aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Current Directions in Psychological Science
1–7

DOI: 10.1177/09637214221098054
www.psychologicalscience.org/CDPS
SAGE

From Semantic Vectors to Analogical Mapping

Keith J. Holyoak[1], Nicholas Ichien[1], and Hongjing Lu[1,2]
[1]Department of Psychology and [2]Department of Statistics, University of California, Los Angeles

Abstract
Human reasoning goes beyond knowledge about individual entities, extending to inferences based on *relations* between entities. Here we focus on the use of relations in verbal analogical mapping, sketching a general approach based on assessing similarity between patterns of semantic relations between words. This approach combines research in artificial intelligence with work in psychology and cognitive science, with the aim of minimizing hand coding of text inputs for reasoning tasks. The computational framework takes as inputs vector representations of individual word meanings, coupled with semantic representations of the relations between words, and uses these inputs to form semantic-relation networks for individual analogues. Analogical mapping is operationalized as graph matching under cognitive and computational constraints. The approach highlights the central role of semantics in analogical mapping.

Keywords
relations, analogy, embeddings, machine learning

The early Romantic poet Samuel Taylor Coleridge observed that the creative mind needs to become "accustomed to contemplate not *things* only, but likewise and chiefly the *relations* of things" (as quoted in Holyoak, 2019, p. 35; italics in original). If "things" and "relations" have distinct yet interconnected representations in the human mind, then new combinations can be formed, revealing more abstract similarities. Relational similarity underlies the ability to grasp everyday metaphors, such as "Life is a roller coaster," and scientific theories, such as the famous Rutherford–Bohr analogy explaining the structure of an atom (an unfamiliar *target* analogue) in terms of the structure of the solar system (a more familiar *source* analogue). In complex examples, the similarity involves not just individual relations, but *patterns* of relations (e.g., very crudely, electromagnetism makes an electron revolve around the nucleus of an atom, similarly to the way gravity makes the earth revolve around the sun).

Cognitive scientists aim to explain how people can detect such patterns of relational similarity and use them to make plausible inferences based on analogy. A general approach is to build computational models that can actually accomplish these tasks. A computational model has to operate on some basic inputs—in the case of analogy, representations of "things" and the "relations" between them. Here we sketch one general approach to building a computational model that takes elementary building blocks and puts them together to create knowledge structures—*semantic-relation networks*—that make it possible to find systematic correspondences, or *mappings*. By finding sensible mappings, it is possible to solve complex analogies, such as that between the solar system and the atom. As we show, this approach combines research in artificial intelligence (AI) with research in psychology and cognitive science.

An overarching goal is to model complex reasoning while avoiding hand coding of the inputs (i.e., having the modeler create representations based on the modeler's own beliefs about what knowledge should be included in them). Besides creating a danger that apparent successes of the reasoning model will actually depend on unrealistic inputs, hand coding is prohibitively labor-intensive (in practice, impossible for

Corresponding Author:
Keith J. Holyoak, Department of Psychology, University of California, Los Angeles
Email: holyoak@lifesci.ucla.edu

large databases). Moreover, hand coding may miss opportunities to exploit rich semantic associations that can inform relational reasoning, as a modeler is unlikely to anticipate all the subtle aspects of meaning that may be encoded in the human semantic system. Current machine-learning algorithms can create high-dimensional vectors that implicitly encode detailed aspects of word meanings in ways a modeler could not anticipate.

The general approach we describe in this review is to start from machine-generated representations of the meanings of words (e.g., "read" and "book"), learn representations of the semantic relations between words (e.g., "read" is an action performed on a "book"), systematically combine words and their relations to form semantic-relation networks, and then use these networks for analogical reasoning. Figure 1 gives an overview of the approach, using the Rutherford–Bohr analogy as an example. We describe a specific instantiation of the general computational framework, but any component could potentially be realized in a different way.

Word Meanings as Embeddings

In recent years, advances in machine learning have enabled the creation of a new kind of mathematical "dictionary" in which the meanings of individual words are represented by high-dimensional vectors of continuous-value features, termed *embeddings* (for a general overview, see Günther et al., 2019). Embeddings correspond to activation states in a hidden layer (i.e., a layer that lies between the input and output layers) of a neural network that has been trained to predict patterns of words in sequence as they appear in large text corpora (e.g., billions of articles in Google News). This general approach to word meaning is termed *distributional semantics*: Meanings of individual words are derived from their statistical distribution in texts. Many embedding models have been developed in machine learning, including GloVe (Pennington et al., 2014), BERT (Devlin et al., 2019), and GPT3 (Brown et al., 2020). Here we focus on Word2vec (Mikolov et al., 2013). In Word2vec, every word in English (or any other language) is typically coded by a vector of values on a common set of 300 feature dimensions. Ideally, an embedding captures core semantics of the input so that semantically similar words are placed close together in the embedding space. For example, "cat" and "dog" would be in close proximity (i.e., the meanings are highly similar), whereas "cat" and "microscope" would be far apart (the meanings are dissimilar).

Embedding models have proved broadly applicable to predicting psychological phenomena that depend on sensitivity to similarity of word meanings. Examples include judgments of lexical similarity or association, neural activity triggered by processing of words and relations, and high-level inferences (for a review, see Bhatia & Aka, 2022; for a discussion of and response to critiques of embeddings as psychological models, see Günther et al., 2019). As building blocks for an analogy model, embeddings can be used to code the things (words). Figure 1 illustrates the extraction of major concepts, corresponding to keywords, from texts describing the solar system (source) and the atom (target).

Semantic Relation Vectors

The next step is to code *relations* between concepts. As is the case for individual word meanings, relations can be represented as vectors. Once vector representations of semantic relations have been created, relational similarity can be computed in much the same way as lexical (word-to-word) similarity. But how should relation vectors be defined?

To reduce the need to hand-code relations, analogy modelers have developed methods that exploit large-scale knowledge bases, large text corpora, and methods for learning relations. One general approach is *search based*. For example, the CogSketch system (Forbus et al., 2011) creates representations of hand-drawn sketches using relations obtained by a search of a large prestored knowledge base, OpenCyc, which includes more than 58,000 concepts, 8,000 relations (both physical and conceptual), and 1.3 million facts. CogSketch uses retrieved relations to build propositional representations that support a symbolic method for performing analogical mapping. Latent Relational Analysis (LRA), though very different from CogSketch, also implements a search-based approach (Turney, 2008, 2013). LRA forms vector representations of semantic relations for individual word pairs without relying on predefined relations (Turney, 2008, 2013). In this analogy model, the semantic relation between two words is represented by a vector in which the elements are computed from frequency counts of short relational phrases containing the two words, identified in a search through a large text corpus. Statistical techniques are used to reduce this co-occurrence table (word pairs on rows, phrases on columns) to a relation vector with 300 dimensions. Turney's model achieved human-level accuracy solving standard sets of verbal analogy problems. As an AI system, LRA was not intended as a psychological proposal—humans probably are not capable of searching a large text corpus for each word pair in the process of solving verbal analogies.

A second general approach is *embedding based*. Rather than using search to create relation vectors,

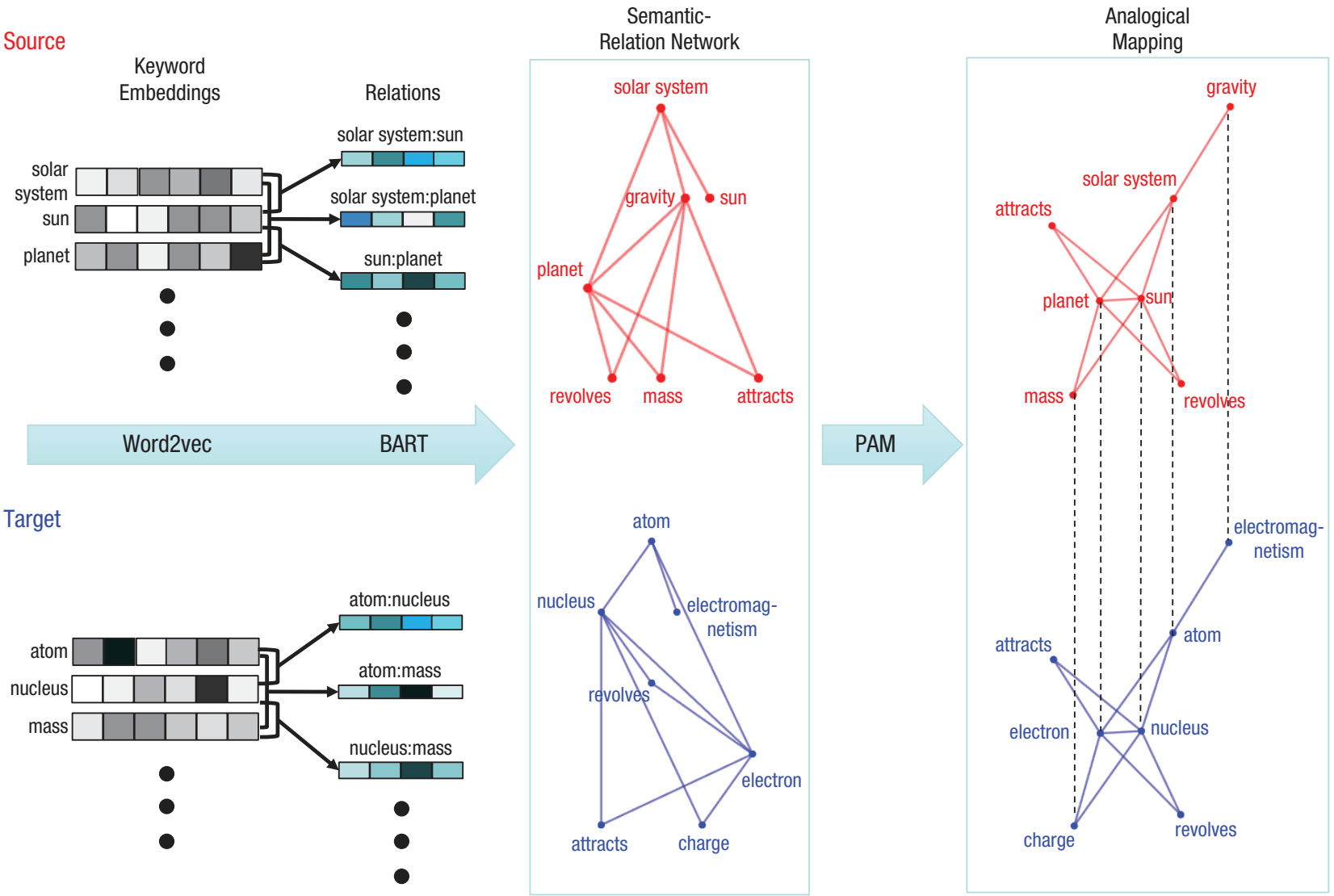

Fig. 1. An illustration of analogical reasoning based on semantic mapping, using the Rutherford–Bohr analogy as an example. Comprehension processes extract main concepts in the source (solar system) and target (atom) analogues, as well as basic syntactic relations. (For illustration purposes, only some of the keywords are shown here.) A word embedding (provided by Word2vec in this example) is obtained for each keyword; relation vectors (from Bayesian Analogy with Relational Transformations, or BART, in this example) are obtained for pairs of keywords. These embeddings and relation vectors are used to generate semantic-relation networks, in which nodes are keywords and edges are semantic relations between keywords. Word embeddings are assigned as node attributes, and relation vectors are assigned as edge attributes. Probabilistic Analogical Mapping (PAM) uses these networks to find optimal analogical mappings between the keywords in the source and target.

models implementing this approach use a generic computation to derive relation vectors directly from the embeddings for individual words. Within these embeddings, some features carry relational information, but in an implicit and "entangled" manner: There is no simple correspondence between individual dimensions of a vector and meaningful attributes, such as shape, basic category, or animacy (although meaningful attributes can be extracted from embeddings by statistical techniques; Hollis & Westbury, 2016). For example, the embeddings for the words "rich" and "poor" may each include features associated (in a probabilistic manner) with such relational concepts as "money," "continuous quantity," and "relative extremity."

It is possible that word embeddings contain enough implicit relational information to solve some analogy problems. A basic intuition is that in a valid verbal analogy in the form A:B::C:D (e.g., "king:queen::man:woman"), the vector representing A:B should be roughly parallel to the C:D vector. Word2vec (or any other embedding model) provides a simple way to create a generic relation vector: Take the difference between the relevant word vectors (e.g., the relation vector for "king:queen" is simply the difference between the embeddings for "king" and "queen").

Using this method, Word2vec achieved some success for analogies based on semantically close concepts, such as this example (Mikolov et al., 2013); however, it fails to reliably solve problems based on more dissimilar concepts (Peterson et al., 2020). A limitation of Word2vec embeddings is that the features—including those that carry relational information—are highly entangled. Feature entanglement is inherited by Word2vec difference vectors, and this obscures information about specific meaningful relations.

A third type of approach is *learning based*. In this approach, learning algorithms are applied to pairs of individual concepts in order to acquire representations of relations between them (Doumas et al., 2008, 2022). As an example, here we focus on a computational model developed by our group, Bayesian Analogy with Relational Transformations (BART; Lu et al., 2012, 2019). BART operates on word embeddings, taking Word2vec embeddings for pairs of individual words as inputs. From these embeddings (which carry relational information within their entangled features), the model learns dimensions of disentangled relation vectors in a transformed space. The dimensions in BART's relation vectors are meaningful semantic relations that have been identified in classic psychometric and psycholinguistic research (Bejar et al., 1991; Chaffin & Hermann, 1988), such as *class inclusion* ("tree:oak") and *part-whole* ("hand:finger"). The model learns individual semantic relations from a set of word pairs (taken from norms collected by Jurgens et al., 2012) consisting of a mix of positive examples of the target relation and negative examples, for which the relation does not hold. BART applies statistical methods to identify a predictive subset of the embedding features, and to estimate associated weights on these features that predict the probability that the relation holds. BART depends on supervised learning, using word pairs that were selected to exemplify each of a set of predetermined relations. But once acquired, its learned weights can be applied to *any* pair of words represented by their embeddings, which enables significant generalization.

After learning a set of specific relations, BART can compute a relation vector for any pair of words by calculating the posterior probability that the pair instantiates each of the learned relations (i.e., the probability derived by combining the prior probability of the relation with the evidence provided by the given word pair). The specific relation between any two words is thus coded as a distributed representation (e.g., the vector for "friend:enemy" might have relatively high values for multiple relation dimensions, perhaps both contrast and similarity). This representation is disentangled, as each element in the relation vector corresponds to the posterior probability that a particular meaningful relation holds between the concepts. BART's distributed representations enable the model to generalize to new word pairs that may be linked by specific relations on which the model was not trained. By comparing the similarity between relation vectors (assessed by cosine distance), BART can solve verbal analogies in the A:B::C:D format, can predict human judgments of the degree to which a word pair is a typical example of a given relation (Lu et al., 2019), and also can predict judgments of the similarity between relations expressed as word pairs (Ichien et al., 2022). In addition, BART has been used to predict patterns of similarity in neural responses to relations during analogical reasoning (Chiang et al., 2021). In the example shown in Figure 1, BART is used to create semantic relation vectors linking pairs of keywords in each analogue (e.g., "solar system" and "sun" in the source).

BART's relation representations are thus compositional, in that the specific relation between a pair of words is approximated by a pattern of values across a set of more elementary relations, as has been suggested in previous theoretical work both in psychology (Chaffin & Hermann, 1988) and in linguistics (Jackendoff, 2007). However, BART is not committed to any particular set of elementary relations; rather, it makes use of whatever relations the model has been trained on. BART's set of learned relations is partially based on a theoretical taxonomy (Bejar et al., 1991), but we do not assume that this set is sufficient to encode the specific aspects of all semantic relations between words. Nonetheless, we have explored the extent to which BART's

approximations to specific relations may prove useful in finding analogical mappings between relatively complex situations.

Semantic-Relation Networks

Armed with vector representations of words and the semantic relations between them, a computational model can solve analogies using more complex knowledge involving several concepts and their interrelationships. As an example, we use an algorithm developed by our group, Probabilistic Analogical Mapping (PAM; Lu et al., 2022). As illustrated in Figure 1, a semantic-relation network has the form of an *attributed graph* (Gold & Rangarajan, 1996) in which nodes and edges (i.e., connections between nodes) are assigned numerical values (*attributes*) that capture the semantic meanings of individual concepts and their pairwise relations. In PAM, the attribute for each node is the Word2vec embedding of a key concept word, and the attribute for each edge is the corresponding relation vector generated by BART. The relation features in the BART vectors used by PAM are augmented by additional features representing the probability that the first word in a pair fills the first role of the relation (e.g., for "finger:hand," the vector includes features coding the probability that "finger" fills the *part* role in a part-whole relation). These role features, which correspond to role-governed categories that human reasoners use to represent relational information (e.g., Goldwater et al., 2011), aid in finding mappings between words that fill similar roles across multiple relations.

Semantic-relation networks can also easily incorporate constraints on relation structures. By default, semantic relations between concepts are treated as bidirectional (e.g., for the words "finger" and "hand," one direction represents the relation that a finger is part of a hand, and the other direction represents the relation that a hand has a finger as a part of it). But to capture basic syntactic structure, such as subject-verb-object, unidirectional connections can be used to form semantic-relation networks. For example, a description of the solar system analogue might include a sentence stating that "the earth revolves around the sun," which of course means something very different from "the sun revolves around the earth." In forming a semantic-relation network as depicted in Figure 1, the intended sentence structure can be coded by unidirectional links between the keywords (i.e., earth → revolves, revolves → sun, earth → sun).

Mapping Based on Semantic-Relation Networks

Using the semantic-relation networks created for the source and target analogues (e.g., solar system and atom), PAM performs analogical mapping using a probabilistic approach (Fig. 1). The basic procedure is to find the mapping between nodes (keywords) that jointly maximizes the similarity of both nodes and their corresponding edges, with the further constraint that one-to-one mappings between nodes are preferred. A parameter in PAM controls the relative importance of lexical similarity (nodes) versus relational similarity (edges) for mapping. A fundamental assumption is that words (nodes) and relations (edges) constitute two separable pools of semantic information (entity based and relation based), which then drive judgments of similarity between analogues. This assumption is consistent with a wide variety of evidence for separable contributions of entity-based and relational similarity in human comparison judgments (e.g., Goldstone et al., 1991).

Our initial work (Lu et al., 2022) has shown that the PAM model is able to solve complex analogical mappings using verbal materials. For example, it finds seven mappings between keywords associated with the Rutherford–Bohr analogy (some of which are shown in Fig. 1). Even without coding the subject-verb-object configuration by unidirectional links, PAM is able to find five of these seven mappings. More generally, across a set of 20 science analogies and everyday metaphors coded by keywords, PAM achieved 85% accuracy in mapping (as compared with 88% accuracy obtained in a human experiment) without representing sentence syntax at all. The model's global relative emphasis on lexical concepts (node similarity) versus relations (edge similarity) can be varied so that it can account for the human developmental shift toward increased sensitivity to relations (Gentner & Rattermann, 1991). This shift likely reflects an asymmetry in processing demands: A relation (edge) can be computed only after the concepts (nodes) that it relates are represented, whereas an individual concept can be represented without necessarily forming a relation representation. In addition, the model provides a measure of global similarity between analogues, which can be used to support the retrieval of plausible source analogues from memory. By building a reasoning model on top of learning mechanisms grounded in distributional semantics, the model has drawn closer to the goal of automating analogical reasoning for natural-language inputs.

Limitations and Future Directions

The project of linking semantic vectors to analogical reasoning is still at an early stage, and many open questions remain. Human assistance is still required to identify key concepts and their relations in texts. Future work should be aimed at more fully automating the generation of semantic-relation networks from text inputs. The current assumption that basic syntactic structure, such as

subject-verb-object, is encoded by unidirectional connections between keywords is simply a convenient heuristic. Adopting natural-language processing techniques that create relatively flat syntactic parses (perhaps with a version of dependency grammar; see Jurafsky & Martin, 2021, Chapter 14) might make it possible to extract syntactic relations automatically and use them to augment semantic-relation networks. Such advances could contribute to efforts to automate the discovery of analogies in online databases. In addition, analogical models based on semantic vectors have yet to be extended to address the later stages of analogical reasoning, in which a mapping is used to develop plausible inferences about the target and to learn a general schema that integrates the source and target. The PAM model so far deals only with verbal analogies. But in principle, semantic-relation networks could be used to perform mapping given any system for assigning vectors as attributes of nodes and edges in a graph. The system could therefore be adapted to solve mappings using perceptual inputs such as pictures, once relevant object features and perceptual relations have been identified (for a preliminary effort to apply PAM to visual analogies, see Fu et al., in press). We hope that future work will foster the evolving synergy between theoretical ideas drawn from AI and from cognitive science, in order both to provide a fuller understanding of human reasoning and to enhance the reasoning capacities of machines.

Recommended Reading

Bhatia, S., & Aka, A. (2022). (See References). An overview of how word embeddings can be used to model human judgments that depend on semantic similarity.

Günther, F., Rinaldi, L., & Marelli, M. (2019). (See References). An introduction to vector-space models based on distributional semantics, with consideration of common criticisms of their use as psychological models.

Lu, H., Ichien, N., & Holyoak, K. J. (2022). (See References). A detailed presentation of Probabilistic Analogical Mapping (PAM), with simulations of representative sets of psychological data involving analogical mapping in adults and children, as well as analogue retrieval.

Lu, H., Wu, Y. N., & Holyoak, K. J. (2019). (See References). A presentation of Bayesian Analogy with Relational Transformations (BART), a model of how semantic relations can be learned from word embeddings, with simulations of human judgments of the degree to which word pairs are typical examples of various relations and of analogical reasoning for simple verbal problems.

Transparency

Action Editor: Robert L. Goldstone
Editor: Robert L. Goldstone
Declaration of Conflicting Interests
The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.
Funding
Preparation of this article was supported by National Science Foundation Grants BCS-1827374 and IIS-1956441.

ORCID iD

Keith J. Holyoak https://orcid.org/0000-0001-8010-6267

References

Bejar, I. I., Chaffin, R., & Embretson, S. (1991). *Cognitive and psychometric analysis of analogical problem solving*. Springer-Verlag.

Bhatia, S., & Aka, A. (2022). Cognitive modeling with representations from large-scale digital data. *Current Directions in Psychological Science*, *31*(3), 207–214. https://doi.org/10.1177/09637214211068113

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., . . . Amodei, D. (2020). *Language models are few-shot learners*. arXiv. https://arxiv.org/abs/2005.14165

Chaffin, R., & Hermann, D. J. (1988). The nature of semantic relations: A comparison of two approaches. In M. Evens (Ed.), *Relational models of the lexicon* (pp. 289–334). Cambridge University Press.

Chiang, J. N., Peng, Y., Lu, H., Holyoak, K. J., & Monti, M. M. (2021). Distributed code for semantic relations predicts neural similarity during analogical reasoning. *Journal of Cognitive Neuroscience*, *33*(3), 377–389. https://doi.org/10.1162/jocn_a_01620

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In J. Bustein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (Vol. 1, pp. 4171–4186). Association for Computational Linguistics. http://doi.org/10.18653/v1/N19-1423

Doumas, L. A. A., Hummel, J. E., & Sandhofer, C. M. (2008). A theory of the discovery and predication of relational concepts. *Psychological Review*, *115*(1), 1–43. https://doi.org/10.1037/0033-295X.115.1.1

Doumas, L. A. A., Puebla, G., Martin, A. E., & Hummel, J. E. (2022). A theory of relation learning and cross-domain generalization. *Psychological Review*. Advance online publication. https://doi.org/10.1037/rev0000346

Forbus, K., Usher, J., Lovett, A., Lockwood, K., & Wetzel, J. (2011). CogSketch: Sketch understanding for cognitive science research and education. *Topics in Cognitive Science*, *3*(4), 648–666. https://doi.org/10.1111/j.1756-8765.2011.01149.x

Fu, S., Holyoak, K. J., & Lu, H. (in press). From vision to reasoning: Probabilistic analogical mapping between 3D objects. In J. Culbertson, A. Perfors, H. Rabagliati, & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Meeting of the Cognitive Science Society*. Cognitive Science Society.

Gentner, D., & Rattermann, M. J. (1991). Language and the career of similarity. In S. A. Gelman & J. P. Byrnes (Eds.),

Perspectives on language and thought (pp. 225–277). Cambridge University Press.

Gold, S., & Rangarajan, A. (1996). A graduated assignment algorithm for graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *18*(4), 377–388. https://doi.org/10.1109/34.491619

Goldstone, R. L., Medin, D. L., & Gentner, D. (1991). Relational similarity and the nonindependence of features in similarity judgments. *Cognitive Psychology*, *23*(2), 222–262. https://doi.org/10.1016/0010-0285(91)90010-L

Goldwater, M. B., Markman, A. B., & Stilwell, C. H. (2011). The empirical case for role-governed categories. *Cognition*, *118*(3), 359–376. https://doi.org/10.1016/j.cognition.2010.10.009

Günther, F., Rinaldi, L., & Marelli, M. (2019). Vector-space models of semantic representation from a cognitive perspective: A discussion of common misconceptions. *Perspectives on Psychological Science*, *14*(6), 1006–1033. https://doi.org/10.1177/1745691619861372

Hollis, G., & Westbury, C. (2016). The principals of meaning: Extracting semantic dimensions from co-occurrence models of semantics. *Psychonomic Bulletin & Review*, *23*(6), 1744–1756. https://doi.org/10.3758/s13423-016-1053-2

Holyoak, K. J. (2019). *The spider's thread: Metaphor in mind, brain, and poetry*. MIT Press.

Ichien, N., Lu, H., & Holyoak, K. J. (2022). Predicting patterns of similarity among abstract semantic relations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(1), 108–121. https://doi.org/10.1037/xlm0001010

Jackendoff, R. (2007). Linguistics in cognitive science: The state of the art. *Linguistic Review*, *24*(4), 347–401. https://doi.org/10.1515/TLR.2007.014

Jurafsky, D., & Martin, J. H. (2021). *Speech and language processing* (2nd ed.). Prentice Hall.

Jurgens, D. A., Mohammad, S. M., Turney, P. D., & Holyoak, K. J. (2012). SemEval-2012 Task 2: Measuring degrees of relational similarity. In E. Agirre, J. Bos, M. Diab, S. Manandhar, Y. Marton, & D. Yuret (Eds.), **SEM 2012: The First Joint Conference on Lexical and Computational Semantics (*SEM)* (pp. 356–364). Association for Computational Linguistics. https://aclanthology.org/S12-1047

Lu, H., Chen, D., & Holyoak, K. J. (2012). Bayesian analogy with relational transformations. *Psychological Review*, *119*(3), 617–648. https://doi.org/10.1037/a0028719

Lu, H., Ichien, N., & Holyoak, K. J. (2022). Probabilistic analogical mapping with semantic relation networks. *Psychological Review*. Advance online publication. https://doi.org/10.1037/rev0000358

Lu, H., Wu, Y. N., & Holyoak, K. J. (2019). Emergence of analogy from relation learning. *Proceedings of the National Academy of Sciences, USA*, *116*(10), 4176–4181. https://doi.org/10.1073/pnas.1814779116

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In C. J. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 26 (NIPS 2013)* (pp. 3136–3144). Curran Associates.

Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1532–1543). Association for Computational Linguistics. https://aclanthology.org/D14-1162/

Peterson, J. C., Chen, D., & Griffiths, T. L. (2020). Parallelograms revisited: Exploring the limitations of vector space models of simple analogies. *Cognition*, *205*, Article 104440. https://doi.org/10.1016/j.cognition.2020.104440

Turney, P. D. (2008). The Latent Relation Mapping Engine: Algorithm and experiments. *Journal of Artificial Intelligence Research*, *33*, 615–655. https://doi.org/10.1613/jair.2693

Turney, P. D. (2013). Distributional semantics beyond words: Supervised learning of analogy and paraphrase. *Transactions of the Association for Computational Linguistics*, *1*, 353–366. https://doi.org/10.1162/tacl_a_00233