Fostering Metaphorical Creativity Using Computational Metaphor Identification

Eric P.S. Baumer[1], Bill Tomlinson[1], Lindsey E. Richland[1], Janice Hansen[1]

University of California, Irvine

[1]Department of Informatics [2]Department of Education

Irvine, CA, USA

{ebaumer, wmt, lerich, hansenj} @ uci.edu

ABSTRACT

Metaphor is often seen as a mode of creative thinking or as a means of fostering creativity. However, little work has studied creative generation of novel metaphors. This paper explores the use of computational metaphor identification (CMI) to foster creative generation of novel metaphors. CMI is a technique for analyzing textual corpora to identify potential conceptual metaphors. Drawing those metaphors to readers' attention can provide an opportunity to consider alternatives to current metaphors. This paper describes results from a study using CMI to foster metaphorical creativity in the context of science education. The results show that CMI leads to more creative mappings within metaphors. The key contributions of this paper are a demonstration that CMI can be used to foster more original metaphorical reasoning, and, more generally, implications for the study of metaphorical creativity.

Author Keywords

Metaphor, creativity, science education, computational metaphor identification.

ACM Classification Keywords

K.3.1 Computer Uses in Education: Computer-assisted instruction; H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

Experimentation, Human Factors.

INTRODUCTION

Metaphors and the ability to use them are strongly linked to creative thinking processes. Gordon [15] describes how a number of scientific and technological discoveries and inventions, including Gutenberg's printing press, Kepler's laws of planetary motion, Laplace's field equation, Whitney's cotton gin, and Bell's telephone, were all greatly influenced by metaphorical thinking. These achievements

resonate with Poincaré's emphasis on analogical reasoning in the creative process [31]. Drawing on Macbeth's expositions, Lakoff and Johnson describe how the metaphor of life as a tale told by an idiot causes "us to try to understand how it could be true, [and] makes possible a new understanding of our lives" [22:175]. Such metaphors are found not only in literature. Dunbar and others have drawn explicit connections to the role of cross-domain mapping in scientific creativity [6,7,11,18,30].

Metaphorical and analogical thinking play an instrumental role in everyday academic creativity. In a detailed analysis of thousands of hours in a well-respected university biology laboratory, Dunbar found metaphor and analogy throughout all creative aspects of professional science, from design of experiments to interpretation of research results and dissemination language [2,6,7]. Facility with noticing, recognizing, and making use of metaphorical thought is thus a crucial part of developing everyday academic and scientific creativity.

Most previous work on creativity that mentions metaphor presents it either as a useful means for promoting creativity, or as a type of creative thinking [1,15,29]. However, less research has explored novel metaphor generation, potentially because of the difficulty of creating new metaphors. Doing so requires an awareness of current metaphors on which new metaphors can build and from which new metaphors will likely differ. However, "because they can be used so automatically and effortlessly, we find it hard to question [conceptual metaphors], if we can even notice them" in the first place [23:65], thereby impeding the process of new metaphor generation.

This paper explores how one technique, computational metaphor identification (CMI), might be used to help overcome these difficulties and foster creative metaphor generation. CMI can be used to analyze large textual corpora and identify potential conceptual metaphors at work in that text. The purpose of this technique is not to state definitively *the* metaphors being used in a text, but rather to identify potential metaphors, as well as the patterns of language serving as evidence thereof, and to draw those patterns and metaphors to readers' attention. These computationally identified metaphors may help increase awareness of current conceptual metaphors and thereby scaffold the process of generating new metaphors.

Thus, this work represents an inversion of more classical approaches in artificial intelligence. Rather than using people as a model to build metaphorical creativity into computers, we instead seek to use computers as a means to promote metaphorical creativity in people.

This paper presents results from a study in which CMI was incorporated into a 7th grade (ages 12-13) science education module on cell biology. Students' writing was analyzed for potential conceptual metaphors. Identified metaphors were presented back to the students to determine potential impacts on students' metaphor generation. The results presented here serve as a contribution not only in terms of understanding the types of metaphorical creativity CMI can be used to foster, but also in terms of working toward a more general understanding of metaphorical creativity.

RELATED WORK

Conceptual Metaphor Theory
The research presented here is informed largely by the work of Lakoff and colleagues [21,22,23], which views metaphor not as a literary or poetic device but rather as a fundamental aspect of human cognition. For example, when discussing money, we might say "he *poured* money *into* his savings account," "they *froze* my assets," or "capital freely *flowed* between investors.*"* Lakoff and Johnson [22] claim that such linguistic patterns evidence the conceptual metaphor MONEY IS A LIQUID, i.e., that we understand the abstract concept of money in terms of our physical experiences with liquids. We use words from our knowledge of liquids to talk about money because the cognitive structure of the metaphor "sanctions the use of source domain language and inference patterns for the target domain" [21:208]. This is not to say that conceptual metaphor is a linguistic phenomenon. Rather, the linguistic patterns serve as evidence for the cognitive phenomenon.

A key aspect of conceptual metaphor theory is that many different metaphors may be used to frame the same concept, a phenomenon referred to as metaphorical pluralism. For example, a wide variety of metaphors can be used to frame the concept of *love*, such as LOVE IS A JOURNEY: "this relationship is[n't] *going anywhere*"; LOVE IS MADNESS: "I'm just *wild* about Harry"; or LOVE IS MAGIC: "she is *bewitching*" [22:44,49]. Each metaphor highlights certain aspects of the concept or situation, while downplaying others. Lakoff and Johnson argue that "successful functioning in our daily lives seems to require a constant shifting of [many] metaphors ... that are inconsistent with one another ... to comprehend details of our daily existence" [22:221]. Moreover, suggestion of an alternative, novel metaphor can provide a different conceptualization that highlights different aspects of the situation, can "cause us to try to understand how [the novel metaphor] could be true, [and] makes possible a new understanding of our lives" [22:175].

However, Lakoff's is not the only account of conceptual metaphor. While, the various approaches are too numerous to review fully here, one line of relevant research is that of structure mapping theory [10], which is a theoretical account of the process of analogical reasoning. Structure mapping makes a key distinction between surface similarity and structural similarity, where surface similarity deals with common attributes of two objects and structural similarity maps complex relationships between sets of objects. For example, in an analogy between the solar system and an atom, the rotational and gravitational relationships between the sun and planets may be mapped on to the orbital and electromagnetic relationships between the nucleus and electrons, a mapping which involves structural attributes. On the other hand, comparing the round shape of the sun and the round shape of the nucleus would involve mapping surface attributes. Gentner et al. [12] summarize work demonstrating the importance of structural relationships in metaphor comprehension, and that structural similarities are more important the surface similarities in making valid and informative analogical inferences (see also [14,17]). This research is relevant to the present work in terms of examining novel metaphors for whether they draw primarily on surface or structural similarity.

Computational Approaches to Metaphor
A significant amount of research has been done on various computational approaches to metaphor, e.g., [8,25,27]. Most such work treats metaphor as a hurdle to overcome during text processing, employing computational methods of differentiating literal text from figurative, then applying special processing to that figurative text in order to infer its literal meaning. For example, one application, designed as a help system for the UNIX command line, would determine that statements such as "how do I get into Lisp?" meant the user wanted to know how to invoke Lisp programming environment [27]. One exception to this trend is CorMet [28], which uses domain-specific textual corpora to extract known conceptual metaphors. For example, CorMet was used to extract the metaphor MONEY IS LIKE A LIQUID by using corpora from the domain of the Laboratory and the domain of Finance. In the Laboratory corpus, verbs such as "pour," "flow," "freeze," and "evaporate" are all associated with words for the concept of *liquid* or *fluid*. In the finance corpus, these same verbs are associated with words for the concept of *money*, including "money," "funds," and "assets." The technique used in this study draws largely on CorMet but extends that work in two important ways. First, CorMet was designed to extract known conventional metaphors, whereas this work involves identifying potential metaphors in relatively arbitrary corpora. Second, little work has explored using such computationally identified metaphors to promote critical thinking about, and creative generation of, metaphors.

Metaphor in Science Education
Conceptual metaphors are well known to play an important

role in many aspects of scientific thinking [5,6,7,13]. In a germinal study, Gentner and Gentner [13] examine the impacts of two different metaphors for electrical circuits, either that of water flowing through pipes or a crowd of people running around a track. Their results demonstrate that consistent use of a single metaphor leads to consistent inferences about series and parallel circuit configurations. Taking a different focus, Cameron [5] explores metaphor comprehension using think-aloud exercises. Specifically, she examines students' understanding of the metaphors THE ATMOSPHERE IS A SHIELD and THE ATMOSPHERE IS A BLANKET, arguing that the lack of specificity and explicitness of these metaphors lead to various misconceptions, and that instruction using such metaphors should include explicit source-to-target mappings.

These and other studies raise, albeit somewhat implicitly, an important aspect of metaphor not often emphasized in the educational literature. While Gentner and Gentner [13] demonstrate important results about analogical reasoning, one important aspect is noted but not examined thoroughly: of the 36 high school and college students in one of their studies, only 15 subjects consistently used a single metaphor. Thus, 21 subjects, over half those recruited, either did not employ the metaphor correctly or employed more than one metaphor. "The responses of subjects who were inconsistent in their use of models were analyzed separately and are not reported" [13:117], so it is unclear how many were screened for errors vs. multiple models. While Gentner and Gentner describe these students' metaphor use as "inconsistent," such inconsistency may lead to reduced inferential errors and actually indicate a better grasp of the material.

Furthermore, the use of multiple metaphors is reminiscent of the notion of metaphorical pluralism [22], wherein different metaphors for the same situation highlight certain aspects while downplaying others. Just as with the examples from Gentner and Gentner's study [13], these different metaphors may often conflict with one another. Cameron [5] emphasizes how multiple metaphors may aid in students' comprehension by helping them determine which are the important aspects to be mapped from various source concepts. Encouraging students to consider multiple alternative metaphors, as well as to generate their own creative, novel metaphors, may ultimately help reduce misconceptions and facilitate learning.

Creativity and Metaphor
The diverse literature on creativity contains little consensus on understanding, modeling, assessing, or even defining creativity, not to mention the varying roles of the creative process, the creative product, the creative person, and the creative situation. Despite these differences, many approaches see novelty and usefulness as pivotal attributes [cf. 4]. While some standard measures of creativity exist (see [1] for a review), Barron and Harrington [1] argue that

creativity may likely be domain-specific, such that different means should be used to assess creativity in different domains. Most previously developed standard measures related to metaphor comprehension or reasoning rather than with assessing the creativity of generated metaphors. Gardner and Winner [9] describe a number of studies analyzing children's use of metaphor, one of which involves metaphor formation. However, their analysis focuses on the appropriateness in terms of metaphors being non-literal and unconventional yet relevant. Kolb et al. [20] explored the possibility of using computationally generated metaphors in the design process, but their study does not include a standard metric for evaluating metaphorical creativity.

One standard measure of potential relevance is Guildford's alternative uses task [16], wherein subjects are asked to name all the uses they can for a common household object, such as a brick or a piece of paper. Responses are scored on a number of criteria, including originality, in terms of the number of other subjects who provided the same response, and elaboration, in terms of the amount of detail in a response. These two criteria resonate with the novelty aspect of creativity, and may be useful in assessing the creativity of an original metaphor. Novelty may be easily assessed similarly to the alternative uses task [16] by simply counting the number of subjects who offered a given response. Assessing usefulness, however, is rather more complex, and may include the examining of the mappings involved the metaphor, the justifications for those mappings, and the inferences sanctioned by them. The assessment of metaphorical creativity used in the study presented here is described in more detail in the methods section below.

COMPUTATIONAL METAPHOR IDENTIFICATION
While space precludes a fully detailed description of the algorithms involved, this section provides a high-level summary of the techniques employed in computational metaphor identification (CMI), which extends previous work in computational linguistics [28].

Metaphors are conceptual mappings wherein a concept from a source domain partially structures the understanding of a concept from a target domain. In the above example, MONEY IS A LIQUID, the target concept *money* is partially framed in terms of the source concept *liquid*. CMI begins by gathering corpora for the source and target domains. In this paper, the target corpus is science students' writing, described in more detail in the methods section below. For the source corpora, we use Wikipedia articles, as they provide a readily available, categorically organized, large source of content on a wide variety of topics. A source corpus for a given domain consists of all the Wikipedia articles in the category for that domain, as well as all articles in its subcategory. All documents in the source and target corpora are automatically parsed to extract sentence structure and grammatical relationships [19,26].

The crux of CMI is selectional preference learning [32], which identifies the tendency of particular words to appear with certain other classes of words in specific grammatical relationships. For example, words for the concept of *food* are often the direct object of the verb "eat." Using the parsed documents, CMI calculates selectional preferences of the characteristic nouns in a corpus, where characteristic means that the noun is highly frequent in the corpus relative to its frequency in general English. Selectional preference is quantified as the relative entropy of the posterior distribution conditioned on a specific noun and case slot with respect to the prior distribution of verbs in general English:

$$S(c) = \sum_v P(v|c) \log \frac{P(v|c)}{P(v)}$$

where *c* is a class of nouns (i.e., a concept like *food)* and a grammatical case slot (such as direct object), and *v* ranges over all the verbs for which *c* appears in the given case slot. While selectional preference captures the "choosiness" of a particular grammatical relationship, selectional association measures the degree to which that grammatical relationship is associated with a particular verb:

$$\lambda(c,v) = \frac{1}{S(c)} P(v|c) \log \frac{P(v|c)}{P(v)}$$

Selectional associations are calculated for classes of words, but the corpora consist of words that may represent many possible classes of nouns. Thus, individual nouns count as partial observations of each class of words that they might represent using WordNet, an ontological dictionary. For example, the words "water," "liquor," and "ammonia" can all represent the concept of *liquid*, as liquid is a parent, or hypernym, of each. WordNet uses synsets (sets of synonyms) to represent classes of words. For example, the synonyms "liquid," "liquidness," "liquidity," and "liquid state" comprise the synset for the liquid state of matter. These word classes are then clustered using two-nearest-neighbor clustering based on the verbs for which they select. Each cluster represents a coherent concept in the corpus.

This approach of using clustered hypernyms resonates with Lakoff's argument that metaphorical mappings occur not at the level of situational specifics, but at the superordinate level. For example, in the metaphor LOVE IS A JOURNEY, the relationship is a *vehicle*. Although specific instantiations of the metaphor may frame that vehicle variously as a train ("off the track"), a car ("long, bumpy road"), or a plane ("just taking off"), "the categories mapped will tend to be at the superordinate level rather than the basic level" [21:212]. This method of counting each word observed as a partial observation of each of the synsets it might represent causes observations at the basic level to accumulate in the superordinate levels they collectively represent.

To identify metaphors, CMI looks for correspondences between conceptual clusters in the source and target corpora. For example, in the LIQUID corpus, the cluster for *container* would select to be the object of the preposition "into" with the verb "flow," the preposition "from" with the verb "flow," the subject of the verb "hold," and so on. In documents about banking or finance, the cluster for *institution*, as in a bank or other financial institution, also selects for those same verbs in the same grammatical relationships. Based on the similarity of these selectional associations, each mapping is given a confidence score to indicate how likely the linguistic patterns are to evidence a conceptual metaphor. One of the strengths of CMI is that it works in the aggregate. While individual instances of phrases like "flowed from the Federal Reserve" and "poured money into my IRA" may not at first glance appear metaphorical, it is the systematicity of these patterns that becomes compelling evidence for the existence of a metaphor.

An important aspect of CMI is that it identifies only linguistic patterns potentially indicative of conceptual metaphors, not the metaphors themselves. As mentioned above, Lakoff [21] emphasizes that metaphor is primarily a cognitive phenomenon, and that metaphorical language serves as evidence for the cognitive phenomenon. CMI leverages computational power to search through large bodies of text to identify patterns of potential interest, then presents those patterns to a human user along with the potential metaphors they might imply to foster critical thinking about metaphor. To reiterate, this places the job of finding patterns in the hands of the computer, and the job of interpreting those patterns in the hands of the human user.

FOSTERING METAPHORICAL CREATIVITY
As described above, metaphor plays in integral role in scientific thought, both in science practice [2,6,7] and science education [5,13]. This section describes a study of incorporating CMI into a 7th grade science classroom lesson on cell biology using the WISE system [http://berkeley.wise.edu]. To help scaffold students' understanding of what a metaphor is, the module included the metaphor A CELL IS A CITY. This metaphor was chosen because it is a complex structural metaphor [cf. 22] involving many component mappings and potential inferences, it was commonly used by teachers whose students participated in the study (more on participants below), it is invoked in the textbook used in the school district in which the study was conducted [omitted], it draws on a source domain with which most students are familiar, and the target concept of a cell allows for many creative alternative metaphors. Those alternative metaphors, as described here, are the locus of this analysis.

Participants
Students in this study came from classes taught by two science teachers at two different suburban middle schools.

Recruitment was done through the teachers, who were contacted via targeted emails to participants in previous studies, informal social connections, and snowball sampling. In total, three teachers responded with interest in the study. Phone conferences were held with each teacher, which led to the referral of a fourth teacher. One teacher was not able to participate due to logistical constraints of working at a year-round school, and another only taught physical sciences, resulting in two teachers whose students participated in the study. Ultimately, 355 students participated in the study, 136 from Teacher A's classes and 217 from Teacher B's. No demographic information was collected during this study.

Methods
The activities in which students engaged took place during four different days over the course of one week. Day 1 involved two activities. The first activity introduced the concept of a cell, the concept of a metaphor, and the metaphor A CELL IS A CITY[1]. This activity also included prompts asking students to describe what they already knew about cells from previous instruction. The second activity described each of six organelles, listed in Table 1 along with their mapping in the city metaphor. Students then completed a "minitest," wherein they were asked to describe the function of each organelle.

Organelle	City Mapping
nucleus	city hall
mitochondrion	power plant
ribosome	factory
endoplasmic reticulum	industrial zone
Golgi body	post office
lysosome	recycling plant

Table 1: Organelle correspondences in the A CELL IS A CITY metaphor.

The next two days were spent analyzing students' written responses to questions from Day 1 using CMI, during which time students were either not in class (Teacher A) or were engaged in alternate instruction (Teacher B). To facilitate CMI analysis, students' responses were spellchecked, and some pronouns were replaced with their antecedents. Specifically, many students began their answers with "it," e.g., "It tells the cell what to do," which was replaced with the organelle about which they were being asked, e.g., "The nucleus tells the cell what to do." Three source domains were used: CITIES, using Wikipedia's Cities category; BUILDINGS, using the Buildings_and_structures category; and FOOD AND DRINK, using the Food_and_drink category. CITIES was

[1]When presented to students, metaphors used "is like," as in, "a cell is like a city," to aid comprehension [3].

used because of the explicit A CELL IS A CITY metaphor in the module, BUILDINGS was selected because cities are composed largely of buildings, and FOOD AND DRINK was chosen as a source domain with which students would likely have rich, experiential knowledge on which to draw.

A full analysis of the computationally identified metaphors in the students' writing is beyond the scope of this paper. Instead, the focus here is on three metaphors, one from each of the above source domains, that were chosen to be included in Day 2 of the module. Each of these metaphors was in the upper one percentile of the metaphors from its domain in terms of confidence, and each metaphor was the strongest for the concept it framed. From CITIES, the metaphor A GOLGI BODY IS LIKE A PORT was chosen, as it was functionally similar enough to the post office metaphor students had seen during the module that it would be readily comprehensible, yet different enough to prompt consideration of the ways in which each of the two metaphors might fit or not fit. From BUILDINGS, the metaphor ORGANELLES ARE BUILDINGS was chosen, because for many organelles it fits with the metaphor A CELL IS A CITY, but for other organelles, such as the endoplasmic reticulum, the buildings metaphor does not fit very well, providing room for generating better-fitting alternatives. From FOOD AND DRINK, the metaphor ENERGY IS FOOD was chosen, partly because it used a very familiar, concrete concept to frame a very abstract one, and partly because it did not fit will with the city metaphor, providing an opportunity to explore how students might grapple with metaphorical pluralism [cf. 22].

Students were split evenly into a CMI condition and a non-CMI condition. During Day 2 of the module, students in the non-CMI condition watched a computer animated video of a cell synthesizing protein and were asked to list the organelles they recognized and describe what each was doing. Students in the CMI condition were presented with the above listed metaphors, noting that each had been identified in their writing by a computer, along with a description of how the metaphor might fit. For example, for A GOLGI BODY IS LIKE A PORT, students were told, "Many people used words like 'transport,' 'send,' 'carry,' and 'move' with the Golgi body, which are words that are often used with ports or harbors," based on the verb-case slots mediating the metaphor. These descriptions were accompanied with paraphrased example sentence fragments from their writing and from Wikipedia, for example, " '*transporting* protein *to* the *Golgi body*' might be like '*transporting* goods and products *to* a *port*'." Students were then asked two questions about each metaphor. First, they were asked if the metaphor made sense to them and if it was similar to their thinking. Second, they were asked to name at least two ways in which the metaphor did *not* fit, e.g., two ways in which a Golgi body might *not* be like a port. The other two metahpors (ORGANELLES AND BUILDINGS and ENERGY IS FOOD) were presented similarly.

All students were then asked to generate an alternative metaphor for a cell. Students were first asked what a cell could be like, other than a city, as well as why a cell might be like that thing. Subsequent questions asked students to describe what each organelle would be like in their new metaphor and why. These "why" questions help examine the types of metaphorical inferences students are making when generating novel metaphors, as well as provide a means of assessing the aptness (as defined below) of the overall cell metaphor through students' ability to explain the component mappings thereof.

Analysis

The analysis of students' metaphors focuses on three aspects thereof: the overall stated metaphor, the mappings given for the organelles within that metaphor, and the justifications used for those component mappings.

Creativity is often defined in terms of novelty and usefulness [cf. 4]. Thus, to we assess students' metaphors through their novelty and usefulness. Each student's metaphor was labeled with a one- or two-word description. For example, the responses "human body," "a cell is like a human," and "cells are like people or animals" were all labeled "body." Similarly, "A cell is like a car," "A cell is like a boat," and "A cell could be like an airplane" were all labeled "transportation." The label "other" was used for seemingly metaphorical but nonsensical replies, e.g., listing metaphors for individual organelles that were not covered in the module and did not form a coherent metaphor for the cell. The label "none" was given if students either supplied no metaphor or stated that they could not think of one. A summary of all labels used is presented below. Novelty here is operationalized in terms of uniqueness with respect to all generated metaphors. Specifically,

$$U_m = \frac{(1 - mPercent) - min(U)}{(max(U) - min(U))}$$

where U_m is uniqueness for metaphor m, $mPercent$ is the percent of students who responded using metaphor m, $max(U)$ is the highest uniqueness, and $min(U)$ is the lowest uniqueness. This approach normalizes uniqueness, such that the most common metaphor has uniqueness of 0 and the rarest metaphors have uniqueness of 1.

The term "usefulness commonly implies usefulness for some purpose, but it's unclear for exactly what purpose students' novel metaphors should be assessed. Instead, we consider the aptness of students' metaphors based on the mappings that they list for the organelles and justifications they give for those mappings. Specifically, a mapping is apt if it both fits with the overall metaphor and relies on functional rather than featural similarities. For example, if the overall metaphor is that A CELL A BODY, then A NUCLEUS IS THE BRAIN fits, but THE GOLGI BODY IS A POST OFFICE does not. The distinction between featural and functional comparisons is informed by work on the role of surface and

structural similarity in analogy [10,12,14,17], where structural similarity is more important for analogical reasoning. Here, we take structural aspects to refer to the functions of the organelles and surface aspects to refer to their features. For example, saying that the nucleus is like the brain because the brain controls the body and the nucleus controls the cell is a functional comparison, whereas saying the nucleus and brain are similar because they are both purple and wrinkly is making a featural comparison. Students' justifications were coded as to whether they used functional arguments, featural arguments, or neither. For each organelle, inter-rater reliability was established between two coders on 30% of the data set using Cohen's kappa, with $\varkappa > 0.8$.

Uniqueness and aptness can be combined to derive an overall creativity score. Here, we give uniqueness and aptness equal weighting:

$$C_m = \frac{U_m + A_m/6}{2}$$

where C_m is overall creativity, U_m is defined above, A_m is aptness in terms of the number of organelles for which apt mappings are provided.

Students' justifications were also coded for whether or not they invoked the city metaphor, which could occur in two ways. First, a student might give a correspondence from the city metaphor for an organelle. For example, if the overall metaphor is that A CELL IS A COUNTRY, the student may say that the ribosomes are factories. Second, a student may reason through the city metaphor when justifying her or his correspondence. For example, with the overall metaphor A CELL IS A SCHOOL, one student said the Golgi body is like the secretary "because the secretary puts mail into the teachers' boxes ... like a post office." Here, even though the mapping given for Golgi body is secretary, the student invokes the city metaphor to justify that mapping. For each organelle, inter-rater reliability was established between two coders on 20% of the data set using Cohen's kappa, with $\varkappa > 0.8$.

Results

Overall, CMI did lead to more original metaphors. Specifically, students in the CMI condition relied less on the city metaphor for the individual items within the larger metaphor. In addition, was associated with more unique metaphors and, somewhat surprisingly, CMI was also associated with less apt metaphors; however, neither the difference in uniqueness nor the difference in aptness was statistically significant. Invocation of the city metaphor was also inversely correlated with more apt metaphors, i.e., students who relied more on the city metaphor generated less apt mappings within their new metaphor. Finally, an inverse correlation was found between uniqueness and aptness, reinforcing the simultaneous requirement of both these criteria when assessing overall creativity. This section addresses these results in more detail. In general, the

analyses in this section only include those students who completed the CMI activity and who provided a new metaphor for the cell (i.e., their metaphor was not labeled as "other" or "none"). These data consist of responses from a total of 243 students. Statistical significance is denoted with asterisks: * is $p < 0.05$, ** is $p < 0.01$, and *** is $p < 0.001$.

An important aspect of metaphorical creativity is the ability to generate novel yet appropriate mappings for a new metaphor, as well as suitable justifications for those mappings. As described above, students' mappings and justifications for each of the six organelles were coded for whether or not they reverted to the city metaphor. Table 2 describes the differences between the CMI and non-CMI groups in the average number of organelles for which the city metaphor was invoked. Those students who saw the computationally identified metaphors were significantly less likely to use the city metaphor, instead generating more creative mappings for the organelles rather than simply reverting to those presented in the module.

	CMI	non-CMI	p-value
city metaphor	0.853	1.298	0.0221*

Table 2: Average number of organelle mappings for which the city metaphor was invoked.

Students provided a wide variety of new metaphors for the cell, as shown in the histogram in Figure 1. The single most common metaphor was that of a human or animal body. Several uncommon metaphors were only given in the CMI condition, including castle, tree/plant, courtroom, and pizza. Some uncommon metaphors were also given only in the non-CMI condition, including atom and monastery.

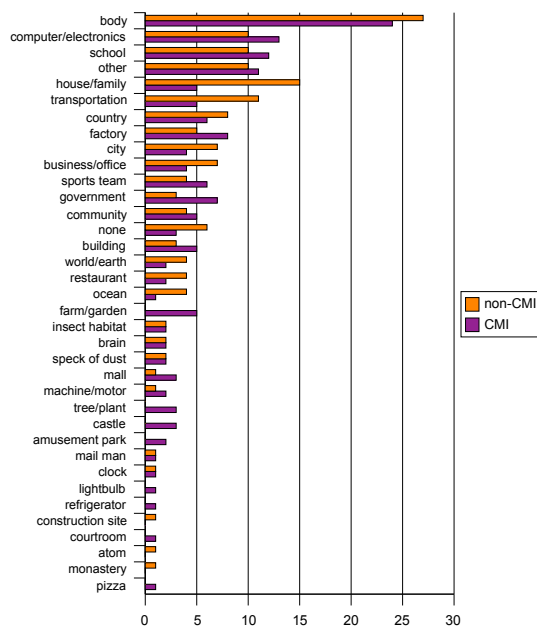Table 3 describes the differences between the metaphors

Figure 1: Histogram of all new cell metaphors.

used by the two groups; p-values are from Fisher's test rather than a χ^2 test, because many of the cells in the contingency table contained few or no observations. Overall, there was not a statistically significant difference between the new metaphors. The only individual metaphor that significantly differed was "house/family," which was more common in the non-CMI condition. It may be possible to conjecture why this metaphor and no others differed significantly (e.g., comparison between a house/family and a city), but the data are not conducive to such exploration.

	CMI	non-CMI	p-value
overall	n/a	n/a	0.102
house/family	1.23%	5.76%	0.00417 **

Table 3: Differences in the new cell metaphor students gave.

As described above, uniqueness of generated metaphors is one potential means of assessing metaphorical creativity. Since students in the CMI condition relied less on the city metaphor, one might assume that their overall metaphors were more creative. Table 4 shows that those students in the CMI condition did generate more unique metaphors, but the difference in uniqueness score was not statistically significant.

	CMI	non-CMI	p-value
uniqueness	0.652	0.599	0.200

Table 4: Average uniqueness scores for new cell metaphor.

The other above-described criterion for metaphorical creativity is aptness, i.e., for how many of the six organelles students both generated mappings that fit with their new cell metaphor and used functional justifications for those mappings. Table 5 lists differences between the two groups in terms of the average number of organelles for which apt mappings were given, the average number of organelles that fit with the overall new metaphor, and the average number of organelles for which functional justifications were provided. Interestingly, the CMI condition resulted in fewer apt mappings than the non-CMI condition, although this difference was not statistically significant. Potential implications of the contrast between uniqueness and aptness are considered further below.

	CMI	non-CMI	p-value
aptness	3.147	3.518	0.185
fit	3.713	3.982	0.367
functional	4.403	4.684	0.169

Table 5: Average number of organelles for which apt mappings were given, that fit with the new metaphor, and for which functional justifications were provided.

Scores for aptness and uniqueness were combined using the above-described formula to derive overall creativity of students' metaphors. Table 6 shows differences between the overall creativity of the two groups' metaphors. We can see that the non-CMI group was slightly more creative, as the greater aptness of the non-CMI group slightly outweighed the greater uniqueness of the CMI group. However, this difference was very small and not significant.

	CMI	non-CMI	p-value
creativity	0.588	0.593	0.882

Table 6: Average creativity of new cell metaphor.

These results indicate that CMI leads to increased metaphorical creativity, but not in the ways expected. Overall metaphors were not significantly more unique or more apt in the CMI condition, but the component mappings relied less on the prior-learned city metaphor. Thus, we next examine correlations of invoking the city metaphor to aptness and uniqueness.

Table 7 lists Pearson correlation coefficients across all participants of use of the city metaphor with aptness and uniqueness. First, we see a negative correlation with aptness, meaning that use of the city metaphor is associated with less apt mappings for the organelles. Looking at the data, we often see a student stating a new overall metaphor but then reverting to the city metaphor for some of the organelles, e.g., one student responded that "a cell could be like a person" and "the nusleus [sic] is like the brain," but then said that the Golgi body "is like the post office." Furthermore, since students in the CMI condition invoked the city metaphor significantly less often, and since invocation of the city metaphor is associated with reduced aptness, it is somewhat surprising that the CMI condition also led to reduced aptness (though not significantly). Second, we also see a positive and even stronger correlation with uniqueness, such that increased use of the city metaphor is associated with increased uniqueness. This result is also somewhat surprising, since the CMI condition led to decreased use of the city metaphor but increased uniqueness (though not significantly). Examining the data anecdotally suggests that some students generated highly unique metaphors for the cell but then reverted to the city metaphor, potentially because of the difficult of finding apt mappings within those unique overall metaphors. Further implications are addressed in the discussion section below.

Correl w/ city	r	p-value
aptness	-0.194	0.00120 **
uniqueness	0.224	0.000219 ***

Table 7: Correlations of use of the city metaphor with aptness and uniqueness.

The difference in these two correlations, particularly that one is negative and one is positive, leads to a question of the relationship between uniqueness and aptness. Informed by previous research on creativity, this analysis took the combination of uniqueness and aptness as a means of assessing the creativity of students' novel metaphors. As Table 8 shows, there is a strong inverse correlation between aptness and uniqueness, i.e., highly unique metaphors are rarely highly apt, and vice versa. This result has important implications for the study of metaphorical creativity in general, as addressed below in the discussion section.

Correl w/ unique	r	p-value
aptness	-0.153	0.00833 **
fit	-0.079	0.110
functional	-0.138	0.0159 *

Table 8: Correlations between uniqueness and aptness.

Discussion

These results demonstrate that those students in the CMI condition exhibited more of novelty, one indicator of creativity, during the process of generating mappings for their metaphors. Specifically, they relied less on the A CELL IS A CITY metaphor that they had learned during the instructional module. Previous research has suggested that using multiple metaphors may increase students' understanding [5,13]. The results presented here align with those conclusions, in that exposure to the computationally identified metaphors increased the creativity of students' metaphorical reasoning and metaphorical mapping skills.

On the other hand, the overall metaphors that students supplied were not significantly more creative in terms of aptness and uniqueness in the CMI condition. This result is not necessarily surprising. Some studies have argued that creativity is linked more closely to self-efficacy than any given intervention [e.g., 24]. Thus, while CMI may have increased the creativity with which students reason about the mappings within metaphors, it did not in this study have a significant impact on creativity of overall metaphors for the cell.

The results contained a number of potentially compelling insights about the relationship between aptness and uniqueness, specifically an inverse correlation between the two. This correlation may, in part, explain the result that the metaphors in the CMI condition were slightly more unique but were also slightly less apt. This result also reinforces the combined criteria of uniqueness and aptness for assessing metaphorical creativity.

Much of the previous work on creativity cited here describes novelty and usefulness (operationalized here as uniqueness and aptness) as important core attributes of creativity. The most creative metaphors should be highly apt and highly unique, a combination which one would expect to be rare, as was the case with these data. Either

students generated highly unique metaphors and then failed to find apt mappings for many of the organelles, or they generated less unique metaphors with highly apt mappings.

When assessing metaphorical creativity, these results suggest that uniqueness and aptness may be partially at odds with one another. This opposition may be due in part to a conflict between the divergent and convergent thinking (cf. [1,4]) involved in each, respectively. Generating a novel metaphor requires divergent thinking, but the more divergent the metaphor the more difficult it will likely be to find suitable component mappings, a task which requires convergent thinking. Furthermore, the variety and complexity of the functions and relationships within a cell constrain somewhat the possible source domains, such that if one insists on choosing an apt metaphor, uniqueness will likely be limited from the outset. The following section on future work considers what other means might be available for assessing metaphorical creativity.

FUTURE WORK
The results presented here offer a view of how CMI may be used to foster metaphorical creativity, but they also point to important future directions for study. First, the CMI intervention had a number of aspects, including the questions asked, the format of the presentation, the larger educational context etc. However, it would be beneficial to know which of these had the most significant impact on students' metaphorical creativity. Was it the specific metaphors chosen? Would other computationally identified metaphors have similar effects? Would using more example sentence fragments and/or showing more mediating verb-case slots alter the impact? Do computationally identified metaphors differ significantly from randomly generated potential metaphors? These and other questions should be addressed to determine which components of the CMI intervention are associated with what aspects of increased creativity.

It could also be informative in future studies to compare metaphorical creativity along gender, ethnicity, or other lines to average performance differences at this age group to determine how well metaphorical creativity aligns with other science skills.

As mentioned above, it may be useful to develop alternative means of assessing a novel metaphor's aptness. For example, novel metaphors could be given to other students to rate; after generating a novel metaphor, students could be required to extend it by addressing an organelle about which they had not previously learned; metaphors could be given to students' teachers, or perhaps to professional biologists, for assessment. Each of these has potential benefits and drawbacks, both theoretically and practically, that should be explored empirically, potentially comparing the aptness arrived at via different methods of assessment. Furthermore, in each case, the resulting relationship between aptness and uniqueness should be examined to determine if a negative correlation occurs.

Lastly, this work has explored the use of CMI to foster creative metaphor generation specifically in the context of science education. However, conceptual metaphors permeate many arenas of human thought and behavior. An important extension of this work is applying CMI to other domains—literature, politics, engineering, religion, etc.—to examine not only what sorts of potential metaphors can be computationally identified, but the impact that exposure to those computationally identified metaphors can have on metaphorical creativity.

CONCLUSION
This paper has explored the use of computational metaphor identification (CMI) to foster creative metaphor generation in the context of middle school science education. The results demonstrate that CMI led to increased creativity in the mappings involved in the metaphor, in that students relied significantly less on the A CELL IS A CITY metaphor they learned during the instructional module. However, in terms of uniqueness and aptness of students' new metaphors for a cell, there were not significant differences between those students who saw computationally identified metaphors and those who did not. Furthermore, the inverse correlation found between aptness and uniqueness suggests that the simultaneous requirement of both aptness and uniqueness, while not often met, may likely be an effective method for assessing metaphorical creativity. This paper presents an important contribution to the study of creativity in two regards. First, it demonstrates that computational metaphor identification can be used to increase creativity in finding mappings during the metaphor generation process. Second, it provides a study of metaphorical creativity in its own right. That is, while thinking *with* metaphors is well established as an important aspect of creativity in many domains, thinking *about* metaphors is less studied and less understood. This paper draws attention to the thinking about metaphors as a viable area for research and makes an important contribution to understanding metaphorical creativity by presenting a means of evaluating and examining aspects of the process of creative metaphor generation.

REFERENCES
1. Barron, F. and Harrington, D. Creativity, Intelligence,

and Personality. *Annual Review of Psychology 32*, (1981), 439-476.

2. Blanchette, I. and Dunbar, K. Analogy use in Naturalistic settings: The influence of audience, emotion and goals. *Memory and Cognition 29*, (2001), 730-735.

3. Bowdle, B. and Gentner, D. The Career of Metaphor. *Psychological Review 112*, 1 (2005), 193-216.

4. Brown, R.T. Creativity: What are We to Measure. In J.A. Glover, R.R. Ronning and C.R. Reynolds, eds., *Handbook of Creativity*. Springer, Berlin, 1989.

5. Cameron, L. Metaphors in the Learning of Science: a discourse focus. *British Educational Research Journal 28*, (2002), 673-688.

6. Dunbar, K. How scientists really reason: Scientific reasoning in real-world laboratories. In *Mechanisms of insight*. MIT Press, Cambridge, MA, 1995, 365-395.

7. Dunbar, K. The Scientist InVivo: How scientists think and reason in the laboratory. In *Model-based reasoning in scientific discovery*. Plenum Press, New York, 1999.

8. Fass, D. Met*: A method for discriminating metonymy and metaphor by computer. *Comp Ling 17*, 1 (1991), 49-90.

9. Gardner, H. and Winner, E. The Development of Metaphoric Competence: Implications for Humanistic Disciplines. *Critical Inquiry 5*, 1, Special Issue on Metaphor (1978), 123-141.

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

11. Gentner, D. Analogy in scientific discovery: The case of Johannes Kepler. In *Model-based reasoning: Science, techonlogy, values*. Kluwer Academic / Plenum, New York, 2002, 21-39.

12. Gentner, D., Bowdle, B., Wolff, P., and Boronat, C. Metaphor Is Like Analogy. In D. Gentner, K.J. Holyoak and B. Kokinov, eds., *The Analogical Mind: Perspectives from Cognitive Science*. MIT Press, Cambridge, MA, 2001, 199-253.

13. Gentner, D. and Gentner, D. Flowing waters or teeming crowds: Mental models of electricity. In D. Gentner and A. Stevens, eds., *Mental Models*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1983, 99-129.

14. Gick, M. and Holyoak, K.J. Schema induction and analogical transfer. *Cognitive Psychology 15*, (1983), 1-38.

15. Gordon, W. Some source material in discovery-by-analogy. *Journal of Creative Behavior 8*, (1974), 239-257.

16. Guilford, J. *The Nature of Human Intelligence*. McGraw-Hill, New York, 1967.

17. Holyoak, K.J. and Koh, K. Surface and Structural Similarity in Analogical Transfer. *Memory and Cognition 15*, 4 (1987), 332-340.

18. Hummel, J.E. and Holyoak, K.J. Analogy and creativity: Schema introduction in a structure-sensitive connectionist model. In T. Dartnall, ed., *Creativity, Cognition, and Knowledge: An Interaction*. Praeger, Westport, CT, 2002, 181-210.

19. Klein, D. and Manning, C. Accurate Unlexicalized Parsing. *Mtg of the Assoc for Comp Ling*, (2003).

20. Kolb, E.M., Hey, J., Agogino, A., and Sebastian, H. Meta4acle: Generating Compelling Metaphors for Design. *Proc of Amer Soc of Mech Engr (ASME), Design Theory and Methodology Conf*, (2008).

21. Lakoff, G. The contemporary theory of metaphor. In A. Ortony, ed., *Metaphor and thought, 2nd. ed.* Cambridge Univ Press, New York, 1993, 202-251.

22. Lakoff, G. and Johnson, M. *Metaphors We Live By*. University of Chicago Press, Chicago, IL, 1980.

23. Lakoff, G. and Turner, M. *More Than Cool Reason: A Field Guide to Poetic Metaphor*. University of Chicago Press, Chicago and London, 1989.

24. Levy, L.H. Originality as Role-Defined Behavior. *Journal of Personality and Social Psychology 9*, 1 (1968), 72-78.

25. Lu, X. and Feldman, A. Workshop on Computational Approaches to Figurative Language. *Human Lang Tech / Conf of the North Amer Chapt of the Assoc for Comp Ling (HLT/NAACL 2007)*, (2007).

26. de Marneffe, M., MacCartney, B., and Manning, C. Generating Typed Dependency Parses from Phrase Structure Parses. *Lang Res and Eval (LREC)*, (2006).

27. Martin, J. *A Computational Model of Metaphor Interpretation*. Academic Press, San Diego, CA, 1990.

28. Mason, Z. CorMet: a computational, corpus-based conventional metaphor extraction system. *Comp Ling 30*, 1 (2004), 23-44.

29. Oxman-Michelli, W. *Critical Thinking as Creativity*. Montclair State, Institute for Critical Thinking, Montclair, NJ, 1991.

30. Penn, D.C., Holyoak, K.J., and Povinelli, D.J. Darwin's Mistake: Explaining the Discontinuity Between Human and Nonhuman Minds. *Behavioral and Brain Sciences 31*, 02 (2008), 109-130.

31. Poincaré, H. *The Foundations of Science*. The Science Press, Lancaster, PA, 1913.

32. Resnik, P. Selection and Information: A Class-Based Approach to Lexical Relationships. 1993, 161.