Title
Investigating Rational Analogy in the Spirit of John Stuart Mill:Bayesian Analysis of Confidence about Inferences across Aligned Simple Systems

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 38(0)

Authors
Rogers, Brad
Landy, David

Publication Date
2016

Peer reviewed

Investigating Rational Analogy in the Spirit of John Stuart Mill: Bayesian Analysis of Confidence about Inferences across Aligned Simple Systems

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

What does it mean for analogy to be rational? John Stuart Mill described a probabilistic underpinning for analogical inference based on the the odds of observing systemic pairwise correspondence across otherwise independent systems by mere chance. Although proponents and critics have debated its validity, Mill's approach has yet to be implemented computationally or studied psychologically. In this paper we examine Mill's approach and show how it can be instantiated using Bayes theorem. Then we describe two experiments that present subjects with partially-revealed, aligned binary strings with varying degrees of intra- and inter-string regularity. Experimental results are compared to a formal rational analysis of the stimuli revealing conditions whereby participants exhibit confidence patterns consistent and inconsistent with Mill's rational basis of analogy.

Keywords: Analogy, Bayes, Confidence, J.S. Mill, Rationality

Introduction

Analogy has been suggested to play central roles in human reasoning activities such as learning facts and concepts (Rumelhart & Norman, 1981), discovering scientific principles (Hesse, 1966), and perceptually interpreting our daily situations (Chalmers, French, & Hofstadter, 1991; Goldstone, Landy, & Brunel, 2011). Previous research has focused on identifying mechanisms of analogy making by comparing people's analogical preferences to predictions made by computational models (Falkenhainer, Forbus, & Gentner, 1989; Goldstone, 1994; Hofstadter, 1995; Holyoak & Thagard, 1989; Hummel & Holyoak, 1997; Lu, Chen, & Holyoak, 2012). These models propose solutions to the problem of finding the best mapping between well described systems based on respective theories of analogy. Inferences have been studied in this paradigm as extensions of possible mappings, but the rationality of such extensions has not been investigated. Although these computational achievements have advanced our understanding of analogy construction, treating analogies as operations over well-defined situations has left an important context under-explored—making predictions and choices in novel and uncertain circumstances where the relevant facts may not be known in advance (Falkenhainer, 1990).

Mill's Rational Basis of Analogy

The idea that analogy has a rational underpinning is not a recent development. In *A System of Logic, Ratiocinative and Inductive*, Mill argued that the probabilistic basis of rational induction does not distinguish relational similarity from property similarity (Mill, 1882). Still he does discuss the nature of their difference:

> *"In the strictest induction, equally with the faintest analogy, we conclude because A resembles B in one or more properties, that it does so in a certain other property. The difference is, that in the case of a complete induction it has been previously shown, by due comparison of instances, that there is an invariable conjunction between the former property or properties and the latter property; but in what is called analogical reasoning, no such conjunction has been made out. There have been no opportunities of putting in practice the Method of Difference, or even the Method of Agreement; but we conclude (and that is all which the argument of analogy amounts to) that a fact m, known to be true of A, is more likely to be true of B if B agrees with A in some of its properties (even though no connection is known to exist between m and those properties), than if no resemblance at all could be traced between B and any other thing known to possess the attribute m."*

Mill describes a key distinction between "complete" property induction and analogical property induction. In the former, a statistic is developed for that specific property by observing it in a sample of many systems thereby establishing a probabilistic connection between instances of that property across those systems generally. In the latter, a statistic is developed according to the pairwise correspondence of a sample of unique properties between two systems thereby establishing a general probabilistic connection between their properties within the two systems or kinds of systems. Mill describes the pairwise correspondence as a proportion:

> *"where the resemblance is very great, the ascertained difference very small, and our knowledge of the subject-matter tolerably extensive, the argument from analogy may approach in strength very near to a valid induction. If, after much observation of B, we find that it agrees with A in nine out of ten of its known properties, we may conclude with a probability of nine to one, that it will possess any given derivative property of A."*

Although Mill's extrapolation of the proportion statistic may be mathematically simplistic, his case for a sampling postulate is clear. It would be terribly unlucky to observe a systemic correspondence in a large sample of pairwise properties and relations across two systems if in fact the true correspondences occur only at a rate of perfect chance. To the extent that we have observed greater correspondence, we would be more surprised to discover that an inferred alignment is untrue.

Mill's approach has drawn some criticism. For example, missing in his argument is the distinction between property and relation in as much as they influence the *independence* of the sample of pairwise correspondences. Along these lines, it seems likely that analogies that occur to us may follow paths toward increasingly higher match rates, a decidedly biased sampling scheme that could lead to systematic over-confidence in analogical inferences (Bartha, 2010). But an argument about biased sampling does not preclude the possibility of an unbiased rational sample, rather it necessitates the existence of such a normative ideal. Another criticism by Bartha of the sampling postulate of analogy is directed not toward its rational basis, but rather the difficulty and bias in representing real systems such that the sampling postulate might be applied. Although analogy in situ undoubtedly relies upon such representations, proponents of this criticism should then be satisfied that a rational solution might exist for any given representation and that this criticism is not about the fundamental analogical form.

Minimal Probabilistic Model of Systemic Pairwise Correspondence

The essence of Mill's proposal is that a proportion can be developed that describes the rate of pairwise matches between two systems, and that this proportion conveys predictive value to inferences made in virtue of that sampled correspondence. Harrod advanced Mill's simplistic calculation by applying the distributional property of systems generated randomly according to a known proportion—if p is the proportion of observable properties of S that are actually shared with T, and n of these properties are observed, then the probability that exactly k of them will be shared with T is given by the binomial distribution (Harrod, 1956):

$$Pr(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Harrod's formulation is a step forward in applying a probabilistic range of a known proportion to an unobserved sample in a system. But it does not fully capture the description of the analogy problem, that of estimating an unknown proportion from an observed sample. This is a problem of inverse probability whereby we seek to understand the probability of different possibly true pairwise match proportions given observation. Since we know the distribution of samples from a hypothesized model, we can apply Bayes theorem to calculate the distribution of hypothetical models (Bayes & Price, 1763; Laplace, 1825):

$$P(H|E) = P(E|H)P(H)/P(E)$$

Applying the binomial distribution for the probability of the sample given the proportion, it has been shown that the solution is beta distributed:

$$p(k) \sim \frac{x^k (1-x)^{n-k}}{B(k+1, n-k+1)}$$

The application of Bayes has brought us to an acceptable model of pairwise correspondence. However, actual reasoning about uncertain situations also involves reasoning about the resulting system. Mill has proposed a basis for reasoning across systems, but any viable approach must also account for this intra-system reasoning. In the analogy literature a classic example of this intra-system reasoning is that an intelligent person who is swayed by the analogy between the solar system and the atom does not then suspect that the nucleus is yellow, large, or causes plants to grow on Earth. Instead, the analogy is integrated with what one knows about each system on its own. Approaches that deal with these concerns have typically separated these two forms of reasoning (Falkenhainer, 1990; Lee & Holyoak, 2008). For the purpose of this initial investigation, it is less important which kinds of relational systems are explored and more important that the systems have structure in some way. In this study, we will analyze systems whose elements also exhibit a proportion. This is admittedly a minimally structured system, but it is adequate to demonstrate Mill's approach.

We apply a Bayesian mixture model that simultaneously evaluates the inferential power of the inter-system relations and the intra-system pairwise proportion (Wasserman, 2000). For computational efficiency, we can implement this using a Dirichlet distribution (or an equivalent hierarchical model) where each category comprises a unique possible combination of paired system properties (i.e., 00,01,10,11):

$$p(k_{i=1}, ..., k_{i=\kappa}) \sim \prod_{i=1}^{\kappa} p_i^{k_i} \frac{1}{B(1+k_{i=1}, ..., 1+k_{i=\kappa})}$$

**Uniform priors were used for purposes of simplicity*

Approach

We investigate Mill's proposed basis of rational analogy in the context of the simplest relational systems and their pairwise correspondences using randomly generated binary strings. Although such systems lack the context and richness often found in human analogy making, they also minimize difficulties and potential sources of biases. This petri dish approach to analogy is a starting place toward evaluating probabilistic aspects of analogical reasoning in more complex situations.

In this study we present two experiments that employ different designs but using stimuli of a common format.

Subjects are presented pairs of randomly generated strings that are only partially revealed and that possess varying degrees of intra- and inter-string regularities. Subjects are given a narrative about the general meaning of the strings, but not the meaning of individual characters. They are asked to evaluate the string pairs, to make a prediction about a target missing character, and to estimate the probability that their prediction would turn out to be true.

Stimuli

Experiments 1 and 2 presented stimuli of a consistent format—pairs of partially revealed binary strings—but employed different designs regarding the generative procedure of the strings. Participants were presented a narrative that we called "Digital Matchmaker" that cast them in the role of a data analyst at a leading online matchmaking website. Their task is to evaluate pairs of incomplete binary strings representing the responses to a questionnaire of two single people who have been matched together for a date.

Single Person 1	0	0	0	1	0	0		0		0	0	0		1	0
Single Person 2	?					1		1	0	1		1	0	1	1

Figure 1: Example "Digital Matchmaker" Stimulus

They are told that matched couples have been made by an algorithm that can look for strong same alignment ("birds of a feather") and for strong opposite alignment ("opposites attract") to responses in the question. But matched couples may also have been made by a professional matchmaker who uses intuitions about photographs rather than survey responses to propose matches. Participants were also told about a study that asked people to fully complete their questionnaires. The study claimed to find that that regularities initially present within a single's incomplete responses and across matched singles' incomplete responses tend to mostly continue in about the same way when fully completed.

Each stimulus contains a target missing survey response that is deemed "crucial to match success." Participants must: 1) select a radio button indicating a prediction of a 0 or a 1 for the target, and 2) use a slider with numeric feedback estimating the probability that their prediction is correct.

Experiment 1

The purpose of Experiment 1 is to evaluate the relationship between people's guesses (and self-reported confidence) and the predictions of the rational analysis. The strings were generated to roughly evenly sample the range of rational predictive strength, from very weak evidence to very strong evidence.

Design

A large pool of stimuli were created using a script in R that creates and combines pseudo-random strings to meet the constraints of target sample sizes, base rates for each string, and match rates across strings. The script then randomly selects the precise positions of elements and matches, and assigns '0' and '1' symbols for a particular stimulus. The large pool of strings was then sampled, and the stimuli were selected whose rational evidence strength was closest to the target. Targets were selected at each centile on the range from 0.50 to 0.92. In this manner, a continuum of stimuli was randomly generated. The set of items was then divided into two lists, by putting odd items in one list and even items in the other.

Participants

We recruited N = 45 participants from Amazon's Mechanical Turk. Participants were paid $2.50 for an average of under 20 minutes of work. Participants were randomly assigned to one of two groups which determined which stimuli they judged.

Procedure

Participants responded to 22 stimuli with varying strengths of regularities. Each participant was randomly assigned one of two sets of stimuli (the even list or the odd list) with roughly equivalent rational probabilities. Participants responded to each randomly ordered stimulus from the list by predicting the target and estimating the probability that their prediction was correct.

Results & Discussion

Figure 2 shows the relationship between the proportion of unique participant predictions and the rational model. Overall, the correlation between choice proportions and the rational approach was fairly strong, but systematically non-linear: as the rational probability for the stimuli increases from 0.50 the mean choice probability quickly increases. After that, median confidence increases slowly or not at all.

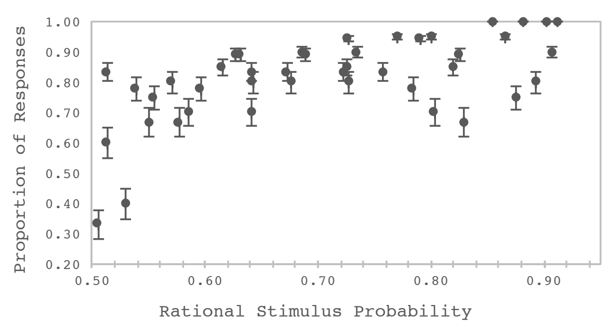

Figure 2: Proportion of Rationally Consistent Responses

**All error bars represent +/- one standard error*

People did systematically deviate from the rational model predictions for some particular stimuli. Upon examination, it appears that the regularities in those stimuli may be more spatially disperse with more interruptions dispersed rather than clustered. These effects may be investigated in future studies.

Figure 3 presents the probability estimates in light grey made by each participant for each stimulus. The median responses are shown by darkened black circles. The dashed line represents the rational model predictions.

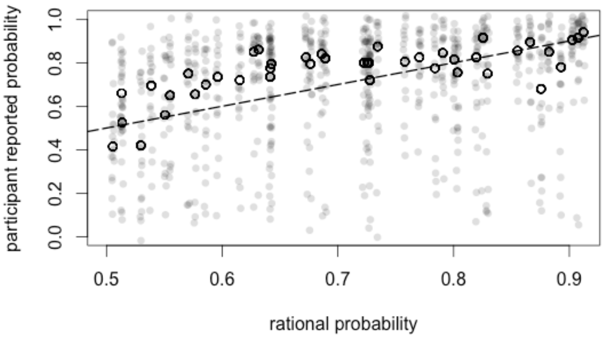

Figure 3: Response Probability vs. Rational Probability

Participants confidence increased with the rational model's probability estimate. The median participant responses by stimulus correlate to the rational probabilities by a factor of 0.71 (S.E.=0.03). The correlation is strongest at the extremes of 0.50 to 0.55 and 0.75 to 0.90, while participants' probability estimates systemically exceed the rational results on the interval of 0.55 to 0.75. This suggests a non-linear relationship with the rational model, where people may behave with overconfidence on this range.

Experiment 2

The purpose of Experiment 2 is to understand how people's predictions are influenced by intra-string regularities versus inter-string regularities.

Design

Experiment 2 used a 2x2x2 factorial design resulting in 8 groups of stimuli that are rationally distinct across groups and rationally similar within groups. The three factors examined were: 1) observed proportion of intra-string regularities (low/high), 2) observed proportion of inter-string regularities (low/high), and 3) agreement or conflict between the rational predictions corresponding to the observed intra- and inter-string regularities. The first two factors vary the degree of evidence coming from within the string, and the degree evidence coming from the analogy or relational correspondence with the other string. The strongest evidence holds when those two sources of evidence point in the same direction (*agree* stimuli); when they conflict rational evidential strength is reduced.

Participants

We recruited N = 52 participants from Amazon's Mechanical Turk. Participants were paid $3.00 for an average of under 30 minutes of work. Participants were randomly assigned to one of six groups, which determined which stimuli they judged.

Procedure

Participants responded to three stimuli from each group for a total of twenty-four responses per participant. Six sets of twenty-four stimuli were created and randomly ordered to minimize the possibility that anomalies in the stimuli or anomalies in the order of presentation would bias results. Participants were randomly assigned one of these six sets. Participants responded to each stimulus by predicting the target and estimating the probability that their prediction is correct.

Results & Discussion

We present results in for Experiment 2 that distinguish between stimuli in which intra- and inter-string regularities lead to predictions that "Agree" versus "Conflict." Agree stimuli elicited responses that are consistent with the presented regularities indicating that participants were mostly able to find at least some basis for their predictions, even when regularities occurred at a low rate. For agree stimuli, participant responses are designated as either rationally consistent or rationally inconsistent since both intra- and inter-string regularities compel the same prediction. Figure 4 shows the proportion of predictions for Agree stimuli that are rationally consistent for each other design factor.

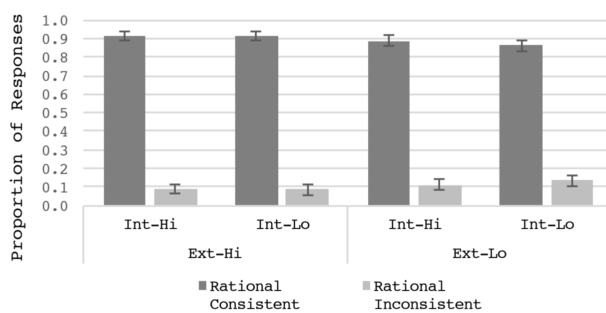

Figure 4: Proportion of Predictions in Agree Stimuli

Participant predictions were generally consistent with the rational predictions with a rate of 0.90 for the highest strength regularities and a rate of 0.85 for the lowest.

Rational probability estimates for the Agree stimuli are shown in Figure 5 below in the plots on the left and right respectively. Participants' self-reported probabilities were systematically higher than the rational model to an extent roughly consistent with the results from Experiment 1 over this rational range. More importantly, though, the main effects between intra-string and inter-string regularities are statistically significant and follow a relative pattern consistent with the rational model predictions.

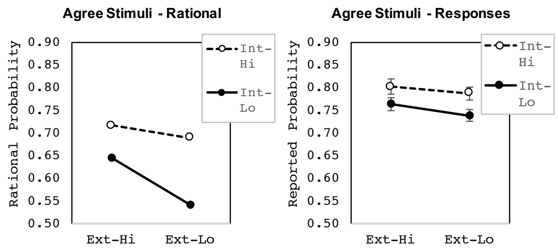

Figure 5: Rational & Response Probabilities in Agree Stimuli

Figure 6 shows the proportion of predictions for Conflict stimuli that are either consistent with the intra-string regularity (internal consistent) or the inter-string regularity (external consistent) for each other design factor.

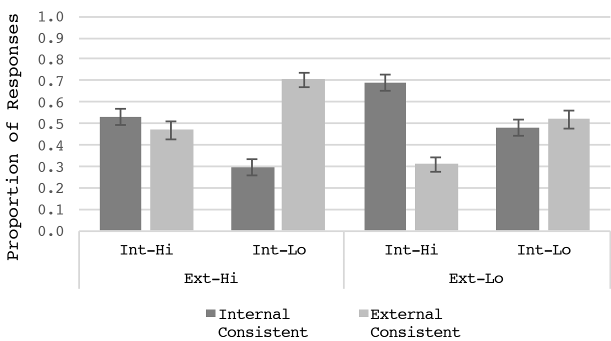

Figure 6: Proportion of Predictions in Conflict Stimuli

Participant predictions are much less consistent with the rational choices when the intra- and inter-string regularities lead to conflicting predictions. This is consistent with the idea that the relative strengths of intra- and inter-string regularities influence the chance that a participant would make a choice consistent with the rational choice. Participant predictions exhibited quantitatively symmetric preference patterns across these two conditions indicating that there does not appear to be a preference for intra- versus inter-string regularities on whole.

Participant probability estimates for Conflict stimuli demonstrate a distinct break from the rational results. Figure 7 shows the rational and participant probability estimates for the Conflict stimuli. Conflict stimuli results are further distinguished by whether the participant predictions were internal consistent or external consistent, respectively.

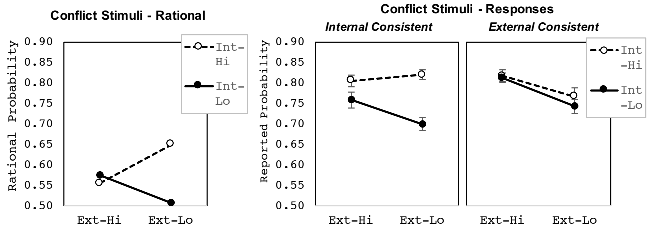

Figure 7: Rational & Response Probabilities in Conflict Stimuli

The rational analysis produces a substantial drop in probability for the hi-hi case, whereby the competing regularities are largely offsetting. Although this effect is borne out in participant predictions, participant confidence does not drop. Rather people indicate probabilities consistent with only integrating the evidence for their chosen prediction, neglecting the evidence for the other prediction. Although some participants were consistent across the three hi-hi stimuli choosing only inter-string or intra-string predictions, just as many people switched the basis of their prediction leading us to conclude that a possible individual preference for either intra- or inter-string regularities could not explain the effect.

General Discussion

This paper describes an initial investigation into the rationality of analogy that relies on a probabilistic basis. First described by Mill in 1882, this approach holds that the odds of sampling pairwise correspondence in otherwise independent systems is greater when a greater degree of pairwise correspondence between those systems truly exists. We developed a plausible minimal Bayesian model that instantiates Mill's proposal and applied this model to a rational analysis of pairs of aligned simple strings.

We compared the results of the rational analysis to the results of two psychological experiments. The first showed that people integrate observed patterns probabilistically and in a manner that correlates strongly with results of the rational analysis. People appear to over-estimate probability as the rational results increase from 0.50, but their estimates converge with the rational model around 0.75 and above.

The second experiment found that when intra- and inter-string regularities lead to the same prediction, people integrate varying combinations of strengths of these regularities in strong correspondence to the patterns of the rational analysis. But when the regularities lead to conflicting predictions, people fail to integrate across the regularities in a manner consistent with the rational analysis, instead only estimating probability according to the regularity that is consistent with their chosen prediction. Moreover, there was no statistically discernable pattern in the choice of the prediction including differences in individual preferences.

Implications for Mill's Theory

Our goal was to construct the simplest possible situation that would allow us to explore the relationship between Mill's theory and human behavior. What we found was that Mill's theory provides a meaningful guide, but discrepancies also point to complexities of the human reasoning process. The positive results for these simple tasks suggest that Mill's theory may provide a useful apparatus to understand peoples' reasoning. Still, the stimuli may be so simple that many different strategies would yield Mill-Like response patterns.

The results of Experiment 1 and the results of Experiment 2 Agree stimuli show that participants' behavior is correlated with predictions based on Mill's theory of rational analogy. These results lend credibility to using a Millian model to gain insight to how and why people use analogies.

The Conflict stimuli in Experiment 2, however, identified an inconsistency between the rational model and observed behavior. If people behave rationally, then equal evidence for competing predictions made on the basis of opposing regularities should effectively "cancel out," lowering the overall probability of a correct response. Instead we observe that people appear to dismiss the evidence for one of the competing theories and make their predictions and confidence statements based largely on the other. However, this behavior may be a more general phenomenon about how people make inferences that is not unique to analogy. First, we did not find any significant bias in predictions consistent with intra- or inter-string predictions in conflict stimuli suggesting that people are able to consider pairwise regularities across strings in the same way they consider regularities within strings. Second, previous research has shown a similar non-rational behavior in a categorization inference task where participants expressed confidence in their categorizations consistent only with the evidence for their chosen categories, neglecting evidence for fits in other categories (Murphy & Ross, 1994).

Two criticisms have been made about Mill's theory: the "counting problem" and the "problem of bias"(Bartha, 2013). The counting problem is about the seemingly arbitrary choices in defining the representation or schema over which analogies can be made. The problem of bias is one about the search and discovery of analogical mappings. Neither criticism is about analogy fundamentally, but rather about other process (i.e., representation and search) that are necessarily implicated in analogical cognitive process. These processes are independent from the basis of Mill's theory. Further, we do not promote Mill's theory as description of the analogical cognitive process. Rather, a rational basis could help answer why analogical processes operate as they do. Cognitive mechanisms need not have a rational basis, and even when they do, we should not mistake that basis as necessary but rather well adapted for the specified task.

Future Directions

We consider Mill's theory akin to the modern computational theories of analogy, but oriented toward the problem of induction with incomplete information rather than analogical discovery in well described domains. It is not hard to envision, for example, how pairwise probabilistic analysis in a more complex structure might produce qualitatively similar patterns to SMT analysis (Gentner, 1983). Binary strings are a small first step toward integrating probability into pairwise systems, but they minimize the core characteristic of analogy—relational similarity. The next step is to instantiate Mill's theory within more general relational configurations and compare results to people and other models.

Acknowledgments

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*. New York: Oxford Uniersity Press.

Bartha, P. (2013). Analogy and Analogical Reasoning. In *The Stanford Encyclopedia of Philosophy*.

Bayes, T., & Price, R. (1763). An Essay towards solving a Problem in the Doctrine of Chances. *Philosophical Transactions of the Royal Society of London*, *53*, 370–418.

Chalmers, D. J., French, R. M., & Hofstadter, D. R. (1991). High-Level Perception, Representation, and Analogy: A Critique of Artificial Intelligence Methodology. *Journal of Experimental and Theoretical Artificial Intelligence*, (March).

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.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). Structre Mapping Engine. *Artificial Intelligence*, *41*, 1–63.

Gentner, D. (1983). Structure-Mapping: A Theoretical Framework for Analogy*. *Cognitive Science*, *7*(2), 155–170.

Goldstone, R. L. (1994). Similarity, Interactive Activation, and Mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(1), 3–28.

Goldstone, R. L., Landy, D., & Brunel, L. C. (2011). Improving perception to make distant connections closer. *Frontiers in Psychology*, *2*(December), 385.

Harrod, R. (1956). *Foundations of Inductive Logic*. London: Macmillan.

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

Hofstadter, D. R. (1995). *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought*. Basic Books.

Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science: A Multidisciplinary Journal*, *13*, 295–355.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, *104*(3), 427–466.

Laplace, P.-S. (1825). *Philosophical Essay on Probabilities:* Springer Science & Business Media.

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), 1111–1122.

Lu, H., Chen, D., & Holyoak, K. J. (2012). Bayesian analogy with relational transformations. *Psychological Review*, *119*(October 2015), 617–648.

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

Murphy, G. L., & Ross, B. H. (1994). Predictions From Uncertain Categorizations. *Cognitive Psychology*.

Rumelhart, D. E., & Norman, D. A. (1981). Analogical processes in learning. In *Cognitive Skills and Their Acquisition* (335–359).

Wasserman, L. (2000). Bayesian Model Selection and Model Averaging. *Journal of Mathematical Psychology*, *44*(1), 92–107.