Beyond Boolean logic: exploring representation languages for learning complex concepts

Steven T. Piantadosi, Joshua B. Tenenbaum, Noah D. Goodman
{ **piantado, jbt, ndg** } @ **mit.edu**
MIT Department of Brain and Cognitive Sciences
43 Vassar Street, Building 46, Room 3037G, Cambridge, MA 02139

Abstract

We study concept learning for semantically-motivated, set-theoretic concepts. We first present an experiment in which we show that subjects learn concepts which cannot be represented by a simple Boolean logic. We then present a computational model which is similarly capable of learning these concepts, and show that it provides a good fit to human learning curves. Additionally, we compare the performance of several potential representation languages which are richer than Boolean logic in predicting human response distributions.

Keywords: Rule-based concept learning; probabilistic model; semantics.

Introduction

Every cognitive theory requires a hypothesis about mental representation—what structures and operations form the basis for complex ideas. High-level, symbolic theories often characterize this representation using a *representation language (RL)* (Fodor 1975) that specifies primitive elements and composition laws which can be used to form complex cognitive structures. This approach has been extensively studied within two traditions in cognitive science: concept learning and linguistic semantics. In models of rule-based concept learning (Bruner, Goodnow, & Austin 1956, Shepard, Hovland, & Jenkins 1961, Feldman 2000) the representation language has typically been simple Boolean logic, which represents concepts—stable mental representations—using conjunctions, disjunctions, and negation of simple perceptual primitives. Goodman et al. (2008) presented a model of probabilistic learning for rule-based concepts that represents concepts in a simple propositional language and achieves state-of-the-art fits to experimentally-measured difficulty of learning Boolean concepts. The logical complexity of concepts appears to play a crucial role in determining how these concepts are learned: learners are biased to preferentially learn concepts with simpler representations.

However, simple Boolean concepts can capture only a very limited range of the human conceptual repertoire. People readily conceptualize context-dependent meanings such as "happiest," and can form more complex and abstract relational concepts like "everyone with two or more siblings." Semantic theories capture such meanings using primitive operations which manipulate and quantify over sets of objects, rather than simply features and propositional connectives. The denotation of a quantifier like "some," for instance, is a function which takes two sets, A and B, and is true only when the intersection of $A \cap B$ is nonempty[1] (Montague 1974).

In this paper we extend the probabilistic approach to concept induction to representation languages which manipulate sets of objects. We first describe an experiment that explores the difficulty of learning concepts that involve set-manipulation and quantification. Second, we compare human difficulty to the predictions of models with varying RLs. Our modeling work has two goals: the first is to test different RLs to see which provide the best account of people's learning behavior. Each possible RL differs in representational power and the way in which it assigns probability to potential concepts. This means that different RLs make different predictions about people's learning trajectories and we can therefore compare RLs by determining how well they match subjects' empirical response distributions. The second goal of the modeling work is to provide an explicit *learning theory* for these concepts. Work on boolean concept learning has provided a probabilistic model which accounts for subjects' behavior in acquiring boolean concepts, but there is no comparable formal theory for concepts which require a richer representation language. Such a theory would importantly extend rule-based concept learning in cognitive science to richer, linguistically-interesting semantic representations.

Behavioral experiment

The experiment we present aims to extend the rule-based concept learning paradigm to concepts which refer to sets and properties of sets of objects. To do this, we used a learning paradigm where subjects see a set of objects, guess at a labeling of the objects according to the unknown target concept, and receive feedback on their responses. Subjects used this feedback to infer the target concept.

Procedure

Amazon's Mechanical Turk was used to run 381 subjects. Each subject was told that they had to learn the meaning of a novel word, *wudsy*, from an alien language. Subjects were told that aliens use *wudsy*, to refer to some objects in a collection of objects, and that they have to figure out what makes an object *wudsy*. Subjects were informed that what makes an object *wudsy* may depend on which other objects are present.

During the experiment subjects were shown a set of four objects which varied in size, color, shape, and background color. An example set of items is shown below:

[1] In a sentence like "Some boy smiled," the set of boys would be A and the set of things which smiled would be set B. "Some boy smiled" is true if and only if the intersection of boys and things which smiled is nonempty.

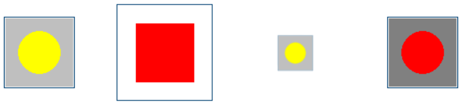

Figure 1: An example set.

After seeing a set of objects, subjects were told to guess which objects were the *wudsy* ones. For each object in the set, they were required to response "Yes," "No," or "NA," and were told to respond "NA" when it is unspecified whether an object is *wudsy*. For this example, subjects might entertain the concept is "red objects," in which case they should respond that the second and fourth objects are in the concept and the first and third are not. However, subjects might also entertain that the meaning of *wudsy* is context-dependent, as in, for instance "unique smallest." Similarly, the concept may also be complex, such as "same shape as the object with the darkest background." The shape with the darkest background is a circle, so subjects should say all the circles are in the concept; if all backgrounds are the same color, subjects may respond "NA." After responding, subjects were told what the correct answer was according to the target concept, but never given explicit instruction on the target concept. Subjects who responded incorrectly to any element of the set were penalized with a 5 second delay, during which they saw the set and the correct responses for each object.

Materials & Concepts

The meanings subjects were required to learn consisted of the concepts shown in Figure 2. These concepts include simple boolean rule-based concepts (e.g. "circles" and "circles or blue objects"), as well as more complex concepts which cannot be expressed in boolean logic ("larger than all the other objects"), and concepts which require several bound variables to express ("Same shape as the largest blue object").

Several of the concepts we studied focus on size predicates. This is because size predicates, such as "largest" and "smallest," are salient properties of objects in sets. They are perhaps the simplest words whose meaning is context-sensitive, and therefore not expressible with only conjunctions, disjunctions, and negation of object features. We included three simple size relations, "there exists a smaller object," "larger than all other objects," and "one of the largest objects." Note that the latter two differ with respect to uniqueness: if there are two objects of the maximal size, then neither is larger than all other objects, but both are one of the largest[2].

Because we included these simple size predicates, it is natural to include complex concepts which are also based on size, such as "same shape as the largest object," "same shape as the largest blue object," and "unique largest blue object." All three of these concepts require finding the largest object and selecting other elements based on the properties of the largest. As such, they require answering *NA* when there is not a unique largest element[3].

Results

The plots in Figure 2 show subjects' accuracy at labeling which objects are *wudsy* (y-axis), as a function of the amount of labeled data they received (x-axis). Subjects who were more than 3 standard deviations below the mean accuracy for each concept were removed in order to exclude subjects who were not performing the task. The vertical error bars show binomial 95% confidence intervals, and the red lines show the best fitting model, which is discussed in the next section. These results reveal several interesting qualitative trends. First, subjects accuracies increase for almost all of the concepts. Importantly, even though the subjects receive labeled data, they are never explicitly instructed on the concept. This means that high accuracy can only be achieved by generalizing from the observed data, which requires inferring abstract rules for these concepts.

Two interesting exceptions to subjects' general ability to learn these concepts are "Everything iff there is a triangle" and "Everything iff there is a single blue object." Subject performance on these concepts does not substantially improve, and these are intuitively somewhat unnatural concepts which require all elements of a set to be selected based on what the set contains. Words do exist with similar denotations in English—for instance, a set is *contaminated* if one element of the set is bad—but subjects find these types of concepts unusually hard to learn.

Figure 2 reveals a number of places where subject performance drops temporarily for a single set–for instance at item 32 of "there exists a smaller object." Post-hoc analysis revealed that many of these dips are caused when subjects see one of the first exceptions to a plausible alternative concept: item 32 is the first time that all objects in the set are the same size. Subjects responded *true* to objects in this item, consistent with a concept such as "not smaller than the rest," but incorrect according to the target concept.

Analysis

We first used a regression to analyze how subjects' learning rate varied across the 12 concepts studied[4]. In each logistic regression the outcome was whether the subject's response to each object in a set was correct, and the independent variable was the number of items each subject had seen so far (0...70). The key prediction we tested is whether slopes (regression coefficients)—which quantify the effect of additional data on accuracy—differed between concepts.

[2] These concepts are interesting in part because it is unclear which of these meanings corresponds to the denotation of "largest" in English, and also what role pragmatics plays in understanding "largest" in normal conversation.

[3] This makes it difficult to compare these concepts with the simple size-predicate concepts since the latter never require *NA*, which may be a difficult response for subjects to learn, independent of the concept.

[4] Because subjects typically were only run on one concept, subject effects are confounded with concept. We therefore performed a separate mixed-effect logistic regression (Gelman & Hill 2007) *within* each concept including slopes and intercepts by subjects. Regression coefficients across concepts were compared using t-tests.

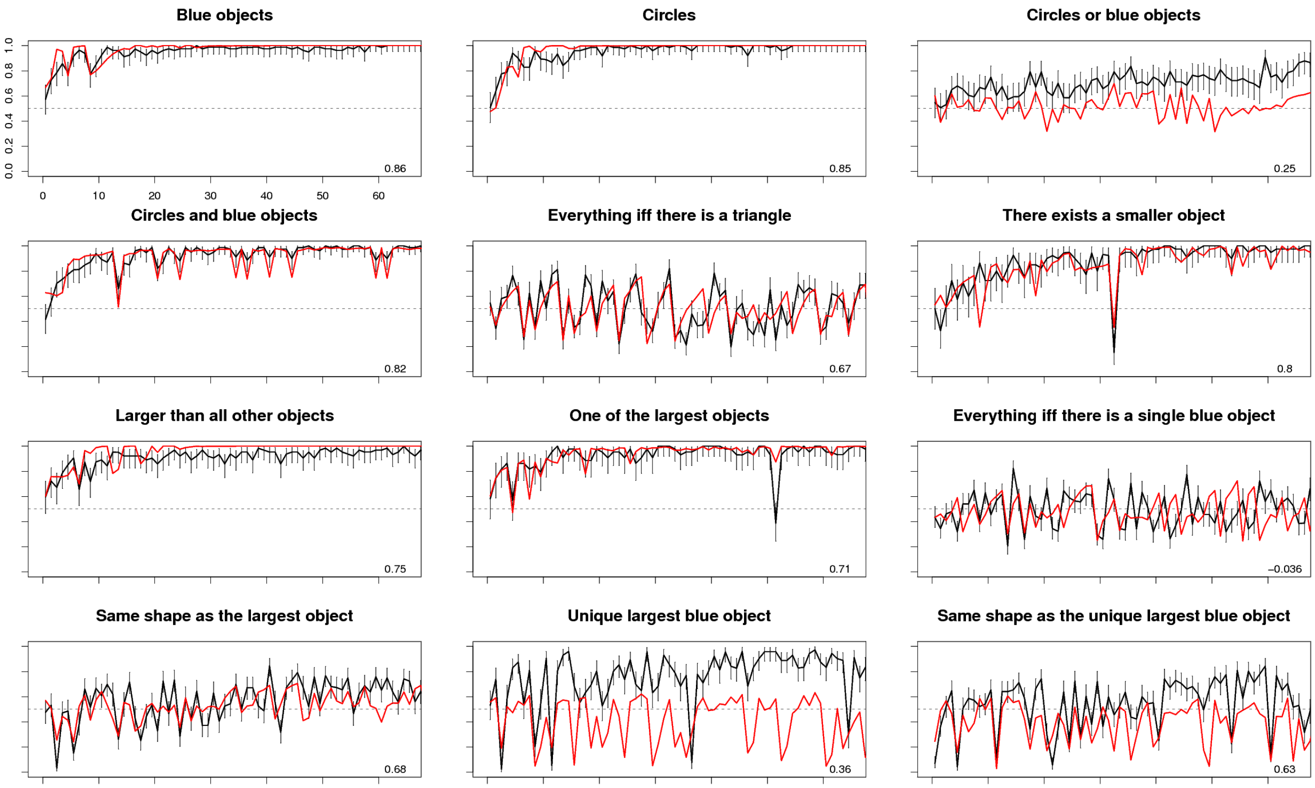

Figure 2: Subject accuracy (y-axis) in labeling the *wudsy* objects as a function of trial number (x-axis). Black lines show subject mean percent correct. Error bars are 95% confidence intervals. The red line shows the best-fitting model, although note that the model is fit based on agreement with the full distribution of responses, not the accuracies shown here. Numbers in the lower right show the correlation between the model and human accuracies.

Our results replicate basic effects in Boolean concept learning (Bruner, Goodnow, & Austin 1956, Shepard, Hovland, & Jenkins 1961). As is clear from Figure 2, simple concepts ("blue objects") are easier to learn than complex concepts ("circles and blue objects") ($t = 2.70, p < 0.01$). In addition, our results replicate that conjunctions ("circles and blue objects") are easier to learn than disjunctions ("circles or blue objects") ($t = 3.10, p < 0.01$). These replications provide validation for our experimental paradigm.

These effects of complexity also generalized to more complex functions than those expressible in Boolean logic. For instance, "The unique largest blue object" was easier to learn than "The same shape as the unique largest blue object" ($t = 2.71, p < 0.01$). This effect is interesting because it shows the additional difficulty associated with more complex set-theoretic concepts. The latter concept requires an additional bound variable to express in first-order logic, or a lambda abstraction to express in lambda calculus, and the effect of this complexity is reflected in subjects' learning rates. The regression revealed no difference between uniqueness presuppositions for concepts involving the largest element of a set: "larger than all other objects" was no more difficult than "one of the largest objects" in either slopes ($t < 1.26, p > 0.20$) or intercepts ($t < 1.83, p > 0.05$).

Importantly, subjects may infer a *different* concept from the one that was used to generate the data—high accuracy on some concepts can be achieved by inferring related concepts. To address this, we compared how well closely-related concepts predicted subjects' responses in the last half of the experiment. For each target concept in Table 1, we looked at data points for which the target concept made a different prediction from the specified alternative hypothesis. For instance, we looked at sets for which "Largest blue object" and "blue objects" made different predictions—that is, when there are multiple blue objects, so not all of them are the largest. We then computed the percent of subjects who responded more than half the time in agreement with the target concept, as well as the overall proportion of time subjects responded with the target concept. These results show that for most of the concepts, subjects typically responded in accord with the target concept and not a close alternative. The only exception to this is the comparison between "One of the largest objects" and "size = 5" (In the items, "5" is the maximal size for objects), which showed that subjects may have been learning to identify objects based on comparing them to absolute size, rather than a context-sensitive measure of "largest". In general, though, these results show that subject's pattern of learning cannot be explained by simpler theories which make reference only to only individual objects' properties. This is especially striking given that many of the alter-

Target	Alternative	Subject pct.	Response Pct.
Larger than all other objects	One of the largest objects	1.00	0.94
	Size = 5	0.81	0.67
	Size $\geq$ 4	1.00	0.82
	Size $\geq$ 3	1.00	0.88
Unique largest blue object	One of the largest objects	0.79	0.66
	Blue Objects	0.79	0.74
	Larger than all other objects	0.79	0.63
Same shape as the unique largest blue object	Same shape as the largest object	0.52	0.58
One of the largest objects	Size = 5	0.36	0.46
	Size $\geq$ 4	1.00	0.67
	Size $\geq$ 3	1.00	0.67
There exists a smaller object	Size = 5	1.00	0.67
	Size $\geq$ 4	1.00	0.73
	Size $\geq$ 3	1.00	0.91

Table 1: Comparison of subject agreement with target concepts compared to alternative concepts. *Subject pct.* shows the proportion of subjects agreeing more than 50% with the target concept, *Response Pct.* shows the overall percent agreement with the target.

native hypotheses are much simpler than the target concepts, and provides strong evidence that subjects are attending to more than simple object properties.

Computational model

The behavioral experiment shows that generally subjects are able to induce these types of set-theoretic concepts from the labeled data. Although it is important that subjects can eventually learn most of these concepts, we are also interested in whether their learning trajectory—their guesses and hypothesized concepts at each point in time—follow sensibly from the observed data. It may be rational to initially learn simpler related concepts which give approximately correct answers. There is no guarantee, for instance, that 70 items are enough to justify learning the correct form of the target concepts. We next present a computational model which can learn these types of set-theoretic concepts.

Our computational model aims to extend the *rational rules* model of Goodman et al. (2008) to a richer hypothesis space—one which is capable of representing these types of set-theoretic concepts. The probabilistic structure of the model and inference algorithm we use are neutral with respect to the RL, meaning that any potential RL can be incorporated and tested to see what distribution of responses it predicts for each object in each concept.

Each potential RL $\mathcal{L}$ defines a hypothesis space of potential concepts corresponding to the set of all ways to compose the RL's primitive functions in order to create functions which map objects to labels (*true*, *false*, *NA*). For instance one concept might be[5]

$$\lambda x.(\mathit{red}\ x) \wedge (\mathit{square}\ x)$$

This expression represents a function which checks whether its argument, x, is red and a square, and returns *true* or *false* for any object (since *red* and *square* are assumed to return *true* or *false*). We might also have concepts which take two arguments, a contextually-relevant set S and an element x:

$$\lambda S\, \lambda x.(\mathit{equal}\ x\ (\mathit{unique\text{-}smallest}\ S))$$

This function checks if x is the object which is the unique, smallest element of S.

We use a probabilistic context-free grammar (PCFG) which assigns probability to every possible composition of primitive elements. This PCFG functions as a prior over concepts and for simplicity, we assume that all PCFG expansions are equally probable[6]. In general, the PCFG assigns high probability to short or "simple" compositions of $\mathcal{L}$'s primitives, and lower probability to complex rules. For instance, a function $\lambda x.(\mathit{red}\ x)$ will be higher probability a-priori than $\lambda x.(\mathit{red}\ x) \wedge ((\mathit{square}\ x) \vee (\mathit{circle}\ x))$. This captures the notion that people should be biased to prefer simple explanations of the labeled data they observe.

The second part of the model is a likelihood function which provides the probability of labels according to a hypothesized RL expression. Specifically, for any composition E of primitives in $\mathcal{L}$, the correct label is generated with probability α, and a label is chosen uniformly at random with probability $1-\alpha$. However, it is also likely that memory factors come into play in remembering past labeled examples. We include this in the model by weighting the log likelihood for the n'th data point back in time by $n^{-\beta}$, where $\beta > 0$. As $\beta \to 0$, the model has perfect memory, and as $\beta \to \infty$ the model quickly forgets past data points. This leaves us with two unknown free parameters: α, which controls how reliably set elements are labeled, and β which controls how much more recent set elements matter than past ones.

Together, the prior and likelihood specify a complete probabilistic model for any RL. Formally, we can score the probability of a hypothesized concept expression E conditioned on a collection of example sets S with corresponding labels L according to Bayes rule:

$$P(E \mid S, L, \mathcal{L}) \propto P(L \mid S, E) P(E \mid \mathcal{L}). \qquad (1)$$

Here, $P(E \mid \mathcal{L})$ is the probability of E according to the PCFG for $\mathcal{L}$ and $P(L \mid S, E)$ scores the likelihood of the labels L under the observed sets of objects S and hypothesized expression E. While Equation 1 scores the probability of any given expression E, it is a complex inference problem to actually determine what expressions are likely given the data. This problem is difficult because the space of possible expressions E is in principle infinite and difficult to search. We solved this problem using a Markov-Chain Monte-Carlo (MCMC) similar to Goodman et al (2008)'s method, which takes samples

[5] We write functions as lambda expressions, meaning that the name for the argument is preceded by λ. We also use prefix notation: a function f applied to an argument x is written $(f\ x)$.

[6] Unlike the rational-rules model, we do not integrate out the PCFG production probabilities. This is because primitives which introduce new bound variables, such as quantifiers, make this integration difficult and potentially not analytically tractable in general.

from the posterior distribution $P(E \mid S, L, \mathcal{L})$. This method takes a biased random walk around the space of hypotheses by making local changes to hypothesized expressions E, and can be shown to, in the limit, draw samples from the posterior distribution. We ran the MCMC algorithm for a range of α and β values for each amount of data, in each sequence, conditioning on the correct, observed labels for all previous sets in the sequence. This gives a distribution $P(E \mid S, L, \mathcal{L})$ on expressions E in the RL $\mathcal{L}$ at each point during learning. These expressions can be evaluated on the next item in order to provide a model prediction of subject's distribution of responses, conditioned on the observed labeled data. Thus, the model was run conditioned on the same labeled data human subjects were given, and—just like human subjects—was asked to make predictions about the correct labels for the next data point. Ideally, subjects' distribution of responses at each point in time during learning should correspond to the predictions of the model, conditioned on the exact same sequence of training data.

One goal of the model is to test different representational languages to see which provide the best theory of people's inductive biases in learning these concepts. We computed the posterior predictive distribution of responses for each representation language $\mathcal{L}$ and saw which assigned the human responses highest likelihood[7]. We compared four different RLs with differing primitives and representational power:

Language	**Primitive Operations**
RESPONSE-BIASED	*true*, *false*, *undefined*
SIMPLE-BOOLEAN	*and*, *or*, *not*, *shape*, *size*, *color*, *background-color*, *equal*
SET-FUNCTIONS	*contains*, *filter*, *only*, *unique-largest*, *unique-smallest*, *set-of-largest*, *set-of-smallest*, *same-object*
QUANTIFIERS	*exists*, *forall*

Each RL is a *superset* of the preceding languages, except that none other than RESPONSE-BIASED contain *true*, *false*, and *na* as primitives. Here, *shape*, *color*, and *background-color* are functions which extract the corresponding properties of objects. *equal* tests if two properties are equal. *contains* returns true if a set contains an element, *filter* removes all elements in a set not satisfying a predicate, and *only* return the only element of a set and *NA* if the set has more than one element. The primitives *unique-largest* and *set-of-largest* return the unique largest element in a set (and *NA* otherwise), and the set of elements for which none are larger, respectively. *same-object* tests if two objects are identical on all dimensions. *exists* and *forall* are first-order existential and universal quantifiers.

Intuitively, the RESPONSE-BIASED language allows learners only to infer a distribution on responses, but not give responses which depend on the current objects. This serves as a baseline, and way to test if subjects are really performing the task. The SIMPLE-BOOLEAN language is one which include basic logical operations and object properties, and implements the representational system studied most in previous rule-based concept learning experiments. The SET-FUNCTIONS language extends the SIMPLE-BOOLEAN language by including primitive operation for testing if sets contain elements, extracting sets or elements with maximal or minimal properties along the size-dimension and filtering sets by elements. The QUANTIFIERS language extends the SET-FUNCTIONS language by incorporating quantification.

Results & Discussion

Table 2 shows the performance of these models in predicting the human distribution of responses across the 12 concepts studied. This shows the average log-likelihood of the human responses for the best-fitting values of α and β within each concept[8]. This table illustrates several key properties of the RLs. First, the RESPONSE-BIASED model is overall the worst predictor of human responses. This is important because it shows that subjects are performing the task, and performing nontrivial inferences about the target concepts.

In addition, this figure shows that while SIMPLE-BOOLEAN is a good predictor for the simple Boolean concepts, SET-FUNCTIONS and QUANTIFIERS provide a better account for the set-theoretic concepts that subjects are able to learn. SIMPLE-BOOLEAN provides the worst account for "same shape as the largest object" and "same shape as the unique largest blue object." While subjects do not learn these concepts especially well, these results show that the SIMPLE-BOOLEAN does not account well for subject responses.

Overall, the best RL is QUANTIFIERS; however, the differences between QUANTIFIERS and SET-FUNCTIONS is small. Richer representation languages not only have the formal power to represent the types of set-theoretic and logical concepts required by human conceptual systems, but also provide a better account of human inductive leaning than the other RLs considered here.

As discussed above, the black line in Figure 2 shows learning curves showing percent accuracy over time for human subjects. This figure also shows a red line, corresponding to the performance of the RL QUANTIFIERS for the best-fitting α and β within each concept. We chose the best-fitting model parameters based on which parameter values assigned highest likelihood to the observed *distribution* of human responses, "true," "false," and "NA." Doing this does not necessarily provide the best fit to the human learning curves in Figure 2 since the model is not fit to human *accuracy* (correct/incorrect). This means that Figure 2 shows a conservative view of the agreement between human accuracies and model accuracies. For the concepts "circles or blue objects" and "unique largest blue object" the model's learning trajectory would increase

[7] Model predictive distributions were smoothed to give each response a minimum possible probability of 0.01, to prevent divergence.

[8] That is, these numbers are the total log likelihood assigned to human responses, divided by the number of responses. This was necessary for cross-concept comparison since concepts may have differing numbers of subject responses.

Concept	RESPONSE-BIASED	SIMPLE-BOOLEAN	SET-FUNCTIONS	QUANTIFIERS
Blue objects	-0.66	-0.18	-0.19	-0.18
Circles	-0.73	-0.17	-0.17	-0.17
Circles or blue objects	-0.81	-0.73	-0.74	-0.74
Circles and blue objects	-0.30	-0.27	-0.27	-0.27
Everything iff there is a triangle	-0.80	-0.73	-0.73	-0.73
There exists a smaller object	-0.81	-0.51	-0.40	-0.41
Larger than all other objects	-0.58	-0.48	-0.46	-0.36
One of the largest objects	-0.80	-0.63	-0.28	-0.28
Everything iff there is a single blue object	-0.85	-0.78	-0.78	-0.78
Same shape as the largest object	-1.10	-1.75	-0.99	-0.99
Unique largest blue object	-1.05	-1.54	-1.06	-1.06
Same shape as the unique largest blue object	-1.08	-1.34	-1.06	-1.04
Mean	**-0.797**	**-0.760**	**-0.594**	**-0.584**

Table 2: Model log likelihoods *per response* for each concept. These represent the model log likelihood assigned to human responses, divided by the total number of responses in each concept to allow comparisons across concepts.

more for other values of α and β, and thus look more like subject's accuracies, but provide a less-good fit to subjects' overall response distribution.

This figure shows good fit between the probabilistic model and human learning. This fit appears especially remarkable for concepts which subjects have a difficult time learning, such as "Everything iff there is a triangle." Because subjects do not learn this concept well, the best-fitting α is low and β is highly negative, meaning that the model is not penalized much for incorrect answers and down-weights old data. The model therefore responds with in simple ways, such as always responding true, or responding true to only the triangles; subjects appear to use similar strategies, and thus both show similar patterns of response accuracies[9]

The model also shows more subtle agreement patterns with human subjects. First, it is capable of learning simple boolean concepts in a way similar to humans, quickly arriving at the correct meaning given the training data. This is also true for concepts like "there exists a smaller object" and the other size-predicates. The model also matches local dips and peaks in reasonably well. This is because the model, like people, may temporarily be led to a concept which is not the target concept, just as subjects (e.g. at item 32 of "there exists a smaller object"). This provides evidence that people make the same rational, statistical inferences given the same data.

Conclusion

While the SIMPLE-BOOLEAN RL provided a good fit to human response data in some cases, it is insufficient to represent some of the complex concepts that subjects learned. Subjects' ability to learn these concepts was demonstrated by their learning curves for several context-dependent concepts. The comparison of different RLs suggests a potentially fruitful approach to discovering the precise form of semantic representations. Recently, Pietroski et al. (2009) and Hackl (2009) have used psychophysical measures to make inferences about plausible representations and computations that underlie semantic meaning for words like "most." Our work provides a complementary approach to the same problem—instead of measuring response times, we studied what RLs provide a good account of human inductive biases during learning. This method may be broadly applicable to discovering the form of semantic representations in natural language.

Of course, the RLs we study here are still incomplete with respect to the full richness of human conceptual systems; however, this work suggests that rule-based concept-learning can be extended to complex concepts which can begin to approach the complexity and context-dependence observed in human linguistic systems. Furthermore, the model provides one potential acquisition theory for semantic concepts. Children may learn semantic meanings like adults in our experiment did—by inducing concepts in a sufficiently-powerful compositional RL.

References

Bruner, J. S., Goodnow, J., & Austin, G. (1956). *A study of thinking*. New York: Wiley.

Feldman, J. (2000). Minimization of boolean complexity in human concept learning. *Nature*, *407*, 630-633.

Fodor, J. (1975). *The language of thought*. Cambridge, MA: Harvard University Press.

Goodman, N., Tenenbaum, J., Feldman, J., & Griffiths, T. (2008). A rational analysis of rule-based concept learning. *Cognitive Science*, *32*, 108-154.

Hackl, M. (2009). On the grammar and processing of proportional quantifiers: most versus more than half. *Natural Language Semantics*, *17*, 63-98.

Montague, R. (2002). The proper treatment of quantification in english. In P. Portner & B. H. Partee (Eds.), *Formal semantics: The essential readings.* Oxford: Blackwell.

Pietroski, P., Lidz, J., Hunter, T., & Halberda, J. (2009). The meaning of most: semantics, numerosity, and psychology. *Mind and Language*, *24*, 554-585.

Shepard, R., Hovland, C., & Jenkins, H. (1961). Learning and memorization of classifications. *Psychol. Monogr. Gen. Appl.*, *75*, 1-42.

[9] The best fitting β also shows a modest negative correlation ($R = -0.55$, $p = 0.06$, $N = 12$) with response accuracies over the 12 concepts, suggesting an interaction between the target concept and the attentional or memory resources people allocate.