Beyond candidate inferences: People treat analogies as probabilistic truths

Brad Rogers (bwrogers@indiana.edu)
David Landy (dlandy@indiana.edu)
Department of Psychological and Brain Sciences, 1101 East 10th Street
Bloomington, IN 47405 USA

Abstract

People use analogies for many cognitive purposes such as building mental models, making inspired guesses, and extracting relational structure. Here we examine whether and how analogies may have more direct influence on knowledge: Do people treat analogies as probabilistically true explanations for uncertain propositions?

We report an experiment that explores how a suggested analogy can influence people's confidence in inferences. Participants made predictions while simultaneously evaluating a suggested analogy and observed evidence. In two conditions, the evidence is either consistent with or in conflict with propositions based on the suggested analogy. We analyze the responses statistically and in a psychologically plausible Bayesian network model. We find that analogies are used for more than just generating candidate inferences. They act as probabilistic truths that affect the integration of evidence and confidence in both the target and source domains. People readily treat analogies not as a one-way projection from source to target, but as a mutually informative connection.

Keywords: **A**nalogy, Bayesian Network, Computation, Confidence, Explanation, Inference, Reasoning

Introduction

A teacher proposes to her class that atmospheric carbon concentration is like the water level in a bathtub (Kunzig, 2009). This science classroom analogy suggests many possible inferences about the atmosphere (and maybe bathtubs) that students can test by collecting evidence. Perhaps the atmospheric carbon level rises or falls based on the difference between carbon "faucet" and "drain" rates. Or maybe once carbon levels hit the upper limit, carbon dioxide will spill over into outer space. These new inferences might be true, or not. But what about the analogy itself? Is it the sort of thing that can be true or false? Does it depend on the inference? If it can be true, what kind of evidence would support it?

Analogy is used in a wide range of uncertain contexts such as contentious negotiation (Loewenstein et al., 1999), ambiguous accounting determinations (Magro & Nutter, 2012), scientific discovery (Gentner, 2002; Hesse, 1966), thinking about astronomical distances (Resnick et al., 2012), and war declaration decisions (Khong, 1992). We use analogies when knowledge is scarce. But does analogy act like other explanations? Can we combine analogy with observed evidence? Like explanations, do we believe in them more when they successfully predict or explain our observations? Currently there is no account for how we integrate analogy and observed evidence when grappling to understand uncertain situations. Even more, there is no psychological account that explicitly affords epistemic value to analogy.

In this paper we examine and affirm the hypothesis that people treat analogies as probabilistic truths. Analogies can be treated as true or false, and people integrate analogies with evidence much like they do for causal explanations.

Candidate Inferences Hypothesis

Analogy is often called "the weakest form of evidence." Indeed, one account is that an analogy does not provide any evidence at all to favor its suggested *candidate inferences* since the act of constructing an analogy does not involve collecting new observations. Proponents of this account suggest that an analogy might render propositions more plausible, but not more probable (Bartha, 2010). Popular theories of analogical inference (Doumas et al., 2008; Gentner, 1989; Hummel & Holyoak, 1997; Lu et al., 2012) largely invoke this candidate inferences account:[1]

> Analogical reasoning in uncertain contexts begins with a well-described source domain and a target domain that requires an inference. A speculative analogy is made from the source to the target, which establishes a structural map between the two situations. With some luck, the source domain might contain useful correspondences to unknown elements of the target, producing candidate inferences that can only be validated by observed evidence in the target domain.

Some computational models treat analogy as a weighted score (e.g., ACME, SME), but this score is typically taken to reflect coherence (Holyoak & Thagard, 1989; Thagard, 1989) or structural consistency (Gentner, 1983) and has not been extended to estimate the truth or rationality of the analogy. The correspondence identified by the analogy is not something that could be true or false. Rather, it is considered an artifact of our thinking about possible target inferences that should only guide our pursuit of evidence.

[1] Our description generalizes across typical candidate inferences approaches (Falkenhainer, 1990; Gentner & Colhoun, 2010; Gentner & Markman, 1997) and other related approaches such as copy with substitution & generation (Holyoak & Hummel, 2000; Lee & Holyoak, 2008). For our purposes, the distinctions matter less than the commonalities.

Analogy as Evidence Hypothesis

In this paper, we explore a stronger account of analogy with an expanded epistemic role. On this account, analogies not only introduce plausible inferences; they create a probabilistic connection between source and target that establishes and conveys inferential confidence.

The basic intuition is suggested by Peirce's notion of abduction (1935) and Harman’s notion of inference to the best explanation (1965). Inference to the best suggests that people have confidence in the explanations that make their observations the least surprising. When an analogy suggests an inference in the target domain, observed evidence for that inference should increase the likelihood of the analogy itself when the analogy is taken as an explanation for the target inference. Conversely, if the target inference turns out to be false, the analogy becomes suspect. Our account builds on this insight to propose that people treat analogies and source knowledge as raising the conditional probability of target inferences.

Some readers may find it easy to consider that analogies act as a kind of theory whose truth can be supported or refuted by evidence. Indeed, some philosophical investigations have proposed statistical bases for analogical rationality (Harrod, 1956; Mill, 1882), and a recent study has found that people are sensitive to these rational statistics (Rogers & Landy, 2016). But this epistemological view of analogy has not been dominant in the literature. Still, we are only interested here in the psychological question of whether people treat analogies as a probabilistic truth, rather than the normative question of whether they ought to.

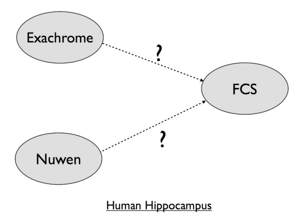
Approach

We conducted an experiment that asked participants to rate their confidence in competing explanations in two domains that may or may not be related. We provided observed evidence in one domain whose coherence with competing explanations was manipulated across subjects. A *statistical analysis* estimates primary effects to determine whether observed evidence influences reported confidence in the analogy. A *Bayesian network model* was used to compare responses with a psychologically plausible instantiation of analogy that integrates with observed evidence.

Experiment

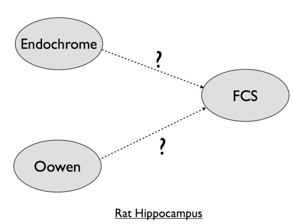
Participants were presented a fictional narrative situation describing two novel scientific phenomena, including simple visual representations.[2] Mutually exclusive explanations are provided for the phenomena. A suggested correspondence between the phenomena is described as leading scientists to develop an explanation and experiment. After receiving the stimulus and a test condition statement, participants rate their confidence in each of the explanations and the analogy.

For 20 years, biological scientists have fought over the relation between FCS, exachrome, and nuwen in the human hippocampus. Some scientists believe that exachrome is produced in cell nuclei, and that exachrome drives up production of FCS. They think that nuwen doesn't matter for FCS production. The more exachrome, the more FCS. Other scientists argue that it is nuwen produced in the cell nuclei that drives production of FCS, and that exachrome is an irrelevant byproduct. Both of these production pathways (the exachrome pathway and the nuwen pathway) are chemically plausible; which is correct is a matter of current debate. It is quite unlikely that both are correct. The following image summarizes the debate:



Rat hippocampi are much less well understood than human hippocampi. Rat hippocampi do exhibit FCS, but they do not contain exachrome or nuwen. Recently, one scientist (who was not attached to either of the other groups mentioned before) has suggested that FCS might be produced in rats in a way similar to that of humans. She identified two chemicals--called endochrome and oowen--that are similar to exachrome and nuwen, and that are produced in the rat hippocampus.

In other non-biological contexts, nuwen is sometimes used to predict properties of oowen, and exachrome is sometimes used to predict properties of endochrome. The following image summarizes the possibilities suggested by this scientist:



Recently, on the basis of the suggested links between exachrome and FCS in humans, and between human and rat FCS production, the scientist and her colleagues tested a novel hypothesis using rat hippocampi. The scientists injected the rats with a hormone that stimulates the production of endochrome. Several days later, they examined the level of FCS in the rat brain, predicting that it would show an increase.

Figure 1. Stimulus narrative presented to all participants

[2] The current stimulus was designed with a near analogy rather than a distant analogy. We expected the homological nature of mammalian brains to make the analogy prima facie plausible. The rat brain to human brain analogy is often used in experimental study, although here we reverse source and target. Other stimuli (discussed later) have produced consistent, but less pronounced effects for analogies across more distant domains.

Participants

We recruited N=300 adults living in the US from Amazon's Mechanical Turk where participants can volunteer to complete short studies and other tasks in return for 35 cents.

Design

Each participant was presented the same narrative (Fig. 1). Additionally, they were presented a single statement regarding the outcome of the scientific experiment implemented on the rat brain. This statement varied between subjects for three balanced conditions:

- As it turned out, increasing endochrome led to a large increase in the level of FCS in the rat hippocampus. (Confirm)
- The experiment results haven't been released yet, so we don't know how it worked out. (Neutral)
- As it turned out, increasing endochrome did not increase the level of FCS in the rat hippocampus at all. (Disconfirm)

Procedure

Following presentation of the narrative and conditional statement, participants were asked to indicate their confidence for each explanation using a 7-point Likert scale:

- Exachrome causes the production of FCS in human hippocampus.
- Nuwen causes the production of FCS in human hippocampus.
- Endochrome causes the production of FCS in rat hippocampus.
- Oowen causes the production of FCS in rat hippocampus.

They were also asked to indicate their confidence that the situations are analogous:

- The production of FCS in rat hippocampus works similarly to that of human hippocampus.

Two balanced question orders were used. No response differences were observed on the basis of question order, so the factor was removed from subsequent analysis. All conditions contained a simple attention check. About ¼ of participants failed the attention check and were removed from the analysis.

Statistical Analysis & Results

We analyzed the participant responses by regressing each response item against the between-subject condition statements with each condition coded as a dummy variable. Since the assumptions violated linearity, we used resampling with 10,000 replications to evaluate statistical significance. For comparison, we also calculated Cohen's d to corroborate the significance of the observed effect sizes.

As expected, participant confidence in this explanation increased for the confirmation condition and decreased for the disconfirmation condition ($p<0.0001$, $d=2.5$). Since the procedure asserted that the two explanations about the rat hippocampus were unlikely to be simultaneously true, we predicted that the competing explanation would follow the opposite pattern. Indeed, when Endochrome ⇒ FCS was supported, confidence ratings for the Oowen ⇒ FCS explanation decreased ($p<0.0001$, $d=2.5$).

Participants confidence ratings in the source domain explanations were also influenced by the observed evidence in the target domain. Confidence in the corresponding source explanations about the human hippocampus changed in the direction consistent with the correspondence structure of the analogy. For the Endochrome ⇒ FCS explanation confidence increased with positive evidence and decreased with negative predictions ($p<0.0001$, $d=0.75$). Confidence in the competing source explanation Nuwen ⇒ FCS was inversely affected ($p<0.0001$, $d=-0.80$). Finally, successful predictions made participants more confident in the idea that the two domains were analogous ($p<0.0001$, $d=0.50$).

Participant responses strongly supported our hypothesis people treat the analogy as evidence for the inferences is suggests. New successful predictions made on the basis of a mapping from the source to the target increased confidence in the commonality of the domains, as well as in the untested scientific explanation that generated them.

Bayesian Network Model & Results

We further analyzed the data using a Bayesian network model (Pearl, 2009) to estimate the influence of the suggested analogy on the response item confidence statements in a way constrained by a plausible causal structure. In the model, each causal explanation is represented as a single node and assigned a prior baseline probability. Since it was stated in the stimulus that the two explanations within a domain were unlikely to be simultaneously true, the model places a negative correlation between the explanations. Without an analogy, the source and target domains (i.e., human and rat hippocampus, respectively) have no causal linkage. On the other hand, if there is a known analogy that is taken as certain, strong causal linkages are present from the source to the target domain.

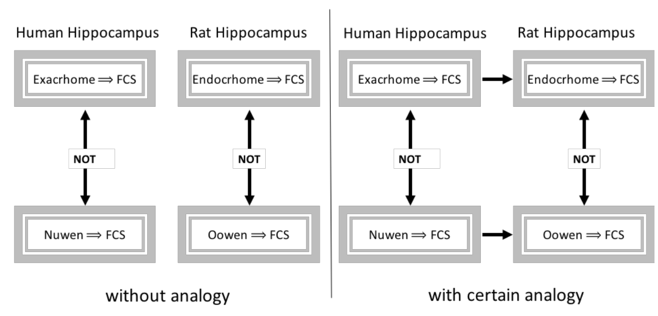


Figure 2. Bayesian network structure without analogy and with certain analogy

With an uncertain analogy, though, the structure itself becomes probabilistic. To capture this, we take the model a step further by representing the analogy itself as a single node. In this way, we can gauge the evidentiary influence of the analogy and participants' confidence in it using their confidence ratings. If the domains were sufficiently complex that multiple mappings were possible, it might be necessary to include structural evaluations in the model such as rankings from a model of structural correspondence (Landy & Hummel, 2010). But in this case, the mapping from source situation to target situation is plainly obvious and can be treated as a single node.

Although the distinction is often drawn between superficial and deep analogies, how people consider the truth of an analogy has not been investigated to the best of our knowledge. As a starting place, the analogy was modeled as a Boolean variable—true or false. Participant confidence in the analogy was estimated by a Beta-distribution..

Now the probability of a target domain explanation prior to observing the experimental results depends on both the probability of the truth of the source domain explanation and the probability of the truth of the analogy. If the analogy is true, then what is true or false in the source domain is also true or false in the target domain. However, if the analogy is false, the truth of the target explanation is independent of the source domain knowledge. In other words, an analogy guarantees correspondence, but a failed analogy does not guarantee non-correspondence. This approach effectively introduces a probabilistic switch between the no analogy and certain analogy network structures.

The prior probabilities of the Bayesian network were fit without including the evidence obtained by the experimental results (i.e., the test condition statement). So each individual is taken to have an estimate of the prior probability of each of the source explanations, the analogy, and the target explanations. The prior probabilities provide an associated estimate of participant confidence that the experimental results will be confirmed or disconfirmed.

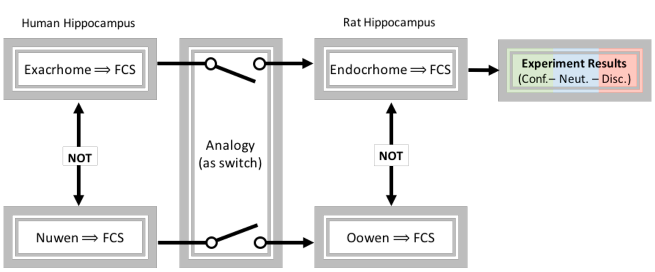


Figure 3. Bayesian network structure with uncertain analogy and evidence from experimental results (i.e., test condition)

The participant data was fit using a hierarchical model. Participants were assumed to have been randomly selected from a population having a single distribution of subjective priors for each node. The priors were estimated using Dirichlet distributions for the domain explanation probabilities[3] and Beta distributions for the analogy and the evidence probabilities. The model had 14 population-level free parameters, fit to 1500 participant responses. Participants from the neutral condition were assumed to respond based on these prior parameter estimates without any additional evidence. Participants from the evidence condition were modeled by updating the Bayes net given the appropriate evidentiary outcome, and these posterior estimates were fit to the responses.

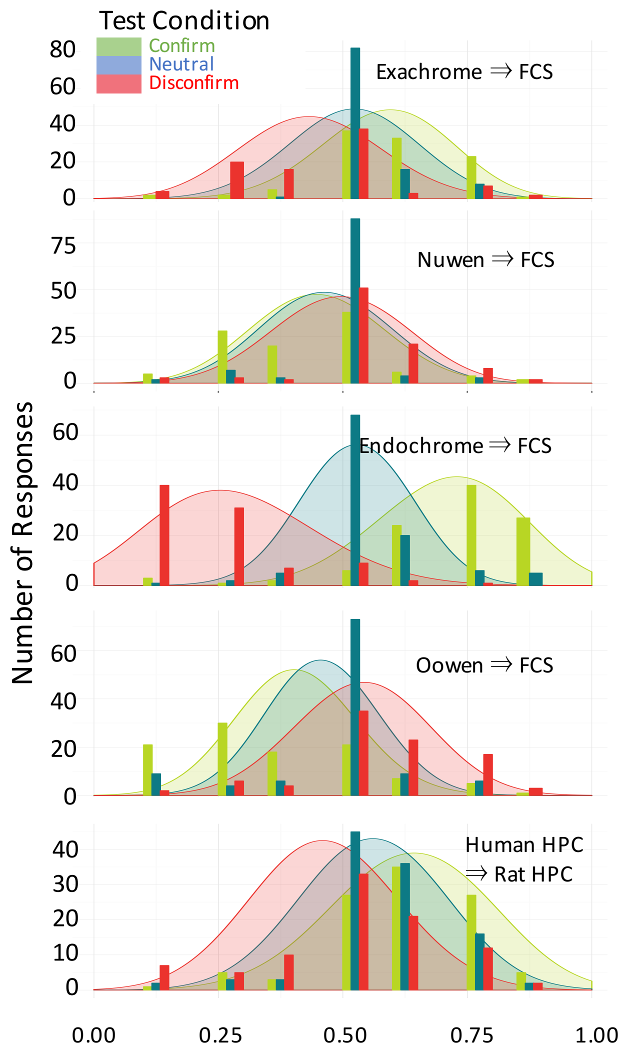


Figure 4. Posterior predictive distributions by condition for each explanation compared with response distributions

We solicited confidence ratings using a Likert scale rather than explicit probability estimates. So a final step in the model was to translate posterior probabilities from the

[3] We intended the explanations in each domain to be interpreted by participants as mutually exclusive and exhaustive, but we did not assume this in their response structure. We allowed the model to account for the possibility that both explanations within a domain are correct or that both are incorrect. This compelled the use of the multivariate Dirichlet distribution rather than the Beta distribution.

Bayesian network into Likert response values. We treated Likert values as ordered and evenly distributed from 0 to 1. Responses were then treated as beta-distributed among these values, with mean at the subjective probability. This allowed variance from the specific posterior subjective probabilities, minimized degrees of freedom in the model, and afforded a limited flexibility in translating posterior probabilities into Likert scale responses. The model was fit in Stan via R: 1,000 posterior samples proved sufficient for model convergence with population-level $\widehat{R}$ values all less than 1.1 (Bates et al., 2015).

Figure 4 indicates posterior predictions of each Bayesian network node overlay with the fit participant Likert responses for each condition. The major patterns in the data were generally well-captured by the model, suggesting that people were integrating evidence from the target prediction success into their confidence in the analogy, and were doing so in a manner that approaches rational behavior. Predictions matched the direction of the observed effects for all five model nodes. If the analogy were rejected by participants, we would expect no differences between conditions in the responses about the analogy and about the source explanations.

	Participant	Model Predicted 95% HPI	
	(Likert fit)	*Lower bound*	*Upper bound*
Exachrome ⟹ FCS	0.14	0.13	0.18
Nuwen ⟹ FCS	-0.12	-0.07	-0.02
Endochrome ⟹ FCS	0.47	0.40	0.46
Oowen ⟹ FCS	-0.21	-0.15	-0.10
Human HPC ⟹ Rat HPC	0.09	0.11	0.20

Figure 5. Difference between confirm and disconfirm participant confidence ratings versus model predictions

The model fit can be evaluated by comparing differences between the distribution of participant responses and the simulated posterior predictions of the population (i.e., 1,000 samples of the posterior for each of 300 participants). Although the model matched the direction of the empirical results in every case, the outcome of this analysis revealed a systematic bias (discussed later) that could not be accounted for by this computational approach.

Discussion

What does it mean to be confident in an analogy? What does it mean for an analogy to be assigned a probability value at all? This is an important open question. Analogies are rarely exact correspondences. Useful analogies are sometimes even known from the outset to be poor, such as "atoms are like solar systems." Alternatively, models and simulations in the social sciences are often presented as valid simplifications of complex phenomena. It seems, then, that we can be confident in an analogy's validity even when we do not believe the correspondence to be exact. This paper takes a first step toward answering these open questions by establishing a basic fact: people *do* treat analogies as probabilistic truths and integrate them with evidence.

Implications for analogical inference

If analogies just generate candidate inferences, then people's confidence in explanations in one domain would be unaffected by observations in another. In contrast, we found that analogical mappings do raise posterior estimates of the likelihood of candidate inferences. Moreover, when uncertainty exists in the source domain knowledge, confirmed analogical inference in the target domain raises confidence in the corresponding source knowledge. This effect suggests that people treat analogies not as a one-way projection from source to target, but mutually informative.

Results show that the effective confidence of the analogy itself is influenced by the success of its inferences suggesting that people evaluate the analogy on more than its degree of structural correspondence. The analogy seems to have a causal property that can be integrated with and influenced by observed evidence. To that point, no evidence was ever presented in the source domain that could arbitrate between the proposed explanations, so evidence confirming a target domain inference could not possibly strengthen the structural correspondence between the domains. And yet, if the new information confirmed inferences made by the analogy, differences by condition in participant confidence ratings suggest they credited the analogy for the success.

It is worth noting that while the candidate inferences account is implied in many extant studies of analogy, the authors of those studies may not wish to explicitly commit to it. For the most part, we believe that the role of evidence in influencing the value of the analogy has been deferred rather than denied. We see these results as extending rather than negating extant approaches.

Deviations from rationality

Although the observed confidence differences are quantitatively close to the model predictions, the observed differences are not completely compatible with rational allocation of probabilities under the assumption that analogies act as evidence for their inferences. Participants attributed success or failure of the analogy more to the veracity of the source explanations and less to the analogy than would be expected by the model structure. In other words, we expected confidence in the analogy to justify shifts in confidence in the source domain explanations. But the observed shifts in the source domain outpaced participant reported confidence in the analogy. One possible explanation is that participants may have interpreted the analogical statement more broadly than intended, so that the possibility of any related dissimilarity would reduce their confidence in the analogical statement. Another possibility is that people use different cognitive processes to rate confidence in analogical statements than they use to rate domain-specific statements. If true, then it may be necessary to apply a simple transformation to reported analogical confidence when modeling analogy in a Bayesian network.

Limitations of the present study

One limitation of the experiment is that the relationship between mammal brains is not only a near analogy, it is also a biological homology. Rats and humans evolved from a common ancestor, so similarities between them may reflect properties of their ancestor rather than attribution of evidence to the analogy per se. Indeed, scientists regularly use animal models to predict properties of human beings on this basis. Because the inference of the experiment may have a biological explanation, shifts in confidence may reflect an alternate process of inference about the cause rather than about the analogy. In subsequent experiments using more distant domains—such as suggesting a link between ion behavior in "super-cooled glass" and macro-economic decisions by nations—we find consistent, but less pronounced effects to those presented here.

Future study

Even though we can conclude that people are willing to treat an uncertain analogy effectively as a probabilistic truth, it is not clear what cognitive processes underlie this effect. Two alternate hypotheses are:

1. People may treat the analogy as a kind of theory whose truth can only be supported by evidence in the source and target domains. This is the most straightforward interpretation of the experiment and the approach taken by the ERIC model of explanatory reasoning under uncertainty (Landy & Hummel, 2010).
2. Success of an inference may imply a stronger structural correspondence than is actually observed. Confirming evidence for an inference in one domain may improve an implied estimate of unobserved, but still predictive, structural correspondence (Rogers & Landy, 2016).

More investigation is needed to distinguish between these possibilities. We still await a fully integrated account of reasoning across correspondences among structures about which people have probabilistic beliefs.

Acknowledgment - *This material is based upon support of BR by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1342962.*

References

Bartha, P. (2010). *By Parallel Reasoning*. Oxford University Press.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistics Software*, *67*(1).

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).

Falkenhainer, B. (1990). A unified approach to explanation and theory formation. In J. Shrager & P. Langley (Eds.), *Computational Models of Scientific Discovery and Theory Formation* (pp. 157–196). Morgan Kaufmann Publishers, Inc.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2013). *Bayesian data analysis*. CRC Press.

Gentner, D. (1983). Structure-Mapping: A Theoretical Framework for Analogy. *Cognitive Science*, *7*(2).

Gentner, D. (1989). The mechanisms of analogical learning. In *Similarity and Analogical Reasoning*. Cambridge University Press.

Gentner, D. (2002). Analogy in Scientific Discovery: The Case of Johannes Kepler. In *Model-Based Reasoning* (pp. 21–39). Springer US.

Gentner, D., & Colhoun, J. (2010). Analogical Processes in Human Thinking and Learning. In *Towards a Theory of Thinking* (pp. 35–48). Springer Berlin Heidelberg.

Gentner, D., & Markman, A. (1997). Structure Mapping in Analogy and Similarity. *American Psychologist*, *52*(1).

Harman, G. H. (1965). The Inference to the Best Explanation. *The Philosophical Review*, *74*(1).

Harrod, R. (1956). *Foundations of Inductive Logic*. Harcourt, Brace & Co.

Hesse, M. (1966). *Models and Analogies in Science*. Notre Dame Press.

Holyoak, K. J., & Hummel, J. E. (2000). The Proper Treatment of Symbols in a Connectionist Architecture. In *Cognitive dynamics: Conceptual change in humans and machines*. MIT Press.

Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science: A Multidisciplinary Journal*, *13*.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, *104*(3).

Khong, Y. F. (1992). *Analogies at War*. Princeton University Press.

Kunzig, R. (2009). The Carbon Bathtub. *National Geographic*.

Landy, D., & Hummel, J. E. (2010). Explanatory Reasoning for Inductive Confidence. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*.

Lee, H. S., & Holyoak, K. J. (2008). The role of causal models in analogical inference. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *34*(5).

Loewenstein, J., Thompson, L., & Gentner, D. (1999). Analogical encoding facilitates knowledge transfer in negotiation. *Psychonomic Bulletin & Review*, *6*(4).

Lu, H., Chen, D., & Holyoak, K. J. (2012). Bayesian analogy with relational transformations. *Psychological Review*, *119*.

Magro, A. M., & Nutter, S. E. (2012). Evaluating the Strength of Evidence: How Experience Affects the Use of Analogical Reasoning and Configural Information Processing in Tax. *The Accounting Review*, *87*(1).

Mill, J. S. (1882). *A System of Logic: Ratiocinative and Inductive*.

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference* (2nd ed.). Cambridge University Press.

Peirce, C. S. (1935). *Collected Papers of Charles Sanders Peirce, vols 1-6*. (C. Hartshorne & P. Weiss, Eds.). Harvard University Press.

Resnick, I., Shipley, T. F., Newcombe, N., Massey, C., & Wills, T. (2012). Examining the Representation and Understanding of Large Magnitudes Using the Hierarchical Alignment model of Analogical Reasoning. In *Proceedings of the 34th Annual Conference of the Cognitive Science Society*.

Rogers, B., & Landy, D. (2016). Investigating Rational Analogy in the Spirit of John Stuart Mill: Bayesian Analysis of Confidence about Inferences across Aligned Simple Systems. In *Proceedings of the 38th Annual Conference of the Cognitive Science Society*.

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences*, *12*.