Modeling Semantic Cognition as Logical Dimensionality Reduction

Yarden Katz, Noah D. Goodman, Kristian Kersting, Charles Kemp, Joshua B. Tenenbaum
{yarden, ndg, kersting, ckemp, jbt}@mit.edu
Department of Brain and Cognitive Sciences & Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02139

Abstract

Semantic knowledge is often expressed in the form of intuitive theories, which organize, predict and explain our observations of the world. How are these powerful knowledge structures represented and acquired? We present a framework, *logical dimensionality reduction*, that treats theories as compressive probabilistic models, attempting to express observed data as a sample from the logical consequences of the theory's underlying laws and a small number of core facts. By performing Bayesian learning and inference on these models we combine important features of more familiar connectionist and symbolic approaches to semantic cognition: an ability to handle graded, uncertain inferences, together with systematicity and compositionality that support appropriate inferences from sparse observations in novel contexts.

Problems of semantic cognition

A person's store of common-sense knowledge about the world is vast, but it is more than just a vast collection of facts. In many domains, the mind organizes what it knows into large-scale systems, with structure at multiple levels of abstraction. These abstract systems of representation—called schemas, or theories—are crucial for our ability to make inferences that go beyond the sparse and noisy data of perceptual experience.

How are such powerful knowledge representations acquired, structured, and used? These problems lie at the heart of semantic cognition. Early proposals focused on symbolic structures, as in the semantic networks of Collins and Quillian (1969) (hereafter CQ) for organizing categories of objects and their properties. Categories are placed in a tree-structured taxonomy, with properties located at nodes of the tree and assumed to inherit down to all categories below them. Each property needs to be stored only at the highest node of the tree where it holds generically, leading to a compact encoding of objects' properties and the ability to project known properties to novel objects.

Later work emphasized the limitations of symbolic representations in handling noisy data or exceptions, and in accounting for graded effects of similarity on people's inductive judgments. Symbolic approaches were also criticized for not providing a working account of how their abstract representations could be learned from experience.

An alternative approach to modeling abstract semantic knowledge using connectionist networks emerged in the 1980's. Hinton proposed a connectionist network that could learn abstract systems of kinship relations (Hinton, 1986), while RM explored a similar architecture for learning about categories and properties (Rogers & McClelland, 2004)—essentially the same problem that CQ treated from a symbolic perspective. While these models can handle noise and uncertainty, they also have important limitations, complementary to those of symbolic approaches. They do not naturally display the *systematicity* and *compositionality* that characterize people's intuitive theories and common-sense reasoning (Fodor & Pylyshyn, 1988). This limitation shows up most clearly when making inferences about novel entities that are only sparsely observed. Suppose that we encounter two new kinds of organisms, tufas and kibos, and we are told that kibos have some novel property (e.g., they have omulums). If we then learn that tufas are kibos, it is likely that tufas also have omulums. The CQ model makes this prediction, via property inheritance down through the taxonomy of categories. The RM network, however, does not automatically yield this inference, if it is trained on the two facts $\mathsf{is\_a}(\mathsf{tufa}, \mathsf{kibo})$ and $\mathsf{has\_a}(\mathsf{kibo}, \mathsf{omulum})$. Because the concept 'kibo' plays different roles in these two propositions—it is the object of $\mathsf{is\_a}$ and the subject of $\mathsf{has\_a}$—it is represented in different populations of units, and the effects of training on these two facts appear in non-overlapping sets of weights. The network does not equate 'kibo' in the first proposition with 'kibo' in the second proposition, and so fails to draw the obvious inference.

This example does not imply that it would be impossible to design a connectionist semantic model whose inferences did respect basic principles of systematicity and compositionality. Our point is only that the connectionist approach does not naturally capture these aspects of human inference, which are no less essential than the statistical capacities it does capture well. While it is possible that either the connectionist approach or the structured symbolic approach could be extended in some way that makes for a satisfactory solution, our aim here is to explore new alternatives for modeling abstract semantic knowledge.

We describe an approach that combines valuable capacities of both traditional paradigms: an ability to represent abstract knowledge respecting systematicity and compositionality, and hence to make appropriate inferences from sparse data in novel situations; and an ability to learn from noisy data and generalize in graded fashion based on the statistics of observed data. Our approach is based on a hierarchical Bayesian model over logical representations. This is similar in spirit to proposals in *inductive logic programming* (Muggleton & De Raedt, 1994), although these are not typically formalized explicitly as hierarchical Bayesian models (for a review of this approach see, Tenenbaum, Griffiths, and Kemp (2006)). A close relative of our approach is (Conklin & Wit-

ten, 1994), where logical theories are learned using a complexity prior. However, that approach does not attempt to construct theories with novel unobservable predicates, which is crucial to understanding many real-world domains and to our work here.

Our framework can be viewed as a kind of *dimensionality reduction* for structured logical theories. We observe data in the form of relations and attributes over a set of objects, and we infer a representation of the abstract structure underlying these data, expressed in terms of a subset of first-order predicate logic. The observations are high dimensional, and the inferred underlying abstract structure can be seen as a low dimensional 'space' into which the observed objects are embedded.

While the model instantiates a structured domain theory, it supports statistical inference via the probabilistic generative process linking the domain theory to the observed data. Bayesian inversion of this generative model allows us to infer the low-dimensional abstract structure underlying the observed data.

We illustrate our approach on simple versions of the kinship and taxonomic categorization domains where previous connectionist approaches have been developed. We show that given an appropriate probabilistic domain theory, we can make successful inferences from sparse data in novel contexts—a setting which has proven challenging for connectionist models. The problem of learning an abstract domain theory remains more difficult for our approach than for connecionist models, but we show that at least in some simple cases, our hierarchical Bayesian formulation allows the correct abstract domain theory to be inferred from observations.

Theories: structure and form

We use the term *theory* formally as a specification of a set of relations and an associated set of laws and types that govern them. Theories are instantiated by *models*, or possible worlds, corresponding to ways in which a theory may hold for a particular set of objects. For example, we can apply the theory of kinship to reason about a set of individuals in a family, and later apply it to an entirely different family. We will refer each of these collections of objects that a theory can apply to, and their associated abstract structures, as *contexts*. Relations include 'father', 'child', or 'spouse', while a law might be that 'the child of an individual's spouse is also their child.' Every family forms a context, and two models might be one where Alice is the spouse of Bob, and another in which Alice is the spouse of Carroll.

The general framework we follow is shown in Figure 2(a). In a generative fashion, our theories produce models, each of which in turn generates observations. Each theory specifies a set of *core relations*, whose values are not directly observable. These are analogous to the lower-dimensional space in numerical dimensionality reduction. The laws of the theory then relate the core relations to an observable set of *derivative* or *observable* relations (the 'high-dimensional space' of the theory.) Laws in our theories take the form of typed *Horn clauses*, commonly used in the logic and inductive logic programming literature. A proposal for Horn clauses as a psychologically plausible representation of theories is found in (Kemp, Goodman, & Tenenbaum, 2007).

A feature of our framework is that the values of derivative relations (such as mother, in kinship) can be compressed into a particular assignment of values for the core relations (such as parent), via the theory's laws. The probabilistic generative model we define favors theories that adopt as few core relations as possible, seeking the optimal compression of the given set of observations.

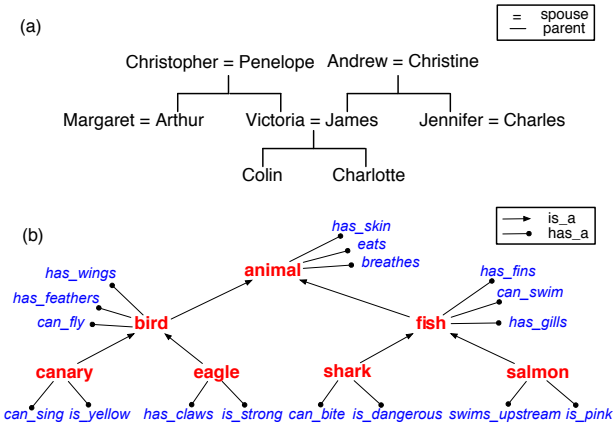

Figure 1: (a) A segment of a family tree for the kinship theory used in Hinton (1986). (b) A Collins & Quillian-like taxonomy. Categories in bold red, properties in italics blue.

(a)

Core rels: $\text{female}_C : \text{Person} \times \text{Person}$
$\text{spouse}_{CF} : \text{Person} \times \text{Person}$
$\text{child}_{CF} : \text{Person} \times \text{Person}$

Types: Person

Laws:

$\text{female}(X) \leftarrow \text{female}_C(X)$
$\text{spouse}(X, Y) \leftarrow \text{spouse}_{CF}(X, Y)$
$\text{spouse}(X, Y) \leftarrow \text{spouse}(Y, X)$
$\text{child}(X, Y) \leftarrow \text{child}_{CF}(X, Y)$
$\text{child}(X, Y) \leftarrow \text{child}(X, Z) \wedge \text{spouse}(Z, Y)$
$\text{mother}(X, Y) \leftarrow \text{female}(X) \wedge \text{child}(Y, X)$
$\text{father}(X, Y) \leftarrow \neg\text{female}(X) \wedge \text{child}(Y, X)$
$\text{daughter}(X, Y) \leftarrow \text{female}(X) \wedge \text{child}(X, Y)$

(b)

Core rels: $\text{is\_a}_{CF} : \text{Cat} \times \text{Cat}$
$\text{has\_a}_C : \text{Cat} \times \text{Prop}$

Types: Cat, Prop

Laws:

$\text{is\_a}(X, Y) \leftarrow \text{is\_a}_{CF}(X, Y)$
$\text{has\_a}(X, Y) \leftarrow \text{has\_a}_C(X, Y)$
$\text{is\_a}(X, Y) \leftarrow \text{is\_a}(X, Z) \wedge \text{is\_a}(Z, Y)$
$\text{has\_a}(X, Y) \leftarrow \text{is\_a}(X, Z) \wedge \text{has\_a}(Z, Y)$

$\text{son}(X, Y) \leftarrow \neg\text{female}(X) \wedge \text{child}(X, Y)$
$\text{wife}(X, Y) \leftarrow \text{female}(X) \wedge \text{spouse}(X, Y)$
$\text{husband}(X, Y) \leftarrow \neg\text{female}(X) \wedge \text{spouse}(X, Y)$

Table 1: Logical representations of (a) portion of the kinship theory, and (b) taxonomy theory. Core relations and functions are denoted with C and F subscripts, respectively.

A generative model for theories

In this section we expand the basic generative model (Figure 2(a)), describing in more detail how theories are represented,

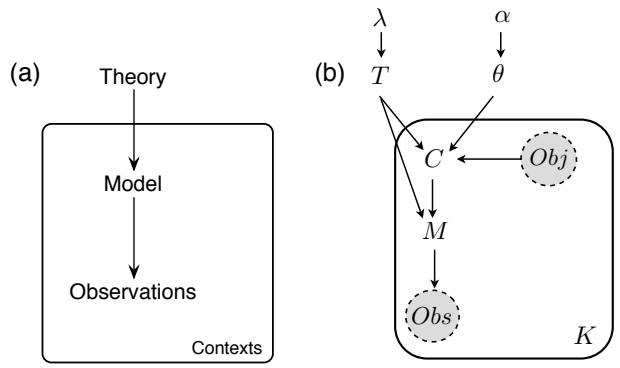

Figure 2: (a) The generic structure of a theory learning framework based on dimensionality reduction. (b) An instantiation of the generic framework in K contexts, for theories where the core relations are generated independently. Shaded nodes denote observed variables for inference given an existing theory.

how a theory generates a model of the relations over a particular set of objects, and how observations are generated from a model.

Formally, a theory T is a triple $\langle Core, Laws, \tau \rangle$ of core relations[1], laws, and types, respectively. The core relations specify an unobservable, compressed representation of the domain, while the laws are a set of rules for recovering the observable properties of the domain from this core representation. Because the laws determine the observable relations given the core relations, fixing the extension of all core relations uniquely determines a model. The core relations themselves are generated independently according to *extension weights* θ_i (each core relation R^i has its own extension weight determining the fraction of "core facts" that are expected to be true). As a further compression, we allow that some core relations are *functions*: if R is a functional relation, then for each i, $R(i, j)$ holds for at most one j (note that this reduces the number of independent core facts, since columns of R are no longer independent). More formally, the core relations are generated as follows:

1. For each $R^i \in Core$, draw a $\theta^i \sim \text{Beta}(\alpha, \beta)$.
2. For every context $k \leq K$,
   (a) Choose a group of objects $O_t \subseteq Obj^k$ that belong to each type $t \in \tau$.
   (b) Generate the core extension C_k^i for R^i: for every $a, b \in Obj^k$, R^i holds with probability $P(R^i(a, b)) = \theta_i$ when the types of a, b match the type signature of R^i (and probability 0 otherwise).
   (c) Complete the model M for C_k, by iterative application of every $L \in Laws$ to C_k, until no additional inferences are made.
   (d) Sample a set of *positive observations*, Obs^k, from M:

$$P(Obs^k) = \prod_i P(o_i \in Obs^k)$$
$$= \prod_i \begin{cases} \frac{1}{size} & o_i \in M, \\ \epsilon & o_i \notin M. \end{cases}$$

where $size$ is the number of true facts in M. A small non-zero value of ϵ allows the theory to tolerate noise in the observed data.

[1]For simplicity, we describe the generative process only for the case of binary relations; similar processes describe unary or higher-arity relations.

Note that while we focus here on the problem of learning from positive observations only, the model can easily be extended to learn from negative observations as well. Learning from positive data is often considerably more difficult than learning from both positive and negative data, and we show that promising generalization is possible even in this less richly observed setting.

Model inference

Given this generative process, we can compute the probability that a query q—a logical atom, such as female(mary) or spouse(mary, jon)—is true given a set of contexts, a theory, and a setting of the hyperparameters α=(α, β). Summing over models of the theory (and restricting to $K = 1$ for clarity), we see that:

$$P(q \mid Obs, Obj, T, \alpha) = \sum_{C,M} P(q \mid M) P(M \mid C) P(C \mid Obs, Obj, T).$$

The laws of T uniquely determine M given C, so the sum over M and the term $P(M \mid C)$ can be dropped. The remaining sum can be expressed via Bayes' rule as follows:

$$P(q \mid Obs, Obj, T) \propto \sum_{C} P(q \mid C) P(Obs \mid C, T) P(C \mid Obj, T). \quad (1)$$

When there are few objects, the sum over the extensions of the core relations can be computed exactly. In practice, we must often approximate the sum using Gibbs sampling or other Monte Carlo methods; these methods can also be used to search for single most probable model.

Inference in sparse contexts

Once a learner acquires a systematic theory, such as kinship or taxonomy (as shown in Table 1), a number of inferences about novel, sparsely observed objects can be made. When we place probabilities over these logical representations, three general classes of inferences are possible: *deductive*, *inductive*, and *deductive consequences of inductive inferences*. We show examples of all three for the theories of taxonomy and kinship.

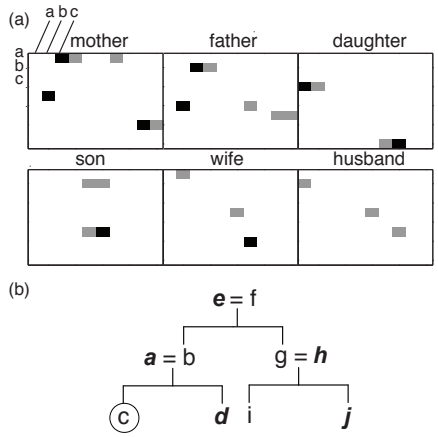

Figure 3: (a) $(a \ldots j \times a \ldots j)$ matrices of observable relations in the kinship domain. Black and grey entries indicate observed and inferred relations, respectively. White entries indicate relations inferred to be *false*. (b) Inferred family tree, compactly representing all inferences in (a). Females shown in bold italics, males in ordinary font (c's gender is unknown, indicated by a circle.)

Kinship Consider the family relations shown in Figure 3. In light of the logical theory of kinship, many inferences could be made. For example, the gender of the family members can be deduced with certainty, since mothers are always female, sons are always males, and so forth.

In addition to deductive inferences, the observations invite several plausible inductive inferences. Since a and b are both observed to be parents of c, it is plausible that a is the spouse of b. In light of this inductive inference, one can infer—by application the logical laws of kinship—that d is the daughter of b, since d was already observed to be the daughter of a. This is an example of a deductive consequence of an inductive inference. Note how our initial inductive leap, that a and b are married, led to a deductive inference that allowed us to make efficient use of our observations. The best model found for this context via greedy stochastic search contains all of these inferences, deductive and inductive. The core relations for this best-scoring model correspond to the family tree shown in Figure 3(b); the observable relations that it predicts will be true (or false) are indicated by the squares colored gray (or white, respectively) in Figure 3(a).

Essential to these inferences is the fact that the identity of objects remains the same even if they play different 'roles' in distinct scenarios. In our observations, we see two roles for a: one in which it is the mother of c (as first argument to the mother predicate), and another in which it is the parent of d (as the the second argument to daughter.) The effective integration of information from both the deductive and inductive inferences just shown relies heavily on the identity of a in these two distinct roles.

Taxonomy Consider the full taxonomy given in Figure 1(b). Suppose that we observe only the is_a links corresponding to direct edges in the hierarchy, along with the properties true of the leaf-node categories. For instance, we observe that canaries (a leaf-node category) can sing, are yellow, have wings, and have skin, that eagles also have wings and skin, and that canaries and eagles are both birds and are animals (along with many other facts). We make no direct observations about the properties of birds, fish or animals, though. If we then search for the best-scoring model, we recover the configuration of core relations shown in Figure 1(b): the extension of $\mathsf{is\_a}_{\mathsf{CF}}$ and $\mathsf{has\_a}_{\mathsf{C}}$ includes only the minimal set of is_a and has_a links needed to capture all observations under the theory's laws. Each property is attached to only one category in the is_a hierarchy, the lowest superordinate of all categories with that property. We compress out the correlations in the observed properties of leaf-node categories, by positing properties true of abstract superordinates which are not themselves ever directly observed.

Now suppose that we learn of two new objects, a and b, by making the following minimal observations: $\mathsf{is\_a}(\mathsf{a}, \mathsf{fish})$, $\mathsf{is\_a}(\mathsf{b}, \mathsf{animal})$. The classic CQ approach would infer many common-sense inferences about a, b, e.g. that both breathe, that a can swim, and so on. Our best-scoring model does as well. Conditioned on the inferences described in the previous paragraph, that familiar properties like swimming and breathing are probably true of the unobserved superordinate categories fish and animal, we now infer that these properties also hold for the new species a and b, for which we have observed only a single is_a relation each. This example shows how we capture a general feature of common-sense reasoning by combining the power of induction and deduction: an inductive leap to the likely properties of unobserved superordinate categories, with deductive inference of the consequences that follow.

Property induction We now show how intuitive patterns of graded, uncertain inference, usually thought to weigh against symbolic representations of human semantic knowledge, can also be captured in our probabilistic logical framework. We consider a simple case of *property induction*. For tractability and ease of presentation, we explore these phenomena using a pared-down version of the Collins & Quillian example, shown in Figure 4(a). We call this structure a *balanced taxonomy*, where both properties and nodes are distributed evenly along all branches of the tree. In this case, each node has a single unique property, in addition to properties it inherits from its parent nodes. For instance, a's unique property is p3, while f's unique property is p5, and both inherit p1 from their parent node g.

Given observations of direct is_a links and properties of leaf node category, we can then query for the probability that each property is true of each category in the hierarchy. Figure 4(c) shows the probabilities of all queries for the property p3, which is observed only at one leaf node category, a. As expected, the probability that a has p3 is near certain. However, other less certain generalizations are found. It is plausible but not likely that e has p3, and very unlikely that g or f

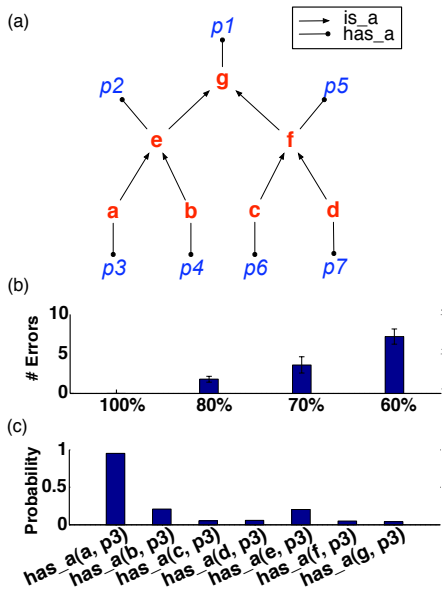

Figure 4: (a) Balanced taxonomy. (b) Simulations of CQ contexts with different sparsity levels. (c) The probabilities of various queries under full observations of leaf node properties. Results obtained by Gibbs sampling core extensions of the theory (using 6000 samples.)

has it. This is explained by the $\frac{1}{size}$ factor in our likelihood. If g had p3, all nodes in the hierarchy would inherit p3, making the fact of our first and only observation of the property at one specific node a relatively less likely. The same size principle weighs against generalization to e, but less so, since e is not so far up in the taxonomy. We also see a pattern of similarity-based generalization along the leaf node categories of the is_a hierarchy. The closer a category is in the tree to a, the more likely it is to have p3.

Systematic study of generalization

We now study more systematically how well our approach generalizes beyond the observed data, as a function of the sparsity of the data it receives. Inspired by the simulations of Hinton (1986) for the theory of kinship, we ran simulations in which we observed randomly sampled sets of the observable facts in the balanced taxonomy domain and searched for the best scoring model under the taxonomic theory. In this domain there are a total of 98 observable propositions ($7 \times 7 = 49$ is_a propositions, and the same number of has_a propositions), of which 27 are true and 71 are false. For instance, $\mathsf{has\_a}(\mathsf{a}, \mathsf{p2})$ is true, while $\mathsf{has\_a}(\mathsf{b}, \mathsf{p5})$ is false. In keeping with our focus on learning from positive examples, we observed a fraction of the true observable propositions—100%, 80%, 70%, or 60% of the 27 possible positive examples—and searched for the best scoring model under the taxonomic theory. Rather than merely memorizing the given true facts, we seek to compress the data via inferring the underlying taxonomy and then predict the unobserved propositions this model entails.

We evaluated performance by computing the number of incorrect observable propositions entailed by the recovered model. This number includes two types of errors: "false alarms" (false propositions predicted to be true) and "misses" (true propositions predicted to be false). Figure 4(b) shows the results, averaged across five trials at every level of sparsity. Generalization is perfect when all 27 true propositions are observed. This means that none of the 71 false propositions were incorrectly predicted to be true. Generalization decreases gradually as the data become sparser, with errors distributed among both misses and false alarms. Even at the highest levels of sparsity, generalization is quite good: we observe only 16 ($\approx 60\%$ of 27) of the 98 observable propositions and we make about 7 errors on average, inferring the correct truth values for (on average) 91/98 observables.

Learning a logical theory

We now address the problem of learning the theories we've described. We do so by defining a prior distribution $P(T)$ over theories. Following Kemp et. al., we take a *representation length* (RL) approach . Intuitively, given the choice between two theories, a RL prior will favor the one that is less complicated to write. The precise definition of RL is tied to a choice of language, which in our case is the language of Horn clauses.

As before, the distribution is described as a generative process. Given an assignment of values to the hyperparameters α, λ, theories are generated as follows:

1. Generate a number γ of core relations in T: $\gamma \sim \text{Poisson}(\lambda)$.
   (a) For every generated core relation R, choose its arity, where $P(R \text{ is binary}) \sim \text{Bern}(\theta_a)$ and whether it is functional, $P(R \text{ is functional}) \sim \text{Bern}(\theta_f)$. We assume θ_a, θ_f are given.
2. Draw a set of laws, scored according to their RL. The RL in our case is a count of the number of total predicates, variables and clauses that appear in the laws:
$$P(Laws) \propto 2^{-rl(Laws)}$$
where $rl(Laws) = \#_{cp} + \#_{vars} + \#_{clauses}$. We assume every $L \in Laws$ is syntactically well-formed (otherwise, $P(Laws) = 0$.)

We now consider a case of competing theories, by evaluating the true taxonomy theory discussed earlier against seven variants, shown in Table 2. These were generated by considering all theories that can be constructed by inclusion/omission of the following features: (1) the is_a transitivity law (L1), (2) the property has_a inheritance law (L2), and (3) having the is_a relation be a function. L1 and L2 are the third and fourth laws in Table 1(b), respectively. To see which theory is favored on a given data set, we first

	T1	T2	T3	T4	T5	T6	T7	T8
L1	√	√	×	×	√	√	×	×
L2	√	√	√	√	×	×	×	×
is_a func.	√	×	√	×	√	×	√	×
log score	**−150**	−180	−163	−185	−171	−200	−164	−206

Table 2: Log scores for true taxonomy theory (shown in bold) and its variants.

look for the best scoring model of each theory using greedy stochastic search. We then score each theory together with its best model according to the joint posterior probability $P(T, M, C \mid Obs, Obj, \lambda, \alpha)$. Table 2 shows the log scores of the eight theories, taking as data all true observable propositions in the balanced taxonomy domain.

Note that all the variants of the true theory are favored by our prior, as they are less complex. We can also see that the score ordering for these theories is sometimes structured and monotonically decreasing as we go from more to less complex theories. While the prior favors simplicity, the posterior ought to favor the more complex of the variants we consider, since these have greater predictive power. For example, T2 is penalized for lacking is_a as a function, and T3 for lacking L1, but T4 is penalized more heavily than either for lacking both.

Similarly, theories having is_a as a function are preferred to those that do not. A learner who believes is_a is only a relation and not a function might believe that a shark is both a fish and a bird, leading to a penalty in the likelihood of our model. Fixing is_a as a function rules out these cases and so gives a more compact encoding of the observations.

This monotonicity property does not hold of all theories (compare T3, T5, and T7.) This can be explained by the fact that L1 and L2 are not independent. A learner who does not know about the inheritance of properties (T7) would gain little by adopting L1, in the current data set, and vice versa, depending on the statistics of the observations.

Our purpose was to demonstrate that our generative model can be used as a evaluation metric for theories. We leave open the orthogonal question of how actual learners select hypotheses to evaluate from this vast space. Though we used a stochastic search for hypothesis selection here, more work is needed to see if any search-based account could provide a solid foundation for theory learning. For any search-based proposal to work, the probability landscape of theories must have tractable properties, such as the monotonicity property we considered above. An area of future work is determining when these conditions hold for theories in general.

Conclusions and future work

In a classic paper on distributed representations, Hinton outlined three questions to ask of all systems of knowledge representation: (1) Does the representation create "sensible internal representations" for the entities it processes, in a way that is sensitive to the (potentially latent) regularities in its input? (2) Does it generalize to unobserved truths of a domain? And finally, (3) Does it compress and generalize information from distinct but isomorphic context?

We believe that our proposed framework fares well on all three. First, the use of a logical framework makes our representation transparent, allowing for direct and interpretable comparisons of the internal structure of two learned representations. It is not obvious how to make such comparisons in a connectionist setting. At the same time, Bayesian inference over these structured representations allows us to make inductive leaps from sparse and noisy data, overcoming a major flaw of traditional symbolic approaches. An advantage of our framework is that the engine of deduction (logical representations), and the engine of induction (Bayesian inference), cooperate and supplement each other to reason in sparse contexts. We have shown the benefits of this approach for the theories of kinship and taxonomy.

In future work, we would like to develop more scalable inference algorithms and a more expressive language, allowing unobserved entities and probabilistic generative mechanisms. This should let us explore more realistic theories, such as the dynamic theory of Mendelian genetics. A second direction is to study how well our approach captures people's reasoning in the laboratory, following (Kemp, Goodman, & Tenenbaum, 2008).

Acknowledgements We thank Vikash Mansinghka, Brian Milch, and Daniel Roy for helpful discussions. This work was funded in part by AFOSR grant FA9550-07-1-0075 and the James S. McDonnell Foundation Causal Learning Research Collaborative.

References

Collins, A., & Quillian, M. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*.

Conklin, D., & Witten, I. H. (1994). Complexity-based induction. *Machine Learning*, *16*(3), 203-225.

Fodor, J., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture. a critical analysis. *Cognition*, 3-71.

Hinton, G. E. (1986). Learning distributed representations of concepts. In *Proceedings of the Eights Annual Meeting of the Cognitive Science Society.*

Kemp, C., Goodman, N. D., & Tenenbaum, J. B. (2007). Learning and using relational theories. In *Advances in Neural Information Processing Systems 20.*

Kemp, C., Goodman, N. D., & Tenenbaum, J. B. (2008). Theory acquisition and the language of thought. In *Proceedings of Thirtieth Annual Meeting of the Cognitive Science Society.*

Muggleton, S., & De Raedt, L. (1994). Inductive logic programming: Theory and methods. *Journal of Logic Programming*, *19/20*, 629-679.

Rogers, T., & McClelland, J. (2004). *Semantic cognition: A parallel distributed processing approach*. MIT Press.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*, *10*, 309–318.