of California–Merced. The same patterns were found in both samples, so we report pooled data. Gathering data from the two populations allowed us to get a somewhat broader cross-section of the general population. This seemed important since people's conceptions of social issues like crime are likely to differ as a function of factors like socioeconomic status and personal experience.

Materials The survey was included in a larger packet of questionnaires that were unrelated to this study. Participants filled out the packet individually in a quiet room. Our survey consisted of a single page which included a short paragraph about crime in the fictional city of Addison and some follow-up questions. The paragraph mostly contained crime statistics, which were the same in the two conditions. The two conditions differed only in two sentences in the paragraph which were used to embed either the crime-as-predator metaphor or the crime-as-disease metaphor. Half of the subjects were given the crime-as-predator version and half the crime-as-disease version. The report read:

> Crime is a (wild beast preying on / virus infecting) the city of Addison. The crime rate in the once peaceful city has steadily increased over the past three years. In fact, these days it seems that crime is (lurking in / plaguing) every neighborhood. In 2004, 46,177 crimes were reported compared to more than 55,000 reported in 2007. The rise in violent crime is particularly alarming. In 2004, there were 330 murders in the city, in 2007, there were over 500.

Below this report, subjects were asked to briefly describe the crime situation in Addison (to make sure they read and understood the story), and were then instructed to answer the following two questions: 1. In your opinion what does Addison need to do to reduce crime? 2A) What aspect of the report was most influential in your decision? 2B) Please underline the part of the report that was most influential in your decision.

Fifty-two participants were given version A of question 2, and 411 participants were given version B. This question was aimed at discovering if participants explicitly noticed or made use of the metaphor, and we wanted to allow participants different opportunities to report this.

Results

The solutions participants proposed to the crime problem in Addison differed depending on the metaphorical frame used in the crime report. Interestingly, most participants were not consciously aware of this influence. In response to the second question (which part of the report was most influential in your decision?), only 3% of the participants reported that the metaphorical framing influenced their decision.

Participants' proposed solutions to the crime problem in Addison were coded into two categories: 1) social environment and 2) law enforcement / punishment. Responses were categorized as "social environment" if they suggested a social reform (e.g., healthcare or educational or welfare programs) or investigating the cause of the problem (e.g., "look for the root cause"). Responses were categorized as "law enforcement / punishment" if they suggested restructuring the police force (e.g., hiring more officers, retraining officers, calling in the national guard) or modifying the penal structure (e.g., instituting harsher penalties, building more jails).

Each participant's response was coded as either 1 point for the social environment category, 1 point for the enforcement/punishment category, or split .5 points for each category if both types of solutions were proposed. Responses were coded by a blind coder. Of all responses, 9% did not fit into either category. In nearly every case this

was because the response lacked a suggestion (e.g., "I don't know", "I need more information", "It should be addressed").

Results are shown in Figure 1. Overall, participants were more likely to suggest an enforcement/punishment solution than a social environment solution (74% enforcement, 23% social environment), $\chi 2 = 98.12$, $p < .001$. However, participants given the crime-as-virus metaphorical framing were more likely to suggest social reform (31%) than participants given the crime-as-predator framing (20%), $\chi 2 = 6.88$, $p < .01$.

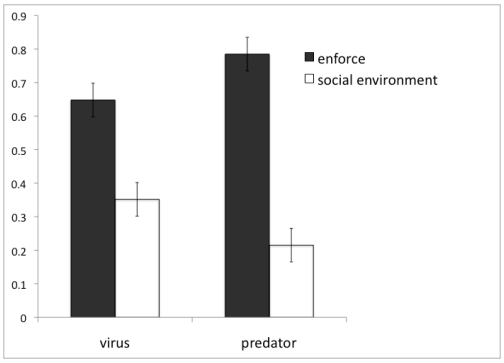

Figure 1: Proportion of proposed solutions broken down by metaphor frame

Discussion

The results suggest that metaphors can influence how people conceptualize and in turn hope to solve important social issues. It appears that even casually encountering one metaphor or another in discourse about crime can lead people to unwittingly propose different types of solutions for the crime problem. Importantly, it appears that the metaphors had a subconscious effect on people's reasoning. Very few of our participants thought that the metaphors influenced their crime-reducing suggestion.

Simulation

To further explore a possible mechanism for the effects observed in this experiment, we created a computational model, using a connectionist architecture. Our goal was to explore how connectionist architectures can model people's reasoning in highly structured and relational domains such as metaphor and analogy, often seen as an area of weakness for connectionist architectures (Fodor & Pylyshyn, 1988; Gentner & Markman, 1993). Recent work by Rogers and McClelland (R&M, 2008) suggests that connectionist networks may indeed be sensitive to the types of relational similarity that defines analogy and metaphor (Gentner, 1983).

To explore metaphors in a connectionist model we considered three domains (crime, predator, and virus). By gathering descriptions of the three domains we were able to create a set of 36 representative phrases (11 for predator, 11 for virus, and 14 for crime). Each phrase is characterized by a "domain-label" (crime, predator, or virus) as the first object of the phrase, a relation (e.g., "caused_by", "infects", "capture_with") as the means of associating objects, and a "completion" (e.g., "poverty", "people", and "trap") as the second object of the phrase. For example, one phrase might read "crime/caused_by/poverty."

This list of 36 phrases included 17 "relations" and 19 "completions". Each relation and completion fell into one of four categories - crime-specific (e.g., "arrest"), predator-specific (e.g., "trap"), virus-specific (e.g., "infect"), or shared (e.g., "cause").

There were four domain-specific relations that reflected suggestions for solving each problem ("solutions units"): these included "treat_by" in the virus domain, "exterminate_by" in the predator domain, and "reform_by" and "fight_by" in the crime domain. These relations reflected the way that people suggested solving each problem – e.g., "solve crime by investigating/reforming the education system" or "solve crime by fighting it with more police." Figure 2 illustrates the set of relations and completions in each domain as well as how they were linked to domains in training. The circles highlight completions that were activated as a function of the domain-specific "solution units."

We trained the model by presenting one of the relations and the corresponding domain-label as inputs and checking the completions. Back-propagation of error was used to adjust the network so that eventually the pairing of a

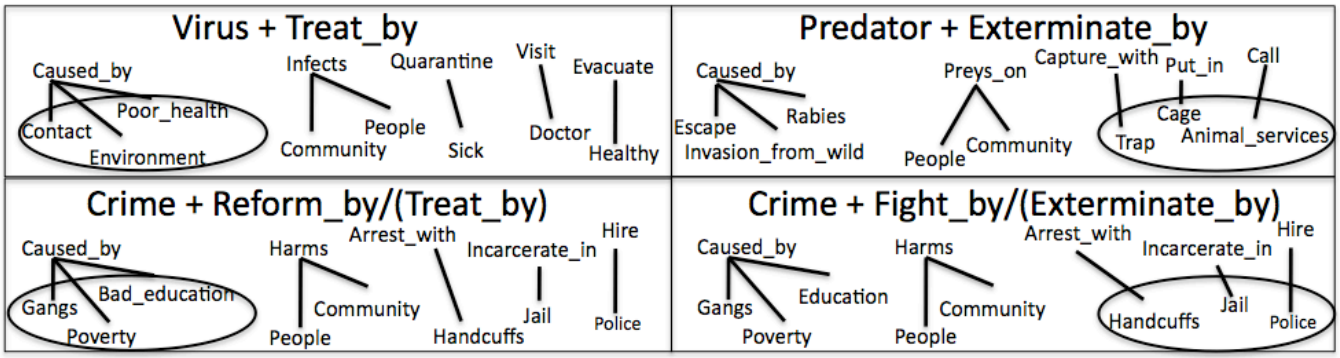

Figure 2: Each quadrant represents a domain (the crime domain is illustrated twice). The upper-left quadrant illustrates the virus domain, the upper-right illustrates the predator domain, and the bottom two illustrate the crime domain. Within each quadrant, the upper row of words lists relations and the lower lists completions. Links between the relations and attributes represent associations learned in the training phase. Circles highlight the completions that are activated when the domain label is paired with the "solve" relation listed at the top of each quadrant.

domain-label and relation activated all of the relevant completions. In the test phase, we presented the model a domain-label and relation that belonged to different domains – pairings that the model had not seen in training (e.g., "crime" and "exterminate_by").

The virus and predator domains both shared some structural similarity with the crime domain. If the model was able to capture this structural similarity in learning, then it should be able to interpret a structurally similar but crossed pairing. If not, it might activate some unsystematic set of irrelevant completions.

Method

Network architecture. The network's architecture is shown in Figure 3. It is modeled on R&M's feed-forward network and is trained through back propagation. There are two input layers (labeled "domain-labels" and "relations"), three hidden layers, and one output layer (labeled "completions"). Each of the input layers projects to an associated set of hidden units - one is the learned representation of the "domain-labels" and the other is the learned representation of "relations." These two sets of hidden units project forward into the third set of hidden units - the learned representation of the item and relation units combined. This hidden layer feeds forward to the output layer – the "completions".

Note that the two input layers and the output layer are divided into distinct domains. This separation is what allows us to explore metaphor processing. That is, in the training phase, the model learns aspects of the three domains separately. In the test phase, previously unpaired relations and domains are crossed.

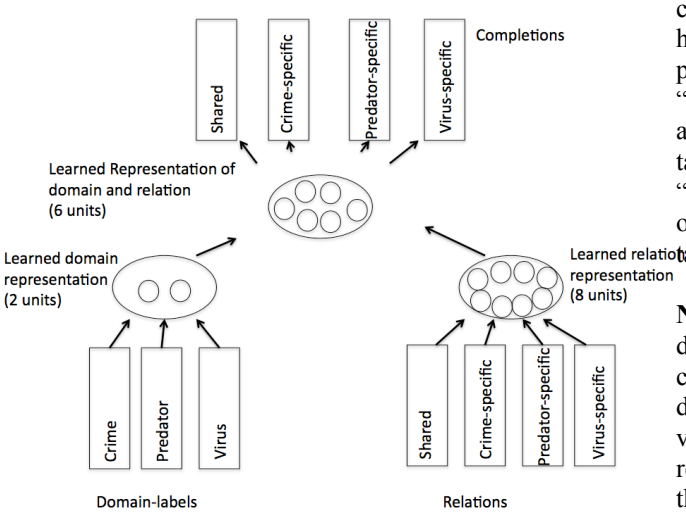

Figure 3: Network architecture

Creating concepts We sought to create ecologically valid representations of our target domains, so we gathered descriptions of the three situations. Forty-one participants from Foothill College completed an online survey in exchange for course credit. The survey asked participants to describe, in at least ten sentences, several situations. Three of the situations were: 1) worsening crime, 2) a wild beast on the loose, and 3) a spreading virus. These descriptions formed the basis of the training patterns that were presented to the model.

We selected a set of relations by reading through the descriptions and collecting lists of the verbs used to describe the situations. For each domain, we identified a set of the most commonly used verbs – often combining counts of verbs that shared a similar meaning (e.g., "capture" and "catch"). We then read through the stories again to identify completions. We created a list of the subjects and objects of the sentences that were commonly associated with the selected verbs. This list was similarly condensed.

Structural Similarity Previous research has demonstrated that structural similarity is a critical feature of a comprehensible metaphor (e.g., Gentner & Clement, 1988). The situation descriptions revealed some critical structural features, which mirrored the findings of the experiment.

Descriptions of the virus situation often emphasized the underlying aspects of the city that enabled the virus to spread (e.g., investigating the water supply and food sources as well as the ways in which people interact socially). Descriptions of the predator situation, on the other hand, often emphasized targeted responses (e.g., hunting down and caging the predator). Descriptions of the crime situation fit two schemas. In one, participants focused on the reforming underlying causes of crime – e.g., social welfare. In the other, participants focused on fighting crime head on – e.g., hiring police officers and building jails.

One relation in the virus set ("treat_by"), one relation in the predator set ("exterminate_by"), and two relations in the crime set ("reform_by" and "fight_by") were critical for highlighting structural similarity. These relations reflect the participants' consensus on how to solve each situation. The "treat_by" relation is associated with the causal elements of a virus. The "exterminate_by" relation is associated with targeted and immediate responses to a predator. The "reform_by" relation is associated with the causal elements of crime and the "fight_by" relation is associated with the targeted and immediate responses to crime.

Network training The training patterns paired a single domain-label with a single relation and all appropriate completions. In some cases, completions were linked to a domain via a shared relation; in other cases they were linked via a domain-specific relation. Figure 2 illustrates the set of relation-completion pairings within each domain (note that there were not two distinct representations of crime).

The connection weights in the network were initialized with very small random values. There were 19 training patterns. All patterns were presented to the model an equal number of times in random order. Back-propagation was used to do online learning after each pattern. The network was trained for 30,000 epochs without momentum with a

learning rate of .05 at which point 99% of the output units were activated to within .1 of their target values.

Network testing In the testing phase, we explored whether the model had learned and could utilize the relational structure of the three target domains. Two test patterns illustrate the critical pairings: 1) Crime + Treat and 2) Crime + Eliminate.

These label–relation pairings were never seen by the network in the training phase. If the model could harness the relational structure of the domains, we would expect the underlying causes of crime activated when crime was paired with "treat_by" (as in the lower left quadrant of Figure 2). We would expect the immediate responses of crime activated when crime was paired with "exterminate_by" (as in the lower right quadrant of Figure 2).

Results

We ran three simulations with different initial starting weights, and the resulting activation in the completions was to ensure that the results do not reflect chance findings from a particular set of starting weights. Figure 4 illustrates the average level of activation in completions in response to the two test inputs.

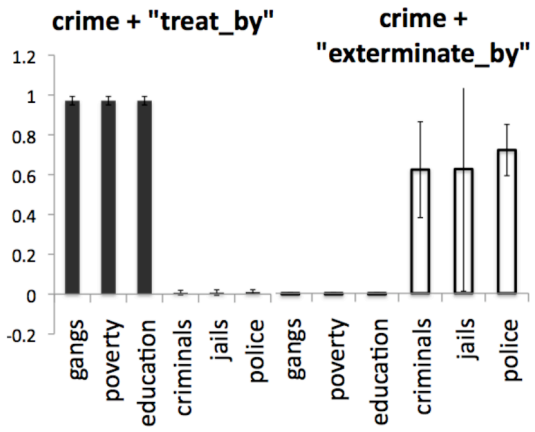

Figure 4: Average activation level of completions when crime was paired with "treat_by" (left) and "exterminate_by" (right) as inputs. No other completion averaged an activation level greater than .1.

The results reveal that pairing "crime" with "treat_by", which were unpaired in the learning phase, activated the underlying causes of crime and not the immediate targeted responses to crime or causes of a virus. Pairing "crime" and "exterminate_by", which were also unpaired in the learning phase, activated the immediate targeted responses to crime and not the underlying causes of crime or the immediate responses to a predator.

A closer look at the learned and distributed representations of the relation-units reveals how the metaphoric influence works. The model naturally represents "treat_by" and "reform_by" as well as "exterminate_by" and "fight_by" similarly because the two pairs of relations serve similar roles in their specific domains and because there are a scarce number of representation units for the relations. That is, both the "treat_by" and "reform_by" relations activate completions that are associated with "cause". Therefore, the model can represent the two domains with very similar patterns of representation. The model does not confuse the two representations because it activates the completions only after integrating the information from the domain-label stream of input (i.e., in the subsequent representation layer).

The similarity between the "exterminate_by" and "fight_by" relations is slightly more difficult to explain because there are no shared relation units that could underlie the mapping. Instead, as in the case of R&M's simulation, the network succeeds because the two domains are nearly identical structurally (identical except for the additional crime-specific relation unit, "reform"). Figure 5 illustrates the learned and distributed representational similarity between the solution units.

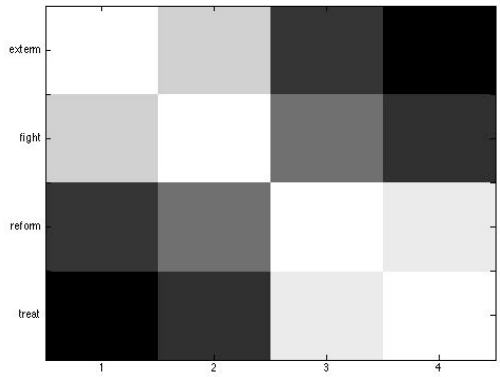

Figure 5: Learned, distributed representational similarity among relations that serve a structurally similar purpose. The columns represent activation levels in the eight units that learn a representation for relations.

Discussion

The model presented here learned and exploited the relational similarity between the three target domains in a way that was consistent with the results from the Experiment. Specifically, the model learned to represent the three domains by learning both the low level particulars of each domain as well as high-level structural similarity. This model builds on earlier simulations by Hinton's "family tree" simulations (1989) as well as R&M's (2008).

Our model is different from previous computational models of analogy/metaphor (e.g., Falkenhainer et al., 1989; Eliasmith & Thagard, 2001; Hummel & Holyoak, 1997) in that the conceptual representations are entirely learned and distributed. In fact, it is precisely because the representations of the relation units are distributed that cross-domain inputs yield appropriate outputs.

Limitations and future directions It should be noted that the model presented here does not fully reflect the experimental paradigm. Participants in the experiment are not asked to, e.g., "treat crime." Future work will attempt to

design a model in which the representation of one domain (e.g., virus) is activated and remains activated while inputs from a distinct domain (e.g., crime) are presented to the model. The model presented here is intended as a first step in closely studying the mechanisms that underlie metaphor processing from a connectionist perspective.

General discussion

In this paper we presented one empirical demonstration of the role of metaphors in shaping people's thinking about an important social issue, and a computational model which explored a possible mechanism for such effects. In the experiment, participants' suggestions for solving a crime problem were systematically influenced by a metaphoric frame. When crime was compared to a virus, participants were more likely to suggest reforming the social environment of the infected community. When crime was compared to a predator, participants were more likely to suggest attacking the problem head on – hiring more police officers and building jails.

In the simulation we explored how neural networks are able to learn and use structural similarity to successfully interpret cross-domain relations. When domain-specific relations activated structurally similar completions, the model naturally learned to represent them in a similar way.

The experiment and simulation confirm recent speculation by policy makers, academics, and journalists who have suggested that the metaphors we use to talk about important issues shapes how we think about the issues and even how we approach solving them. This research suggests that it is vital that we choose good metaphors to frame social issues. Far from being mere rhetorical flourishes, metaphors may have a profound influence on how we act to deal with important societal issues.

References

Boroditsky, L. (2000). Metaphoric Structuring: Understanding time through spatial metaphors. *Cognition*, *75*, 1-28.

Eliasmith, C. & Thagard, P. (2001). Integrating structure and meaning: a distributed model of analogical mapping. *Cognitive Science, 25,* 245-286.

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

Fodor, J. & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition, 28*, 3–71.

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

Gentner, D., & Clement, C. (1988). Evidence for relational selectivity in the interpretation of analogy and metaphor. In G. H. Bower (Ed.), The psychology of learning and motivation: Advances in research and theory (Vol. 22, pp. 307-358). New York: Academic Press.

Gentner, D., & Gentner, D. (1983). Flowing waters and teeming crowds: mental models of electricity. In Gentner, D., & Stevens, A. (ed.) *Mental Models.* Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Gentner, D., and Markman, A. (1993). Analogy-watershed or waterloo? Structural alignment and the development of connectionist models of analogy. In S.J. Hanson, J.D. Cowan, & C.L. Giles (Eds.), *Advances in Neural Information Processing Systems 5* (pp. 855-862). San Mateo, CA: Morgan Kaufmann Publishers.

Guerrero, R., & Concha-Eastman, A. (2001). An Epidemiological Approach for the Prevention of Urban Violence: The Case of Cali, Colombia. *World Health and Population, 4.*

Hinton, G. (1989). Learning Distributed Representations of Concepts. In R. G. M. Morris (Ed.), *Parallel distributed processing: Implications for psychology and neurobiology* (Chap. 3, pp. 46–61). Oxford: Clarendon Press.

Hummel, J.E., & Holyoak, K.J. (1997). Distributed Representations of Structure: A Theory of Analogical Access and Mapping. *Psychological Review, 104*, 427-466.

James, W. (1904). Remarks at the peace banquet. *The Atlantic Monthly,* 94, 845-847.

Kelling, G. (1991). Crime and metaphor: Toward a new concept of policing. *City Journal, 1*.

Kotlowitz, A. (May 4, 2008). Is urban violence a virus? *New York Times Magazine,* 52-59.

Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. Chicago: Univ. of Chicago Press.

Lakoff, G. (2002). Moral Politics: How Liberals and Conservatives Think. Chicago: University of Chicago Press.

Nersessian, N. (1992). How do scientists think? Capturing the dynamics of conceptual change in science. In Giere, (ed.). Minneapolis: Univ. of MN Press.

Rogers, T. McClelland. (2008). Précis of *Semantic Cognition: A Parallel Distributed Processing Approach. Behavioral and Brain Sciences, 31,* 689-749.

Schön, D. (1993). Generative metaphor: A perspective on problem-setting in social policy. In Ortony, A. (ed.) *Metaphor and Thought*. Cambridge, UK: Univ. Press.