Modeling the Acquisition of Domain Structure and Feature Understanding

Amy Perfors, Charles Kemp, Joshua B. Tenenbaum
Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology
Address correspondence to **perfors@mit.edu**

Abstract

Young children face a difficult learning task: without having access to the cues that full language mastery would provide, they must acquire conceptual knowledge about the world; likewise, they must somehow learn words while lacking the full conceptual structure that the words refer to. We present a Bayesian framework that can model aspects of the acquisition of theory knowledge as a function of different types of input. We describe a set of developmental phenomena that our model can address.

Introduction

Acquiring theories of the world, particularly as they vary from domain to domain, is one of the most difficult tasks confronting young children. This is especially the case for infants and children in the earliest stages of language learning, since they are faced with two simultaneous problems: learning words without yet fully grasping the conceptual theory underlying the domain to which those words refer, and learning the knowledge underlying a domain without yet fully understanding the language that would give cues to that knowledge.

What constitutes theory knowledge? The question is the focus of some debate, but for this paper we wish to limit our focus to two aspects that are believed to be important facets of theories. The first is structure: knowledge about the organization of relationships between objects in a domain. For instance, biological objects are organized hierarchically with respect to each other, but political views may be better characterized by a linear spectrum [2, 12]. The other is knowledge about feature weights: which features are the most important or central in that domain. For example, the generalization of novel nouns for animals is best done on the basis of shape or internal features, while for foods it depends more on color [10, 13, 21].

Of course, feature biases and domain structure are not independent of each other. Important features are those that are the most coherent and informative in that domain; when that domain is structured, this means that important features are those that are most coherent with respect to the structure. For instance, animals organized in a taxonomic structure would be expected to share important features (e.g., warm-bloodedness) only with those animals close to them on the structure (i.e., other mammals); unimportant features, such as skin color, could be shared – or not – with objects arbitrarily far away in the structure.

The interdependence between structures and features means that learning about structures and learning about features goes hand-in-hand: initially knowing that some features are more important than others can constrain the best structures to ones on which those features to covary coherently. Similarly, initially knowing that some structure is most appropriate limits the important features to those than cluster tightly on that structure.

Though the empirical phenomena underlying the acquisition of knowledge about domain structure and feature biases are well studied (e.g., [10, 15, 21, 23]), there are few computational models for them. The main models involve connectionist networks (e.g., [4, 19]), which have proven quite useful to cognitive psychologists in many ways, but which suffer from the disadvantage of opacity. It is often difficult to understand exactly what the model is doing and what its representation actually consists of. Particularly if one is interested in the representation and acquisition of *structure* – such as the tree-structured taxonomy in folk biology – an unstructured connectionist approach may not only be less transparent, but may also perform worse on tasks that specifically rely on that structure [12].

The Bayesian model of concept acquisition we present here allows for a tractable and clear exploration of the interplay between the acquisition of information about the world and the development of domain theories. Because we operate in a framework for Bayesian inference, we incorporate one of the primary advantages of connectionist algorithms – their domain-general statistical learning character – while still providing the clarity and explicitness of a more structured approach.

In the following sections we first describe a set of phenomena that arise while children acquire domain-specific theory knowledge; we then explore how our approach can naturally model them.

Learning about structure Little is known about at what age and to what extent children realize that some domains are organized taxonomically. Some argue that the tendency to group living kinds into hierarchies reflects an "innately determined cognitive structure" [2]. However, children's behavior changes in ways that suggest they may not initially begin with an accessible taxonomic structure. Very young children have a hard time learning superordinate categories [9], especially if they already have an appropriate basic-level term in their vocabulary [14]. There is also a clear developmental trend in the ability to use superordinate or subordinate categories, either in induction [8], classification [23], or as the possible target in a word-learning task [15]. In sum, though the data are unclear about whether very young children *have* taxonomic categories, there is strong evidence that at least they have a difficult time *using* taxonomic information until they are older.

At that point, children can apply their knowledge about taxonomic structure to new learning situations such as word learning and object classification. 4-year-old children given multiple examples of a novel word are capable of generalizing its meaning based on its location in a taxonomic structure [22]. Also, learning superordinate labels appears to enable children to improve their performance on classification tasks in taxonomic domains [23].

Structural knowledge can also be helpful in property induction, since explicit understanding of taxonomic structure can guide the ability to correctly reason about novel features [12]. For instance, knowing that whales are more closely related to elephants than they are to fish in a taxonomic structure can help one realize that whales might be warm-blooded like elephants rather than cold-blooded like fish. Children readily make inductions about the internal features of novel objects in taxonomic domains based on knowledge of domain structure, and older children make inferences at the superordinate level more often than younger children [8].

In sum, there is evidence that children learn to represent or use the correct underlying structure in taxonomic domains over the first five years of life. They are capable of using this structure knowledge to correctly classify novel words and objects as well as to correctly generalize about hidden properties of objects. Our model gives one account for how children's developmental shift towards the robust use of structural information can arise, both in a strongly taxonomic domain (animals) as well as one with a less clear hierarchical structure (foods). The model also demonstrates how more accurate generalization of hidden features in older children may be a result of their more accurate underlying structural knowledge about that domain.

Learning about features Children often learn domain-specific feature biases even before realizing what the correct structure for that domain is. By the age of two, they correctly generalize novel count nouns on the basis of whether the original referent is a solid or a non-solid [21]. If the referent is a solid, children show a distinct shape bias: a tendency to generalize the new word on the basis of shape rather than size, color, or texture [10, 21]. When the original referent is a non-solid, children have a "material bias" in which they generalize the novel name to objects sharing the same material but having a different shape [21]; however, this bias does not appear strongly until after 30 months [20].

Though the shape bias emerges earliest and applies to many domains, children are capable of learning different biases for different domains. They eventually realize that functionality is an important feature for artifacts but not biological kinds [6] and that color is more important for nonliving natural kinds [11] and food [13].

Just as knowing about structure can help children learn and classify new objects in a domain, learning which features are important can help them generalize that knowledge to new items as well. Recent work by Samuelson & Smith suggests that the acquisition of feature biases can facilitate the learning of new object names [20]. Their explanation is that children are able to use the words they learn to infer that categories are organized by similarity in shape. When presented with a novel word-object label, they are therefore more apt to generalize the word based on its shape rather than on other properties.

In sum, there is evidence that children learn domain-specific knowledge of feature biases during the first few years of life. They are then capable of using this feature knowledge to learn novel words and thus classify novel objects in that domain. Our model illustrates the emergence of the shape bias in a domain where it applies (animals) as well as its lack of emergence in a domain where other features such as color may be equally or more important (foods). Additionally, when objects covary coherently according to some features, our model is capable of recognizing that covariance and using it to correctly classify new objects.

The Bayesian model

There are two aspects to our model: learning domain structure and learning feature weights.

The structure-learning component, described more fully in Kemp, Perfors, & Tenenbaum [12], defines a structure as a graph over objects, such that objects with many common features are closer together in the graph. For instance, a one-dimensional chain is a good structure for the American political domain because objects (politicians) on one extreme of the chain tend to share many features with politicians at their end rather than politicians on the other extreme. We assume that learners come to a new domain equipped with the capacity to represent a range of qualitatively distinct kinds of structures that could describe that domain, including taxonomies (trees), dimensions (chains), and clusters. (A cluster is a grouping of objects such that all the objects within it are expected to share many features, but there is no higher-order relationship between or within clusters). Though people can surely conceptualize more complex structure classes, we restrict ourselves to these three here, as a representative set of simple hypotheses that small children could reasonably learn about in early development.

Our model assumes that learners are presented with data in the form of binary-valued object-feature matrices. We can use Bayesian model selection to evaluate which structure class was most likely to have generated this data. More formally, if D is an object-feature matrix, the posterior probability of any structure class C is proportional to its likelihood under the data $p(D|C)$ times its prior probability $p(C)$. Since we assign equal prior probability to all classes, the best class is the one that makes the data most likely.

Likelihood for trees and chains is calculated based on the intuition that objects that are "close to" each other on a structure will be more likely to share features than objects further away. This is captured formally by assuming that features are generated over structures using a symmetric mutation process, under which the probability of a feature switching values between the beginning and end of any branch b is a function of the length of b and the mutation rate, λ. We assume that features are conditionally independent given the structure, and can then compute the likelihood $p(D|\mathcal{S}, C)$, the probability of the data given structure $\mathcal{S}$ and structure class C, for each feature vector taken individually. The likelihood of any *specific* structure can be calculated by multiplying probabilities for each feature taken individually on that structure; the likelihood of a structure *class* by integrating[1] over the space of all structures, as below:

$$p(D|C) = \int p(D|\mathcal{S}, C)p(\mathcal{S}|C)d\mathcal{S} \qquad (1)$$

Intuitively, this means that a structure class C provides a good account of object-feature data D if the data are highly probable under a range of structures $\mathcal{S}$ in class C, and if these structures themselves have high prior probability within C. For trees and chains, prior probability $p(\mathcal{S}|C)$ is spread uniformly over the space of all possible trees or chains. For clusters, we use a prior over possible cluster partitions that is derived from the Chinese Restaurant Process [3]. This prior admits any number of clusters but favors fewer clusters, while the likelihood favors more clusters in order to fit the data better. The likelihood $p(D|\mathcal{S}, C)$ is defined by a weighted coin flipping process, with distinct weights for each cluster and for each feature in the data. The balance between priors and likelihoods instantiates a Bayesian version of Occam's razor that finds a cluster model with an appropriate number of clusters. This clustering model is related to Anderson's rational model of categorization [1], and a more mathematical treatment can be found in [16].

For all structure classes, we can intuitively understand the "most important" features as the individual features with the highest likelihood. Feature likelihood captures the intuition that the highly weighted features are those that best fit a given structure. For instance, "warm blooded", an important feature, is tightly clumped on an animal taxonomy (only mammals and birds are warm blooded). The feature "is black", on the other hand, is not important and does not fit well (animals of all types may be black, regardless of where they are in the taxonomy). In our model these features would have lower weights because they have low likelihood given that particular taxonomy. This intuition applies equally well to all structure classes: for instance, an important feature for a given cluster would be one that obeys cluster boundaries and thus has a higher likelihood.

[1]In practice, we approximate the integral using Markov Chain Monte Carlo (MCMC) techniques

The Datasets

We used datasets in two domains, animals and foods. We chose these domains because they are of great interest to very young children and contain words that are among the first few learned [7]. Moreover, the domains may differ in terms of which structure best describes them [18] as well as which features are highly weighted [13].

Each dataset is a binary-valued object-feature matrix. The choice of features and objects was inspired by the features and objects in the data collected by Cree & McRae [5]. A subject blind to the hypothesis of the experiment classified the features as *shape*, *surface* (colors and textures), *behavior* (for animals), *smell & taste* (for foods), *is-a* features corresponding to superordinate words (e.g. "is a reptile", "is a vegetable"), and *internal* features corresponding to internal properties (e.g., "has lungs", "has red blood"). In both domains, all *is-a* words were structurally appropriate for a hierarchy. None of the internal features applied to insects or cephalopods, but all other animals had at least two.[2]

The animal dataset contained 60 animals with 112 features, and the foods dataset contained 56 foods with 64 features. Animals consisted of mammals, reptiles, birds, amphibians, and insects; foods consisted primarily of fruits and vegetables (49 objects), but included desserts and other staples (bread, rice, etc) as well (7 objects). Not all objects had the same number of features, but the variance between objects was small and no objects had fewer than eight (animal mean: 27 features, sd: 6.86; food mean: 17.1 features, sd: 4.06).

One way to model development is to alter the nature and number of features the model works with. Early in development, children may have access to perceptual, obvious features, but not to conceptual or hidden ones. A subject blind to the hypothesis of the experiment classified each of the features on a scale measuring its degree of perceptual obviousness (where "is red" is very obvious, but "has Vitamin C" is very nonobvious). When we wished to model differing numbers of perceptual features, we presented the model input composed of the most perceptually obvious features.[3]

Another way to model development is to alter the number of objects the model works with. This was done by ranking each object in the order of the age of the first production of the word corresponding to the object, according to the 50% norms found in Fenson et al. [7].[4] Adding objects in the order of word acquisition might be seen as a way to model word learning or as a reflection of the amount of exposure each child has had to the different objects in the world. Either of these interpreta-

[2]A copy of the dataset may be found at `www.mit.edu/~perfors/cogsci05.html`.

[3]There were enough animal features to make two perceptual datasets. One had the 48 most perceptually obvious features, the other had the 61 most obvious.

[4]A few objects were included if children spoke a similar word early and that word did not correspond to an object in the dataset: e.g. *rat* for *mouse*. Additionally, children learn the words *bird* and *fish* quite early: these were approximated by including the most prototypical examples of each category in the dataset, *robin* for bird and *trout* for fish.

	Objects	# Feat	Percept.	L_{tree}	L_{chain}	L_{clust}
Animals	20	48	48	-460	**-418**	-470
	20	67	48	-607	**-530**	-614
	40	48	48	-859	**-849**	-903
	40	67	48	-1092	**-967**	-1098
	40	61	61	-1108	**-1076**	-1166
	60	61	61	**-1510**	-1522	-1639
	60	80	61	**-1842**	-1929	-2013
	60	80	61	**-2106**	-2157	-2255
	60	112	61	**-2854**	-2994	-3051
Foods	20	37	37	-417	**-405**	-415
	20	44	37	-477	**-443**	-482
	56	64	37	-2087	**-1607**	-1680

Table 1: Log likelihoods on each structure class as a function of type of input. Higher likelihoods are indicated in bold.

tions is consistent with our results, in which we compare datasets composed of 20, 40, and 60 objects. Datasets with fewer objects tend to have a higher proportion of mammals than later datasets do.

Results

We now explore how our model accounts for the phenomena described in the introduction.

Learning about structure

Our model gives one account of how children's developmental shift towards the robust use of structural information can arise, both in a strongly taxonomic domain (animals) as well as one with a less clear hierarchical structure (foods). Table 1 shows the log likelihood values for each structure class considered by our model, as a function of its input. Because log likelihoods are negative, higher likelihood model classes (highlighted in bold) have a smaller absolute value. Because they are log probabilities, differences of the magnitude shown in the table are substantial. Trends in both the foods and the animals domain can be identified.

The trend in the animal domain demonstrates an interesting progression from simplicity to complexity. When there are fewer objects and fewer features, the simpler chain structure class has a higher likelihood than the more complicated tree structure class (though both had a higher likelihood for than clusters). The change from chain to tree is primarily driven by the increased number of objects in the dataset: all of the datasets with 40 objects are best fit by a chain, while all of the datasets with 60 objects are best fit by a tree.

What are the developmental implications of this change in structure? Because a chain is one-dimensional, it cannot represent superordinate structure. This parallels evidence suggesting that younger children do not seem to use the superordinate level in tasks like induction, classification, or word learning. They do so only when they reach the age of 4 or 5 years, suggesting that before then they might not believe that a tree-structure representation is appropriate. [8, 15, 23]. Why did our model find that chains were more appropriate than clusters? This is probably because clusters collapse all within-cluster information. Since reasonable animal clusters are highly heterogeneous (e.g., mammals vary widely in size, shape, and behavior), a cluster structure would lose too much information compared to a chain.

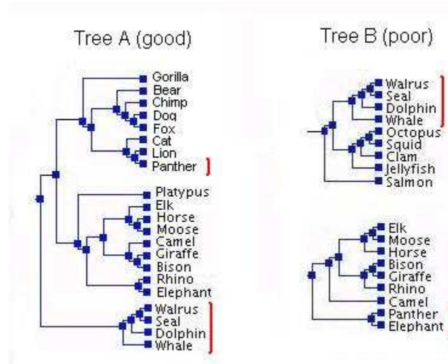

Figure 1: Partial consensus trees from two datasets, each with 60 objects and 61 features. Tree A is built from a dataset containing *internal* and *is-a* as well as perceptual features, and is more structurally accurate; Tree B is built using only perceptual features.

The food domain is an interesting contrast to animals. For all food datasets, no matter how many objects or features the algorithm was working with, the chain class fit the data better than either the tree or cluster model class. This may be because our foods dataset was comprised primarily of fruits and vegetables. This is in line with a recent study that found that subcategories are much less well-differentiated in the domain of fruits than in the animal domain [18]. Indeed, the best chains follow a sensible path from green leafy vegetables on one end, through melons, citruses, and berries, then finally going through legumes and roots before arriving at all the non-plant foods at the other end.

One area in which the interdependence of features and structure becomes most apparent is in comparing structures made using datasets with different features. *Is-a* features like "is a mammal" and *internal* features like "has warm blood" are often learned by children around the same time they begin to realize that the animal domain is organized taxonomically. Similarly, we find that the taxonomies found by the model appear more accurate when made using datasets that incorporate these features compared to datasets incorporating purely perceptual features. Figure 1 compares portions of two 60-object trees found by our model, each made with 61 features. Tree A contains 42 perceptual, nine *is-a*, and ten *internal* features. By contrast, Tree B contained no *is-a* or *internal* features, just the 61 basic perceptual features.

Though both trees are adequate, Tree B has some structural flaws that are not apparent in Tree A. For instance, Tree B incorrectly locates the aquatic mammals with the sea creatures and places the panther far from the other felines. Tree A, which incorporates the information that aquatic mammals are mammals and that panthers are felines, does not have these errors. This supports the intuition that structures incorporating "more important" features will appear to be more accurate than structures that do not.

If structures and features are interdependent, improvement in one should lead to improvement in the other. In this case, does the more accurate structural knowledge in Tree A lead to improved generalization of hidden or unknown features? We can test this by observing the performance of Tree A and Tree B on a property induction task. We compare the inductive predictions using Tree A and Tree B to the human argument ratings collected by Osherson et al [17]. Osherson used a ten-animal domain consisting only of mammals: horse, cow, chimp, gorilla, mouse, squirrel, dolphin, seal, elephant, and rhino.[5] The specific set contains 36 two-example arguments, and the conclusion species is always "horse." The general set contains 45 three-example arguments, and the conclusion category is "all mammals." Unfamiliar (blank) predicates were used for all these arguments. The tree-based Bayesian model rates the strength of general arguments by computing the probability that all ten animals in the domain have the property.

The predictions of the model using Tree B were noticeably more poorly correlated with ratings of human argument strength than were the predictions using Tree A (specific: $r = 0.833$ (Tree A), $r = 0.739$ (Tree B); general: $r = 0.832$ (Tree A), $r = 0.566$ (Tree B)). This poor performance is probably a result of Tree B's less accurate structure. In a similar way, older children's more accurate taxonomic knowledge may underlie their increasingly accurate generalization of hidden features.

Learning about features

We have shown that our model is capable of learning appropriate structural information and applying that information to make inferences about features. But can it learn about features directly by giving more weight to more important features in a domain? As Figure 2 shows, our model correctly realizes that shape features should be given more weight in the animal but not the food domain, and that this realization is a function of how many objects the model has "seen." For all animal datasets with only 20 objects, there was little difference in the feature likelihoods[6] of shape, surface, or behavior features. For all animal datasets with 60 objects, shape features are consistently significantly different than surface (color & texture) features.

Unlike in the animal domain, for the foods there is no significant difference in the likelihoods of any of the features at any stage. Thus we do not see the emergence of a bias towards the shape features; however, there is also no bias toward other types of features, including the surface features like color. This contrasts with the finding that young children consider color features important in the food domain [13]; however, since our dataset included primarily fruits and vegetables rather than a representative sampling of the foods 2-year-olds were likely to be exposed to, it may not be strictly comparable. Indeed, the finding that shape features were likely to have

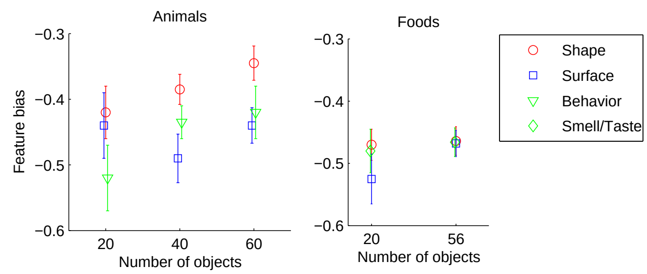

Figure 2: Relative (normalized) log likelihood of each type of feature as the number of objects in the dataset increases. In the food domain, there is no bias towards one type of feature no matter how many objects were in the dataset: all likelihoods are similar. In the animal domain, a shape bias emerges for larger datasets.

significantly higher likelihoods than surface features for animals but not foods suggests that a shape bias will not just emerge automatically given enough features.[7] Rather, shape features for animals simply covary more coherently with the correct structure, and thus the model tends to give them a higher likelihood than features that are less coherent.

Recent work by Samuelson and Smith [20] suggests that acquiring a shape bias can facilitate the learning of more object names. Is this evident in our model? When objects covary coherently according to some features, is our model capable of recognizing that coherence and using it to correctly classify new objects?

We can answer this by adapting a test done by Rogers & McClelland [19]. They trained a connectionist network on 21 biological objects (birds, fish, mammals, and plants) and 57 features. They then presented it with four new test items and four new features. The four new features included two features they called "size" (*large* and *small*) and two dubbed "brightness" (*dull* and *bright*). Size but not brightness mattered for discriminating between trees and flowers, while brightness but not size mattered for discriminating fish and birds (e.g., all birds were bright and all fish dull, but brightness features varied randomly for the plants). Test objects were given values on these features such that items O1 & O3 and O2 & O4 were of the same brightness, but O1 & O2 and O3 & O4 were of the same size. In one run, the test items were also given a feature belonging to plants ("roots"); in another, they were instead given one belonging to animals ("skin"). If the model is able to infer domain-specific feature biases – size for plants and brightness for animals – then it should classify the test items according to their size when they have plant features, but by their brightness when they have animal features.

We presented our model with the same dataset and found that test items were classified appropriately. Figure 3 shows the output tree when test items had the "skin" feature. Items 3 and 4, which are the same bright-

[5]We replaced *cow* with *bison* and *mouse* with *rat* since these two words were not in our dataset; this is conservative because if anything, this would make performance on the model decrease more for the more "correct" trees, rather than the reverse.

[6]All feature likelihoods are calculated with respect to the best structure for that dataset, but the results are qualitatively the same no matter which structure is used.

[7]It was not that case that shape features have an average higher likelihood simply because there are more of them; for instance, in the dataset showing a significant difference between shape and both surface and behavior features, there were 23 shape, 15 surface, and 43 behavior features

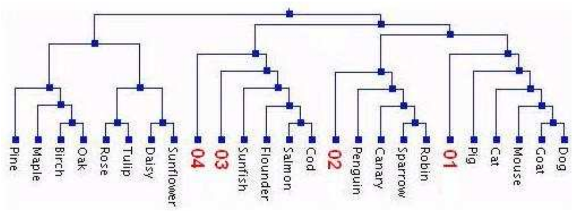

Figure 3: Performance on the object classification task modeled by that in Rogers and McClelland [19]. Test items O3 & O4 and O1 & O2 should be classified together, since they share the feature that is important for animals.

ness (an important feature for animals) are classified together; items 1 and 3, which share size, are not. The run using the "root" feature shows similar results; for space reasons, we did not include it.

This suggests that learning a bias for coherently varying features can actually assist with the generalization of novel items. A learning algorithm that can correctly place novel items in the existing domain structure will be able to learn more of these items than one that cannot. In our model, learning a bias towards certain features – that is, learning that some features have a higher likelihood on the correct structure in that domain – can result in this improved generalization.

Conclusions

The Bayesian model presented here provides an explicit and tractable paradigm in which to explore the interaction of word learning and concept acquisition. We explored developmental phenomena in both feature and structure learning and showed that our model could qualitatively capture the stages of learning of both. Our model can also demonstrate how this learned knowledge might be useful for accurate word/object classification and property induction. Our intent was not to demonstrate that it fully captures all aspects of these phenomena, but rather to give a "proof of concept" – a demonstration that our model can be a useful tool for cognitive scientists seeking to understand the interaction between features and structure in conceptual development, and the role that different types of input may play. The model can qualitatively and quantitatively explain a range of interesting phenomena: the emergence of domain-specific feature biases, the ability to use these biases to correctly classify new objects, the realization that some domains are hierarchically organized, and the ability to use this structure knowledge to improve induction of novel properties. We are optimistic that this modeling approach has the flexibility and transparency to be an important tool for developmental psychologists and cognitive scientists alike.

Acknowledgements AFP was supported by the NDSEG graduate fellowship, and JBT by the Paul E. Newton Career Development Chair.

References

[1] Anderson, J. (1991). The adaptive nature of human categorization. *Psych.Review*, 98(3):409–429.

[2] Atran, S. (1998). Folkbiology and the anthropology of science: Cognitive universals and cultural particulars. *BBS*, 21:547–609.

[3] Blei, D., Griffiths, T., Jordan, M., and Tenenbaum, J. B. (2003). Hierarchical topic models and the nested chinese restaurant process. In *NIPS 16*.

[4] Colunga, E. and Smith, L. (2004). Dumb mechanisms make smart concepts. In *Proceedings of the 26th annual conference of the Cognitive Science Society*.

[5] Cree, G. and McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *JExP: General*, 132:163–201.

[6] Diesendruck, G. (2003). Mapping the sim. space of children & adults' artifact categories. *Cog.Dev.*, 18:217–231.

[7] Fenson, L., Dale, P., Reznick, J., Bates, E., Thal, D., and Pethick, S. (1994). *Variability in Early Communicative Devleopment*, volume 49. Monographs of the Society for Research in Child Development.

[8] Gelman, S. and O'Reilly, A. (1988). Children's inductive inferences within superordinate categories: the role of language and category structure. *Ch.Dev.*, 59:876–887.

[9] Horton, M. and Markman, E. (1980). Developmental differences in the acquisition of basic and superordinate categories. *Ch.Dev.*, 51:708–719.

[10] Imai, M., Gentner, D., and Uchida, N. (1994). Children's theories of word meaning: the role of shape similarity in early acquisition. *Cog.Dev.*, 9:45–75.

[11] Keil, F. (2003). Explanation, association, and the acquisition of word meaning. *Lingua*, 92:169–196.

[12] Kemp, C., Perfors, A., and Tenenbaum, J. B. (2003). Learning domain structures. In *Proceedings of the 26th annual conference of the Cognitive Science Society*.

[13] Macario, J. F. (1991). Young children's use of color in classification: Foods as canonically colored objects. *Cog.Dev.*, 6:17–46.

[14] Macnamara, J. (1982). *Names for Things*. MIT Press.

[15] Markman, E. and Hutchinson, J. (1984). Children's sensitivity to constraints on word meaning: taxonomic vs. thematic relations. *Cog.Psych.*, 16:1–27.

[16] Neal, R. (1992). Bayesian mixture modeling by monte carlo simulation. Technical report, University of Toronto.

[17] Osherson, D., Smith, E. E., Wilkie, O., Lopez, A., and Shafir, E. (1990). Category-based induction. *Psych.Rev.*, 97:185–200.

[18] Rogers, T., Garrard, P., McClelland, J., Ralph, M., Bozeat, S., Hodges, J., and Patterson, K. (2004). Structure and deterioration of semantic memory: a neuropsychological and computational investigation. *Psych.Rev.*, 111:205–235.

[19] Rogers, T. and McClelland, J. (2004). *Semantic cognition: A parallel distributed processing approach*. MIT Press, Cambridge, MA.

[20] Samuelson, L. and Smith, L. (1999). Early noun vocabularies: Do ontology, category organization, and syntax correspond? *Cognition*, 73:1–33.

[21] Soja, N., Carey, S., and Spelke, E. (1991). Ontological categories guide young children's inductions of word meaning: object terms and substance terms. *Cognition*, 38:179–211.

[22] Tenenbaum, J. and Xu, F. (2000). Word learning as bayesian inference. *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*.

[23] Waxman, S. and Gelman, R. (1986). Preschoolers' use of superordinate relations in classification and language. *Cog.Dev.*, 1:139–156.