Generative Inferences Based on a Discriminative Bayesian Model of Relation Learning

Dawn Chen[1] (sdchen@ucla.edu)

Hongjing Lu[1, 2] (hongjing@ucla.edu)

Keith J. Holyoak[1] (holyoak@lifesci.ucla.edu)
Departments of Psychology[1] and Statistics[2]
University of California, Los Angeles
Los Angeles, CA 90095 USA

Abstract

Bayesian Analogy with Relational Transformations (BART) is a discriminative model that can learn comparative relations from non-relational inputs (Lu, Chen & Holyoak, 2012). Here we show that BART can be extended to solve inference problems that require generation (rather than classification) of relation instances. BART can use its generative capacity to perform hypothetical reasoning, enabling it to make quasi-deductive transitive inferences (e.g., "If *A* is larger than *B*, and *B* is larger than *C*, is *A* larger than *C*?"). The extended model can also generate human-like instantiations of a learned relation (e.g., answering the question, "What is an animal that is smaller than a dog?"). These modeling results suggest that discriminative models, which take a primarily bottom-up approach to relation learning, are potentially capable of using their learned representations to make generative inferences.

Keywords: Bayesian models; generative models; discriminative models; relation learning; transitive inference; deduction; induction; hypothetical reasoning

Introduction

Generative and Discriminative Models

Bayesian models of inductive learning can be designed to focus on learning either the probabilities of observable features given concepts (generative models) or the probabilities of concepts given features (discriminative models; Friston et al., 2008; Mackay, 2003). Generative models are especially powerful as they are capable of not only classifying novel instances of concepts (using Bayes' rule to invert conditional probabilities), but also generating representations of possible instances. In contrast, discriminative models focus directly on classification tasks, but do not provide any obvious mechanism for making generative inferences.

In recent years, generative Bayesian models have been developed to learn complex concepts based on relational structures (e.g., Goodman, Ullman & Tenenbaum, 2011; Kemp & Jern, 2009; Kemp, Perfors & Tenenbaum, 2007; Tenenbaum, Kemp, Griffiths & Goodman, 2011). Representations of alternative relational structures are used to predict incoming data, and the data in turn are used to revise probability distributions over alternative structures. The highest level of the structure typically consists of a formal grammar or a set of logical rules that generates alternative relational "theories", which are in turn used to predict the observed data. That is, the set of possible relational structures is provided to the system by specifying a grammar that generates them.

Despite their impressive successes, there are some reasons to doubt whether the generative approach provides an adequate basis for all psychological models of relation learning. Since the postulated grammar of relations is not itself learned, the generative approach implicitly makes rather strong nativist assumptions. Moreover, generative models of relation learning do not fit the intuitive causal direction. For example, it seems odd to claim that a binary relation such as *larger than* somehow acts to causally generate an ordered pair (e.g., <dog, cat>) that constitutes an instantiation of the relation. It seems more natural to consider how observable features of the objects in the ordered pair give rise to the truth of the relation, i.e., to apply a discriminative approach.

Discriminative Models of Relation Learning

Recently, discriminative models have also been applied to relation learning. Silva, Heller, and Ghahramani (2007) developed a discriminative model for relational tasks such as identifying classes of hyperlinks between webpages and classifying relations based on protein interactions. Although their model was developed to address applications in machine learning, the general principles can potentially be incorporated into models of human relational learning. One key idea is that an *n*-ary relation can be represented as a function that takes ordered sets of *n* objects as its input and outputs the probability that these objects instantiate the relation. The model learns a representation of the relation from labeled examples, and then applies the learned representation to classify novel examples. A second key idea is that relation learning can be facilitated by incorporating *empirical priors*, which are derived using some simpler learning task that can serve as a precursor to the relation learning task.

These ideas were incorporated into *Bayesian Analogy with Relational Transformations* (BART), a discriminative model that can learn comparative relations from non-relational inputs (Lu, Chen & Holyoak, 2012). Given

independently-generated feature vectors representing pairs of animals that exemplify a relation, the model acquires representations of first-order comparative relations (e.g., *larger*, *faster*) as weight distributions over the features. Learning is guided by empirical priors for the weight distributions derived from initial learning of one-place predicates (e.g., *large*, *fast*). BART's learned relations support generalization to new animal pairs, allowing the model to discriminate between novel pairs that instantiate a relation and those that do not. Moreover, BART's learned weight distributions can be systematically transformed to solve analogies based on higher-order relations (e.g., *opposite*).

BART has thus demonstrated promise as a discriminative model of relation learning, which does not presuppose an innate grammar of relations. However, the challenge remains to extend the model to tasks requiring generative inferences. For example, people are able to construct actual instantiations of relations, answering questions such as, "What is an animal that is smaller than a dog?" (Although one might suppose that such questions could be answered by undirected trial-and-error, we shall see that people's answers are often systematically guided by their representations of the relation and of the animal provided as a cue.) Another challenging task is purely hypothetical reasoning, which requires making inferences about arbitrary instances of the relation. Comparative relations such as *larger* exhibit the logical properties of transitivity and asymmetry, supporting deductions such as "If *A* is larger than *B*, and *B* is larger than *C*, then *A* is larger than *C*." Children as young as five or six years can make such transitive inferences reliably (Halford, 1992; Goswami, 1995; Kotovsky & Gentner, 1996). In the present paper we describe an extension of the BART model that addresses these challenges of making generative inferences.

BART Model of Relation Learning

Domain and Inputs

We focus on the same domain and inputs used in the initial BART project (Lu et al., 2012): the domain of comparative relations between animal concepts (e.g., a cow is larger than a sheep). To establish the "ground truth" of whether various pairs of animals instantiate different comparative relations, Lu et al. used a set of human ratings of animals on four different continua (size, speed, fierceness, and intelligence; Holyoak & Mah, 1981). These ratings made it possible to test the model on learning eight different comparative relations: *larger*, *smaller*, *faster*, *slower*, *fiercer*, *meeker*, *smarter*, and *dumber*.

Each animal concept is represented by a real-valued feature vector. In order to avoid the perils of hand-coded inputs (i.e., the possibility that the model's successes may be partly attributable to the foresight and charity of the modelers), we use what we call "Leuven vectors." These representations are derived from norms of the frequencies with which participants at the University of Leuven generated features characterizing 129 different animals (De Deyne et al., 2008; see Shafto, Kemp, Mansinghka, & Tenenbaum, 2011). Each animal in the norms is associated with a set of frequencies across more than 750 features. We created vectors of length 50 based on the 50 features most highly associated with the subset of 44 animals that are also in the ratings dataset (Lu et al., 2012). Figure 1 provides a visualization (for 30 example animals and the first 15 of the 50 features) of these high-dimensional and distributed representations, which might be similar to the representations underlying people's everyday knowledge of various animals.

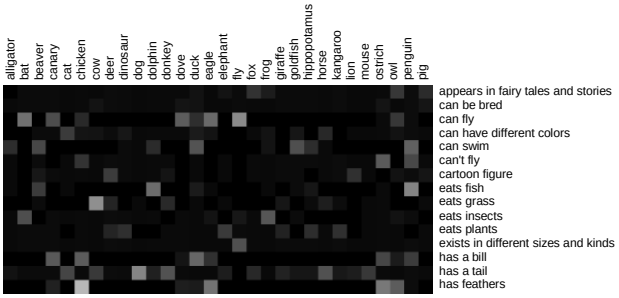

Figure 1: Illustration of Leuven vectors (reduced to 15 features to conserve space) for some example animals. The cell intensities represent feature values (light indicates high values and dark indicates low values).

Relations as Weight Distributions

BART represents a relation using a joint distribution of weights, **w**, over object features. A relation is learned by estimating the probability distribution $P(\mathbf{w} \mid \mathbf{X_S}, \mathbf{R_S})$, where $\mathbf{X_S}$ represents the feature vectors for object pairs in the training set, the subscript **S** indicates the set of training examples, and $\mathbf{R_S}$ is a set of binary indicators, each of which (denoted by *R*) indicates whether a particular object (or pair of objects) instantiates the relation or not. The vector **w** constitutes the learned relational representation, which can be interpreted as weights reflecting the influence of the corresponding feature dimensions in **X** on judging whether the relation applies. The weight distribution can be updated based on examples of ordered pairs that instantiate the relation. Formally, the posterior distribution of weights can be computed by applying Bayes' rule using the likelihood of the training data and the prior distribution for **w**:

$$P(\mathbf{w} \mid \mathbf{X_S}, \mathbf{R_S}) = \frac{P(\mathbf{R_S} \mid \mathbf{w}, \mathbf{X_S}) P(\mathbf{w})}{\int_{\mathbf{w}} P(\mathbf{R_S} \mid \mathbf{w}, \mathbf{X_S}) P(\mathbf{w})}. \quad (1)$$

The likelihood is defined as a logistic function for computing the probability that a pair of objects instantiates the relation, given the weights and feature vector:

$$P(R = 1 \mid \mathbf{w}, \mathbf{x}) = \frac{1}{1 + e^{-\mathbf{w}^{\mathbf{T}}\mathbf{x}}}. \quad (2)$$

The prior, $P(\mathbf{w})$, is a Gaussian distribution and is constructed using a bottom-up approach in which initial learning of simple concepts provides *empirical priors* that guide subsequent learning of more complex concepts. Specifically, BART extracts empirical priors from weight distributions for one-place predicates such as *large* to guide the acquisition of two-place relations such as *larger*. Lu et al. (2012) trained BART on the eight one-place predicates (e.g., *large*, *small*, *fierce*, *meek*) that can be formed using the extreme animals at each end of the four relevant continua (size, speed, ferocity, and intelligence).

After learning the joint weight distribution that represents a relation, BART discriminates between pairs that instantiate the relation and those that do not by calculating the probability that a target pair $\mathbf{x}_T$ instantiates the relation *R*:

$$P(R_T = 1 | \mathbf{x}_T, \mathbf{X_S}, \mathbf{R_S}) = \int_{\mathbf{w}} P(R_T = 1 | \mathbf{x}_T, \mathbf{w}) P(\mathbf{w} | \mathbf{X_S}, \mathbf{R_S}). \quad (3)$$

Although the general framework of the relation learning model is straightforward, the calculations of the normalization term in Eq. (1) and the integral in Eq. (3) are intractable, lacking analytic solutions. As in Silva, Heller, and Gharamani (2007), we employed the variational method developed by Jaakkola and Jordan (2000) for Bayesian logistic regression to obtain closed-form approximations to the posterior weight distribution $P(\mathbf{w} | \mathbf{X_S}, \mathbf{R_S})$ and the predictive probability $P(R_T = 1 | \mathbf{x}_T, \mathbf{X_S}, \mathbf{R_S})$.

Extension to Generative Inference

The goal of the present paper is to endow BART with generative abilities, allowing it (for example) to complete a partially-instantiated relation, answering questions such as, "What is an animal that is smaller than a dog?" We use the weight representation for a relation learned by BART to construct a new generative model for the completion task. When presented with a cue relation (e.g., *smaller*) and a cue object (e.g., dog), the model produces possible responses for the remaining object (e.g., cat) so that the ordered object pair satisfies the relation. More specifically, given the features of an object *B*, $\mathbf{x}_B$, and the knowledge that relation *R* holds for the object pair (*A*, *B*), the model generates a probability distribution for the features of object *A*, $\mathbf{x}_A$, by making the following inference:

$$P(\mathbf{x}_A | \mathbf{x}_B, R = 1) \propto P(R = 1 | \mathbf{x}_A, \mathbf{x}_B) P(\mathbf{x}_A | \mathbf{x}_B). \quad (4)$$

The likelihood term, $P(R = 1 | \mathbf{x}_A, \mathbf{x}_B)$, is the probability that relation *R* holds for a particular hypothesized object *A*, $\mathbf{x}_A$, and the known object *B*, $\mathbf{x}_B$. It is defined using a logistic function, just as in Eq. (2):

$$P(R = 1 | \mathbf{x}_A, \mathbf{x}_B) = \frac{1}{1 + e^{-\mathbf{w}_1^{\mathrm{T}} \mathbf{x}_A - \mathbf{w}_2^{\mathrm{T}} \mathbf{x}_B}}. \quad (5)$$

Relative to Eq. (2), we have only introduced small differences in the notation. The learned relational weights, **w**, are written as two separate halves: weights associated with the first relational role ($\mathbf{w}_1$) and weights associated with the second relational role ($\mathbf{w}_2$). Similarly, the feature vector **x** for a pair of objects is separated into the feature vector for object *A* ($\mathbf{x}_A$) and the feature vector for object *B* ($\mathbf{x}_B$).

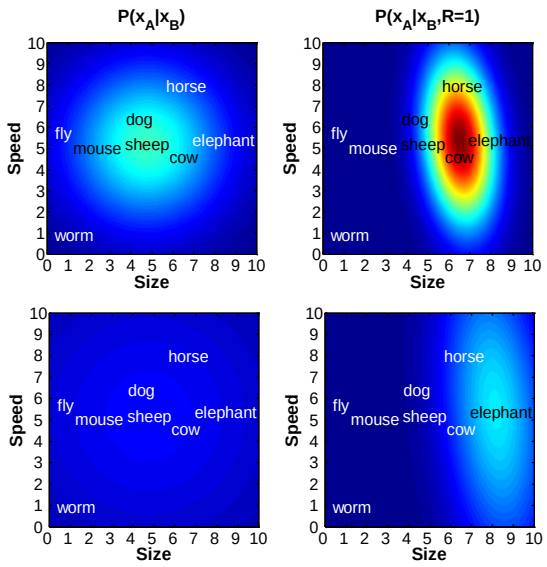

Figure 2: Illustration of the generative model for inferring an animal that is larger than a sheep. Colors annotate probability densities (red indicates high values and blue indicates low values). The top panel shows the prior and posterior distributions with $\sigma^2 = 7$ (favoring similarity-based completions such as *cow*), and the bottom panel shows the prior and posterior with $\sigma^2 = 25$ (favoring "landmark" completions such as *elephant*). Various animals are represented in the two-dimensional space based on their size and speed ratings. The posterior was generated using the relational weights that BART learned from the full ratings input (i.e., all four dimensions).

The prior for the features of object *A*, $P(\mathbf{x}_A | \mathbf{x}_B)$, is the conditional distribution given the features of object *B*. It is defined as the following:

$$P(\mathbf{x}_A | \mathbf{x}_B) = N(\mathbf{x}_B, \sigma^2 \mathbf{I}). \quad (6)$$

We assume that object *B* (the referent) serves a starting point for generating object *A*, so the means of $P(\mathbf{x}_A | \mathbf{x}_B)$ are taken to be the feature values of object *B*, reflecting a certain degree of semantic dependency between the two objects (i.e., people's tendency to think of *A* objects that are similar to *B*). The prior also encodes the assumptions that the features of *A* are uncorrelated and have the same variance σ^2, the value of which is a free parameter reflecting the strength of the model's preference for generating *A* objects that are similar to *B*.

Our generative model infers a feature distribution for object *A* that reflects a compromise between (1) maximizing

the semantic similarity of *A* and *B*, which is reflected in the prior term; and (2) maximizing the probability that the relation holds, which is reflected in the likelihood term. We adapted the variational method to estimate the posterior distribution, using the following update rules for the mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{V}$ of the feature distribution, as well as the variational parameter ξ:

$$\begin{aligned}\mathbf{V}^{-1} &= \frac{\mathbf{I}}{\sigma^2} + 2\lambda(\xi)\mathbf{w}_1\mathbf{w}_1^{\mathrm{T}},\\ \boldsymbol{\mu} &= \mathbf{V}\left(\frac{\mathbf{I}}{\sigma^2}\mathbf{x}_B + \frac{\mathbf{w}_1}{2} - 2k\lambda(\xi)\mathbf{w}_1\right),\\ \xi^2 &= \mathbf{w}_1^{\mathrm{T}}\left(\mathbf{V} + \boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}}\right)\mathbf{w}_1,\end{aligned} \tag{7}$$

where $\lambda(\xi) = \dfrac{\tanh\left(\frac{1}{2}(\xi + k)\right)}{4(\xi + k)}$ and $k = \mathbf{w}_2^{\mathrm{T}}\mathbf{x}_B$.

Figure 2 illustrates the operation of the model in generating an animal (*A*) that is larger than a sheep (*B*). The feature distribution for *A* is updated from a prior favoring some degree of similarity between the two animals (left panel; top: high similarity, bottom: low similarity) to a posterior distribution after taking into consideration the relation (i.e., *larger*) instantiated by the animals (right panel). These distributions are shown in a simplified two-dimensional feature space (the size and speed ratings for animals; Holyoak & Mah, 1981).

Modeling Transitive Inference

Comparative relations such as *larger* exhibit the logical properties of transitivity and asymmetry, supporting deductions such as, "If *A* is larger than *B* and *B* is larger than *C*, then *A* is larger than *C*." Such hypothetical reasoning seems to depend on the ability to generate arbitrary instantiations of the relation without any guidance from object features (as the object representations are semantically empty). Our first test evaluated whether the generative extension of BART enables transitive inferences on comparative relations using arbitrary hypothetical instances.

Operation of the Model

The basic approach to transitive inference is straightforward: The model "imagines" objects *A*, *B*, and *C* that instantiate the two given premises, as in the example above, and then tests the unstated relationship for the pair <*A*, *C*>. If the *larger* relation that BART has learned is indeed transitive, then any such instantiation will satisfy the conclusion, "*A* is larger than *C*." This test is done repeatedly, in essence searching for a counterexample. If no counterexample is ever found, the transitive inference is accepted.

Specifically, for each of the eight comparative relations that BART learned, we first let the model "imagine" an animal *B* (because the statement "*A* is larger than *B*" implies that *B* is the referent against which *A* is being compared) by sampling a feature vector from a distribution representing the animal category. This is a Gaussian distribution with a mean vector and covariance matrix that were directly estimated from the feature vectors of the 44 animals in the Leuven dataset that are included in the ratings dataset.

Given the sampled animal *B*, the generative model constructs a distribution for animal *A* (e.g., to satisfy the premise that "*A* is larger than *B*") by letting *B* fill the second role of the relevant relation. Similarly, the model constructs a distribution for animal *C* (e.g., to satisfy the premise that "*B* is larger than *C*") by letting *B* fill the first role of the same relation. Next, the model creates feature representations for specific animals *A* and *C* by setting their feature vectors, $\mathbf{x}_A$ and $\mathbf{x}_C$, to be the means of the inferred feature distributions for *A* and *C*, respectively. Note that these "imagined" animals are hypothetical: although their features are sampled from the distribution of animal features, the results will seldom correspond to actual animals. To ensure that the premises have actually been satisfied, the model accepts the imagined animal *A* only if $P(R=1|\mathbf{x}_A,\mathbf{x}_B) > 0.5$ and $P(R=1|\mathbf{x}_B,\mathbf{x}_A) < 0.5$, and the imagined animal *C* only if $P(R=1|\mathbf{x}_B,\mathbf{x}_C) > 0.5$ and $P(R=1|\mathbf{x}_C,\mathbf{x}_B) < 0.5$.

Finally, if $\mathbf{x}_A$ and $\mathbf{x}_C$ have been accepted as satisfying the premises, the model calculates both $P(R=1|\mathbf{x}_A,\mathbf{x}_C)$, denoting the probability that *A* is larger than *C*, and $P(R=1|\mathbf{x}_C,\mathbf{x}_A)$, denoting the probability that *C* is larger than *A*. The model concludes that the relation holds for the pair <*A*, *C*> (and not for <*C*, *A*>) if $P(R=1|\mathbf{x}_A,\mathbf{x}_C) > 0.5$ and $P(R=1|\mathbf{x}_C,\mathbf{x}_A) < 0.5$, implying that a counterexample has not yet been found to refute the transitive inference.

We conducted tests of transitive inference using the relational representations that BART learned based on 100 randomly-chosen training pairs. For comparison, we also tested a baseline model that substituted an uninformative prior for the empirical prior that guides BART's relation learning (see Lu et al., 2012). For each of the eight comparative relations, the relation learning model was run ten times, each time with a different set of training pairs and resulting in a different learned weight distribution. For each of these learned weight distributions, we let the model generate 100 *A*-*B*-*C* triads satisfying the premises, testing the relevant relationship between *A* and *C* for each triad. To assess the influence of the free parameter in model predictions, the tests were conducted multiple times with different values of σ^2 ranging from 1 to 1000. The strongest tests are those in which σ^2 is set at low values, creating a strong prior preference that *A*, *B*, and *C* are similar to one another. When the similarity constraint is strong, the model is forced to generate animals that are similar on the relevant dimension, and hence more likely to yield a counterexample. When the value of σ^2 was reduced below 1, the models produced many instantiations that did not satisfy the required premises (i.e., *A* > *B*, *B* > *C*, and not vice versa). We therefore treated the value of 1 as the

minimal value of σ^2 that yields triplets of animals with discriminable values on the relevant dimension.

Results and Discussion

Figure 3 shows the mean proportion correct (i.e., the mean proportion of triads that satisfy the conclusion based on transitive inference) for BART and the baseline model as a function of σ^2. These results are averaged over the eight comparative relations. The critical result is that the BART's accuracy remains constant at 100% as σ^2 is reduced to the effective minimal value of 1. Thus, BART demonstrates what may be considered an inductive approximation to deduction: despite exhaustive search for a counterexample to the transitive inference, no counterexample is ever found. In contrast, the baseline model often fails to infer that $A > C$ (and not vice versa) even when the value of σ^2 is as large as 100.

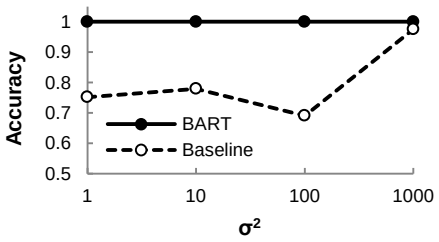

Figure 3: Mean proportion correct on the transitive inference task for BART and baseline model, as a function of the variance parameter. These results are averaged across the eight comparative relations.

Animal Generation Task

A second evaluation of the model involves predicting the distribution of human responses in an animal generation study conducted using Amazon Mechanical Turk. In this free-generation study, participants typed responses to queries of the form, "Name an animal that is larger than a dog." They were instructed to enter the first animal that came to mind. Four comparative relations (*larger*, *smaller*, *faster*, and *slower*) and nine cue animals (shark, ostrich, sheep, dog, fox, turkey, duck, dove, and sparrow) were used. At least 50 responses were collected for each of the 36 relation-animal pairs. To minimize learning across trials, we asked each participant to answer only two questions about a single animal: either *larger* and then *slower*, *slower* and then *larger*, *faster* and then *smaller*, or *smaller* and then *faster*.

The same relation-animal pairs were presented to the model after it had been trained on the relevant relations. For each question, the model produces a continuous posterior distribution for the feature vector of the missing animal using Eq. (4). This distribution was used to calculate the probability densities for the feature vectors of all animals among the human responses that had Leuven vectors. These probability densities were normalized to produce a discrete

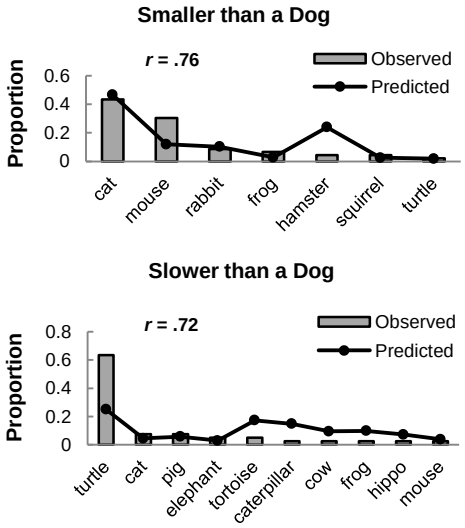

Figure 4: Observed human response proportions and BART's predictions for the queries, "Name an animal that is smaller than a dog" (top), and "Name an animal that is slower than a dog" (bottom).

probability distribution over the animals included in the human responses. The model's predicted probabilities were averaged across the ten runs.

The human results were complex, and here we report only a partial and preliminary attempt to make a comparison with model predictions. Qualitatively, human responses were dominated by two trends: (1) reporting an animal similar to the cue animal and fitting the cue relation (e.g., *cat* for "smaller than a dog"), or (2) reporting a "landmark" animal at an extreme of the continuum (e.g., *turtle* for "slower than a dog"). The landmark animal coupled with the cue animal provides an ideal example of the cue relation. This tradeoff between reporting animals that are similar to the cue animal and reporting animals that are landmarks for the cue relation (and usually more dissimilar to the cue animal) is captured by the single free parameter in the generative module, σ^2. As explained earlier (see Figure 2), a low σ^2 results in a response distribution that favors animals similar to the cue animal, whereas a high σ^2 leads to a preference for response animals that are more likely to satisfy the cue relation with respect to the cue animal (i.e., landmark animals for the cue relation).

To reflect the unique pattern of human responses to each question, the variance parameter in the generative model was chosen separately for each question (from the values, 1, 5, 10, 50, and 100) to maximize the average of Pearson's r and Spearman's ρ (rank-order) correlations between the model's predicted probabilities and the observed response proportions for that question. Here we present results for two illustrative questions. The top panel of Figure 4 shows the model's predicted response distribution and the human response distribution for the request, "Name an animal that is smaller than a dog." The human response pattern reveals a strong influence of semantic similarity between the cue

animal and generated animal. The most common human response was *cat*, followed by *mouse* (the landmark animal for the *smaller* relation). With σ^2 = 10, the correlation between the model predictions and the human response pattern was r = .76.

The bottom panel of Figure 4 depicts the model predictions and human response pattern for the request, "Name an animal that is slower than a dog." For this question, the most common response was the landmark animal *turtle*. With σ^2 = 50, the correlation between the model predictions and the human response pattern was r = .72. The higher variance assumed for this question (relative to that for the *smaller* question) reflects the dominance of the landmark response for the *slower* question.

Note that even though the two questions use the same cue animal (*dog*), different sets of animals were generated depending on the cue relation, revealing that humans do take relations into consideration in this free generation task. The model showed a similar pattern of results.

Conclusions

These results provide initial evidence that a discriminative model of relation learning, BART (Lu et al., 2012), can be extended to yield generative inferences. These inferences can involve relations between either hypothetical (in the case of transitive inference) or actual (in the case of the animal generation task) objects. In the latter free generation task, preliminary analyses indicate that BART achieves some success in modeling human response patterns.

The model's ability to make transitive inferences based on relations it has learned in a bottom-up fashion from examples illustrates the potential power of the discriminative approach to relation learning. Importantly, BART is not endowed with any notion of what a "transitive and asymmetric" relation is (though like a 6-year-old child, it *is* endowed with sufficient working memory to integrate two relations as premises). Rather, it simply uses its learned comparative relations to imagine possible object triads, and without exception concludes that the inference warranted by transitivity holds in each such triad. The model thus approximates "logical" reasoning by a systematic search for counterexamples (and failing to find any), akin to a basic mechanism postulated by the theory of mental models (Johnson-Laird, 2008). The fact that BART achieves error-free performance in the tests of transitive inference is especially impressive given that its inductively-acquired relational representations are most certainly fallible (e.g., the model makes errors in judging which of two animals close in size is the larger; see Lu et al., 2012). It turns out that imperfect representations of comparative relations, acquired by bottom-up induction, can be sufficiently robust as to yield reliable quasi-deductive transitive inferences.

Acknowledgments

Preparation of this paper was supported by grant N000140810186 from the Office of Naval Research.

References

De Deyne, S., Verheyen, S., Ameel, E., Vanpaemel,W., Dry, M., Voorspoels, W., & Storms G. (2008). Exemplar by feature applicability matrices and other Dutch normative data for semantic concepts. *Behavior Research Methods, 40*, 1030-1048.

Friston, K., Chu, C., Mourão-Miranda, J., Hulme, O., Rees, H., Penny, W., & Ashburner, J. (2008). Bayesian decoding of brain images. *NeuroImage, 39*, 181-205.

Goodman, N. D., Ullman, T. D., & Tenenbaum, J. B. (2011). Learning a theory of causality. *Psychological Review, 118*, 110-119.

Goswami, U. (1995). Transitive relational mappings in 3- and 4-year-olds: The analogy of Goldilocks and the Three Bears. *Child Development, 66*, 877-892.

Halford, G. S. (1992). Analogical reasoning and conceptual complexity in cognitive development. *Human Development, 35*, 193-217.

Holyoak, K. J., & Mah, W. A. (1981). Semantic congruity in symbolic comparisons: Evidence against an expectancy hypothesis. *Memory & Cognition, 9*, 197-204.

Jaakkola, T. S., & Jordan, M. I. (2000). Bayesian logistic regression: A variational approach. *Statistics and Computing, 10*, 25-37.

Johnson-Laird, P.N. (2008) Mental models and deductive reasoning. In L. Rips & J. Adler. (Eds.), *Reasoning: Studies in human inference and its foundations* (pp. 206-222). Cambridge, UK: Cambridge University Press.

Kemp, C., & Jern, A. (2009). Abstraction and relational learning. In Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams & A. Culotta (Eds.), *Advances in Neural Information Processing Systems, 22*, 943-951.

Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science, 10*, 307–321.

Kemp, C., & Tenenbaum, J. B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences, USA, 105*, 10687-10692.

Kotovsky L, & Gentner D. (1996). Comparison and categorization in the development of relational similarity. *Child Development, 67*, 2797-2822.

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

Mackay, D. (2003). *Information theory, inference and learning algorithms.* Cambridge, UK: Cambridge University Press.

Shafto, P., Kemp, C., Mansinghka, V., & Tenenbaum, J. B. (2011). A probabilistic model of cross-categorization. *Cognition, 120*, 1-25.

Silva, R., Heller, K., & Ghahramani, Z. (2007). Analogical reasoning with relational Bayesian sets. In M. Mella & X. Shen (Eds.), *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics*.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science, 331*, 1279-1285.