A Rational Constructivist Account of the Characteristic-to-Defining Shift

Francis Mollica, Shirlene Wade, Steven T. Piantadosi
{fmollica | swade | spiantadosi}@mail.bcs.rochester.edu
Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627 USA

Abstract

A widely observed phenomenon in children's word-extensions and generalizations is the characteristic-to-defining shift, whereby young children initially generalize words based on typical properties and gradually transition into generalizing words using abstract, logical information. In this paper, we propose a statistically principled model of conceptual development grounded in the trade-off between simplicity and fit to the data. We run our model based on informant-provided family trees and the real-life characteristic features of people on those trees. We demonstrate that the characteristic-to-defining shift does not necessarily depend on discrete change in representation or processes. Instead, the shift could fall out naturally from statistical inference over conceptual hypotheses. Our model finds that the shift occurs even when abstract logical relations are present from the outset of learning as long as characteristic features are informative but imperfect in their ability to capture the underlying concept to be learned—a property of our elicited features.

Keywords: characteristic-to-defining shift; concept learning; development; computational modeling

Introduction

Children can often comprehend a word and use a word without having a full grasp of its meaning. Consider the following scenario from Keil and Batterman (1984, pp. 226): "This smelly, mean old man with a gun in his pocket came to your house one day and took your colored television set because your parents didn't want it anymore and told him that he could have it." While adults have a strong sense that the man in the scenario is not a robber, young children are willing to label the man a robber. Across multiple domains, young children have been shown to initially privilege perceptually-observable, characteristic information in concept learning. Eventually, children transition to more abstract, conceptually-aligned-upon meanings. This phenomena has been termed the *characteristic-to-defining shift* (Keil & Batterman, 1984).

Previous research has suggested that perceptual similarity (e.g., shape) plays a strong role in young children's early word-concept mappings (e.g. Landau, Smith, & Jones, 1988). As children age, they begin to use deeper, relational properties for concept learning (Imai, Gentner, & Uchida, 1994; Keil & Batterman, 1984). For example, Keil and Batterman (1984) probed kindergartners', second graders', and fourth graders' definitions for several words using a scenario task. In some scenarios, characteristic features of a term were presented without the defining features of the term; whereas, other scenarios provided the defining features of the term without the typical characteristic features associated with the term. Younger children extended a word's meaning to more scenarios lacking defining features—but possessing many characteristic features—than older children.

While the characteristic-to-defining shift is commonly observed in concept acquisition, the process by which this occurs is unclear. One possibility is that the characteristic-to-defining shift is a stage-like transition that occurs in the representational system (Werner, 1948; Bruner, Olver, Greenfield, et al., 1966). For example, the shift could be explained by a transition from representing concepts wholistically—i.e., using all the features of objects, to representing concepts analytically—i.e., narrowing in specific relevant features of objects (Kemler, 1983). Neural network models of conceptual classification inherently capitalize on this idea when demonstrating a shift (e.g., Shultz, Thivierge, & Laurin, 2008). Another possibility is that there is a change in the mechanism by which one learns concepts. For example, concept learning might change from storing exemplars to constructing prototype or rule-based representations. These hypothetical changes in representation or processing might be maturational in nature, such as the development of abstraction (Piaget & Inhelder, 1969). Alternately, they may be driven by inductive inference mechanisms operating over observed data, a la rational constructivism (Xu, 2007).

From the outset we can narrow down this space of hypotheses. The conceptual to defining shift is most likely a function of data, not maturation (Keil, 1983). One prediction of a maturational-shift is that at a single time-point, children should represent all words using characteristic features or defining features, whereas a data-driven shift predicts that both adults and children should have more exemplar-based representations in unfamiliar domains, and more rule-based representations in familiar domains. The former does not explain children's behavior—children seem to possess characteristic representations and defining representations of different words at a single time point. The prediction of the latter—that individuals have more exemplar-based representations in unfamiliar domains and more rule-based representations in familiar domains, is observed in adults (Chi, Feltovich, & Glaser, 1981) and in children (Chi, 1985).

All of the aforementioned hypotheses require a discrete shift in representation or process. However, it is unclear whether a representational or mechanistic shift is entirely warranted. To date, no model has tested whether a characteristic-to-defining shift could be a natural by-product of the continuous data-driven construction of concepts. We evaluate this proposal in the task of learning kinship concepts. While "mommy" and "daddy" are some of a child's earliest produced words, children actually spend many years mastering kin relations (e.g. Haviland & Clark, 1974; Benson & Anglin, 1987; Keil & Batterman, 1984). For example, 7- and 8-year-olds are still unable to provide adequate definitions for a number of kinship terms (Haviland & Clark, 1974).

The Acquisition of Kin Terms

Kinship is an ideal domain for studying the characteristic-to-defining shift because it easily lends to logical representations (e.g. Kemp & Regier, 2012); the domain of kinship is familiar to young children; and the characteristic and defining features behind kinship terms are fairly intuitive and straightforward. Furthermore, several semi-structured interviews attempting to uncover children's knowledge of kinship demonstrate considerable variation in children's definitions. For example, the following is an interview with a six-year-old from Benson and Anglin (1987, p. 48):

I: What is an uncle?
S: A man that's related to ya.
I: Tell me everything you know about an uncle.
S: He knew you when you were a baby...Sometimes they work to build houses... Sometimes they join in for the army.
I: Can you tell me anything else about an uncle?
S: They help you. That's all I know.
I: What kind of a thing is an uncle?
S: A man that's related to you.

Based on children's definitions, researchers have proposed theories weighing the importance of conceptual simplicity (Haviland & Clark, 1974) and the role of sufficient data (Benson & Anglin, 1987) in the acquisition of kinship terms. To explain the order of acquisition of kinship terms, Haviland and Clark (1974) proposed a semantic complexity hypothesis. In this account, the simplicity of a concept is defined as the fewest number of base relations (e.g., up one node on the family tree) required to explain a relationship on a kinship tree with a penalty on the variety of base relations used. Children use these base relations to build concepts in a piecewise fashion. By this logic, adult-like kinship concepts are acquired for semantically simpler terms before semantically complex terms. Haviland and Clark (1974)'s original hypothesis is a learning model whereby children first develop perceptual features to construct a concept and only over time develop abstract, relational features. This formalism is entirely consistent with the formalisms used in Mollica and Piantadosi (2015), which we also adopt and describe below. Furthermore, simplicity, in general, is an empirically grounded principle underlying concept construction (Feldman, 2000). More specifically, the role of simplicity and communicative efficiency in kinship terms has been demonstrated across a variety of the world's languages (Kemp & Regier, 2012).

In addition to simplicity, researchers have proposed that the amount and quality of the observed data drive word learning and conceptual development both in kinship (Benson & Anglin, 1987; Danziger, 1957) and in other domains (e.g., Weisleder & Fernald, 2013). For example, Benson and Anglin (1987) found that the order of acquisition of kinship terms was best predicted by children's experience with their relatives. In his rejection of stage theories, Danziger (1957) proposed that conceptual development is primarily driven by opportunities provided by the environment. To account for the influence of data, we incorporate assumptions about plausible data distributions in our model. Further, we trade off the influence of data with semantic complexity by placing a simplicity weighted prior against a data-driven likelihood.

Our Approach

We approach this problem at the computational level of analysis (Marr, 1982) to demonstrate how an ideal learner would manifest a characteristic-to-defining shift. We start with the model of Mollica and Piantadosi (2015), which demonstrates how a learner could use cross-situational word-referent occurrences to infer the concept that licenses how a word should be extended. We extend the Probabilistic Context-Free Grammar (PCFG) in their model to construct both characteristic and defining hypotheses for kinship terms. We then collected data about the characteristic and logical relationships from two naive informants' own family trees. This is important because the characteristic and logical relationships of real people allows us to test if natural data will contain perceptual and experiential features informative enough to observe a characteristic to defining shift. We ran the model on the informant-provided trees and a simulated tree to generate possible characteristic and defining hypotheses for four kinship concepts: BROTHER, GRANDMA, MOTHER and UNCLE. These hypotheses were then scored using Mollica and Piantadosi (2015)'s Bayesian model according to their simplicity and ability to explain simulated word-referent data. We analyzed (1) whether an ideal learner is most likely to entertain characteristic or defining hypothesis given an amount of data and (2) the accuracy of the hypotheses in explaining the data as a function of the amount of data observed.

We expect that an ideal learner (without any maturational factors) should demonstrate a characteristic-to-defining shift only if the elicited features (both perceptual and experiential) are informative but imperfect in their ability to capture the underlying concept. If the elicited features accurately capture a concept, an ideal learner should never shift from generating characteristic hypotheses to defining hypotheses. On the contrary, if the elicited features are uninformative, and thus poor at capturing a concept, an ideal learner might predominately generate defining hypotheses, predicting either no shift or an implausibly rapid shift from characteristic to defining hypotheses. Therefore, it is crucial that we collect data about the characteristic and logical relationships of real people to test if natural data will contain features within the range of informativity that will show a characteristic-to-defining shift.

Data Collection

To simulate data for the learning model, two informants, who were blind to the experiment, drew their family tree, ranked each member in terms of how frequently they interacted with them as a child (e.g., see Figure 1), and provided ten one-word adjectives for each family member. For each informant, the unique adjectives were used to construct a binary feature matrix (adjective by family member; e.g., see Figure 2). Each informant was presented with the feature matrix and asked to indicate if each feature applied to each family member. Informants made a response to every cell of the matrix: zero if

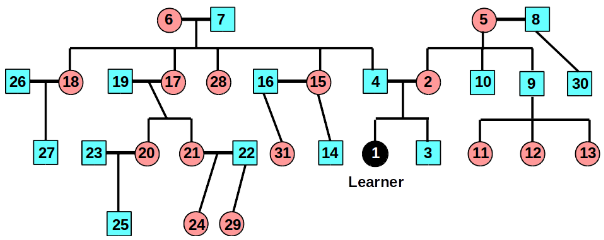


Figure 1: Distance-ranked family tree from Informant One.

START → CHAR	START → DEF
CHAR → union(CHAR, CHAR)	DEF → union(DEF, DEF)
CHAR → intersection(CHAR, CHAR)	DEF → intersection(DEF, DEF)
CHAR → set_difference(CHAR, CHAR)	DEF → set_difference(DEF, DEF)
CHAR → complement(CHAR)	DEF → complement(DEF)
CHAR → feature(VAL)	DEF → up_node(DEF)
VAL → {Yes, No}	DEF → down_node(DEF)
	DEF → lateral_node(DEF)
	DEF → male(DEF)
	DEF → female(DEF)
	DEF → X (i.e., the speaker)

Table 1: The Probabilistic Context Free Grammar used to generate kinship concepts

the feature did not apply; one if the feature did apply. The informants provided 109 and 88 unique features respectively including both experiential features (e.g., *strict*) and perceptually observable features (e.g., *blonde*). Additionally, we simulated data from the extended family tree used in Mollica and Piantadosi (2015). To sample from the extended tree, we ranked distance using Euclidean distance and constructed a feature matrix for the tree based on 29 perceptually observable features following the principles of Mendelian genetics[1].

Extending the model

The model incorporates a PCFG prior with uniform rule probabilities to measure the simplicity of any composition of logical or perceptual features. In the PCFG (see Table 1), we include set theoretical primitives—i.e., union, intersection, set_difference and complement—for both characteristic and defining hypotheses. For defining hypotheses, we include gender primitives—i.e., male and female—and graph theoretical primitives that mimic the abstract primitives proposed by Haviland and Clark (1974): up_node, down_node and lateral_node. The terminal for a defining hypothesis is an argument for the speaker X, as we assume that the kinship term should be processed relative to the speaker. For characteristic hypotheses, we include a primitive for each feature, which takes a binary indicator variable and returns the set of family members with or without that feature. Using a PCFG as a prior penalizes complex hypothetical meanings and, thus, builds in a simplicity bias as discussed above. It is important to note that our PCFG generates characteristic hypotheses—i.e., only containing characteristic information, and defining hypotheses—i.e., only containing logical information (and gender). We leave the exploration of a hybrid characteristic-defining hypothesis space for future research.

Data for the learning model was noisily sampled from a family tree such that 90% of the time the data reflected accurate use of the true concept and 10% of the time the data was entirely random. To construct a data point, which took the form of a speaker-referent pair $\{s, r\}$, we first sampled a speaker s from a Zipfian distribution over all members of the family tree ordered by reported distance from the learner. Consequently, data from speakers ranked closer in distance to the learner were more likely to be sampled than data from speakers ranked distant to the learner, which is in line with the intuition that most input a child receives comes from her immediate family. We then sampled a referent r from the Zipfian distribution conditioned on the speaker and word. Given all possible referents the speaker could be correctly referring to when using the word, referents that are closer to the learner are more likely to be talked about than the learner's more distant relations. This reflects the intuition that a child is more likely to hear about her immediate family than distant relatives. Both intuitions are supported by Benson and Anglin (1987)'s survey of children's experience with kinship terms and relations. During learning, we compute the likelihood of the data under the same model used to simulate the data.

Together the prior and the likelihood specify a model for all possible hypotheses constructed from the PCFG:

$$P(h|\{s,r\}_i^N) \propto \prod_i^N P(r_i|s_i,h) \cdot P(h) \tag{1}$$

With this model we can score the probability of a hypothesis conditioned on simulated data. We then investigate the conditions under which a characteristic-to-defining "shift" will naturally emerge as children learn kinship concepts without positing discrete change.

Methods

Discovering the most likely hypotheses considered by an ideal learner according to Equation 1 is a complex inference problem because the PCFG specifies an infinite set of possible hypotheses. We solved this problem with Markov-Chain Monte-Carlo (MCMC) methods, which provided us with samples from the posterior distribution by walking around the space of hypotheses. In the limit these walks provably draw samples from the true posterior distribution $P(h|\{s,r\}_i^N)$. We implement our model using LOTlib (Piantadosi, 2014).

At different amounts of data, we expected an ideal learner to favor different hypotheses. Therefore, we explored the space varying the amount of data between 10 data points and 250 data points by 10 point increments. At each increment of data, we ran eight chains per hypothesis type for one million steps. We stored the top 1000 hypotheses from each chain and combined the hypotheses discovered across chains to form a finite hypothesis space representing the posterior distribution over hypotheses. We normalized all hypotheses by calculating the likelihood over the same set of 1000 data points generated using the same procedure used to generate data for individual chains. We then divided this value by the amount of data (i.e., 1000) to get a measure of each hypothesis' average log likelihood per data point.

[1] All family trees, feature matrices (and code) can be found at `https://github.com/MollicaF/LogicalWordLearning`

Results

The upper panels of Figure 4 show the posterior probability of entertaining either a characteristic or defining hypothesis (y-axis) as a function of the amount of data observed (x-axis). For all of the words, we observe the characteristic-to-defining shift–i.e., the probability of entertaining a characteristic hypothesis is initially greater than the probability of entertaining a defining hypothesis. This means that a simple conceptual learning model shows a characteristic-to-defining shift when given real data about logical relations and characteristic features.

The lower panels of Figure 4 show the posterior weighted F_1 score conditioned on hypothesis type (characteristic or defining). The F_1 score is the harmonic mean of precision (i.e., the pressure to extend without over-extending) and recall (i.e., the pressure to extend to all the correct referents). An F_1 score of 1 reflects perfect performance. Notably in Figure 4, the model successfully learns BROTHER, GRANDMOTHER, and MOTHER—i.e., posterior weighted F_1 scores all reach 1. With 250 data points, the model does not successfully learn UNCLE yet there still is a shift from characteristic to defining hypotheses on a larger timescale[2] (Note the x-axis in the upper panels).

To help build intuitions about how the model works, Figure 3 presents the three most likely hypotheses an ideal learner trained on Informant One's data would consider for GRANDMA at three time points. Before observing data, an ideal learner should prefer the simplest hypotheses, which often generalize to many referents. In this example the three most likely hypotheses are defining hypotheses that select the speaker X, a male speaker and everybody but the speaker. After observing three data points, the hypotheses considered become much more plausible and shift to characteristic features. At this time, the three best hypotheses for grandma are that grandmas are either outgoing, nosy or small. In general the model is shifting from simple hypotheses that generalize broadly to hypotheses that narrow in a bit more, yet still over-extend. Immediately post-shift (i.e., 13 data points), we observe a mixture of characteristic and defining hypotheses. The best hypothesis is the speaker's parents' parents, which misses the female component of GRANDMA. The next best hypothesis is that grandmas are outgoing. The third best hypothesis is actually the definition of a GRANDMA—i.e., the female parents of the speaker's parents. This glimpse at the hypotheses just after the shift illustrates that without a sufficient amount of data, even the correct hypothesis is unlikely because it is more complex in the prior. As we observe more data, the imprecision of the two leading hypotheses decreases their posterior probability relative to the correct hypothesis, which will make the correct hypothesis the maximum a posteriori (MAP).

The range of hypotheses are similar between the different trees. Across all trees, characteristic hypotheses have very low posterior weighted F_1 scores compared to the defining hypotheses. In other words, characteristic hypotheses mislabel referents more than defining hypotheses. Yet, the posterior probability of characteristic hypotheses suggests that characteristic hypotheses are clearly favored at low amounts of data. Given the perspective that the emergence of defining hypotheses is delayed due to the development of abstraction, it is particularly important to note that even in a model with abstraction available from the beginning, we observe a characteristic-to-defining shift. Further, compared to a neural network model where all features are initially considered (Shultz et al., 2008), a characteristic-to-defining shift is still observed in our model where it is initially more likely to only consider only a few features.

Taken together, this pattern of results demonstrates that the characteristic-to-defining shift could naturally fall out of a single statistical inference process with a single representational language[3]. It is not necessary to propose a discrete change in representation or processing. Characteristic hypotheses are favored early because with little data the prior dominates inference–they generalize well to small data amounts and are comparatively less complex in the prior than the best defining hypotheses. Only when there is enough data to warrant additional complexity will defining hypotheses come to dominate inference.

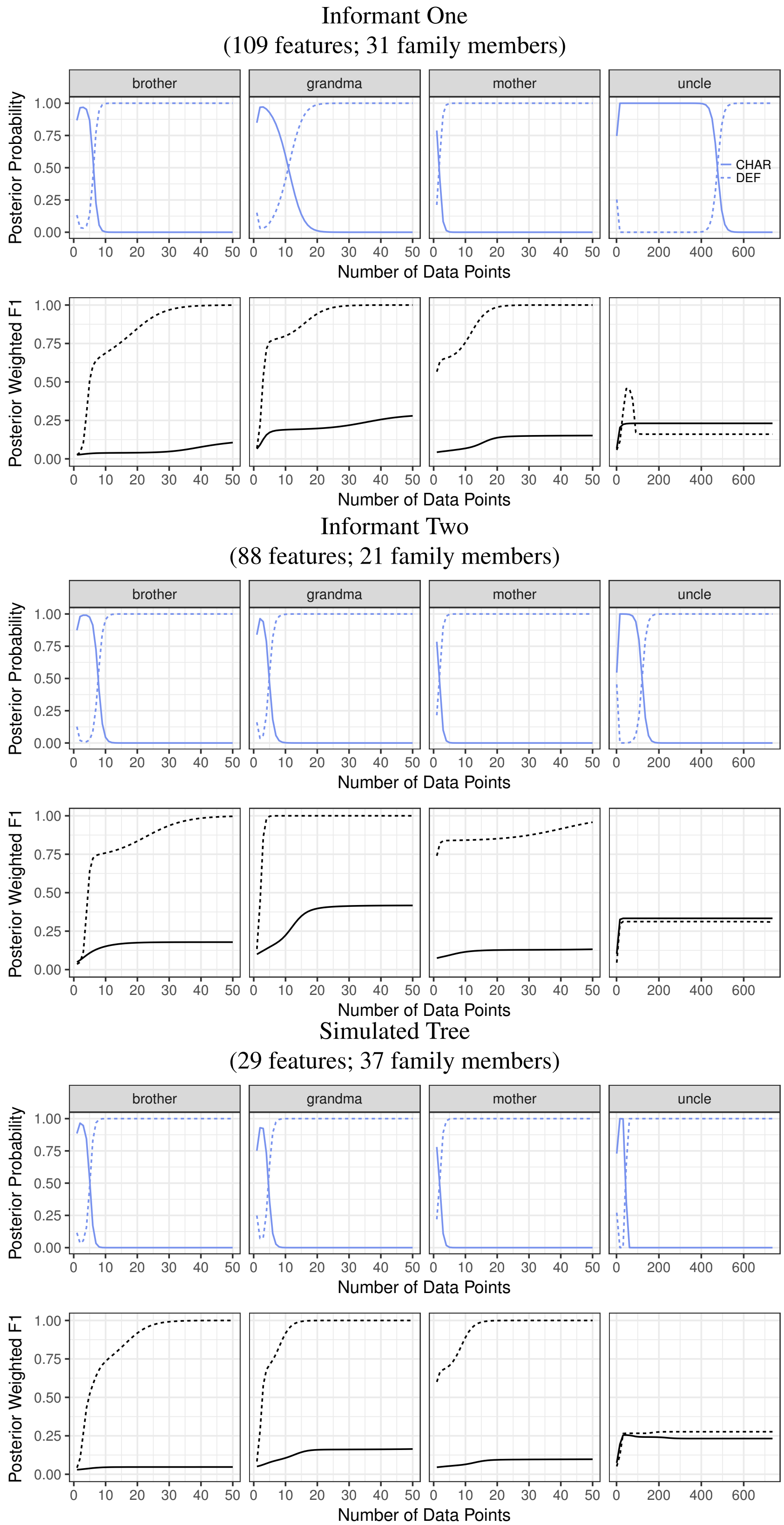
Discussion

In this paper, we tested whether a characteristic-to-defining shift would emerge naturally in a statistically principled learning model without positing a discrete mechanistic or representational shift. In general, the model successfully learns kinship terms and demonstrates a characteristic-to-defining shift using a single representational language of thought and a single statistical inference mechanism. Therefore, while a discrete shift in mechanism or process is possible, it is not necessary to observe a characteristic-to-defining shift during concept learning.

In our model, kinship concepts are developed through statistical inference over word-referent data and observed kinship structures, which could plausibly be developed from statistical learning of structure (Katz, Goodman, Kersting, Kemp, & Tenenbaum, 2008; Kemp & Tenenbaum, 2008). When an ideal learner only observes data about a few referents, there are simple characteristic hypotheses based on perceptual observations that will explain the data; however, as more data is observed, these hypotheses fail to adequately fit the data and warrant a prior-likelihood trade-off, such that more complicated defining hypotheses (which are unlikely

[2]This may be due to data sparsity for UNCLE in the trees. As UNCLE is the most complex concept learned here, it may be that UNCLE requires more unique data points to be learned. Under our Zipfian data sampling, the model receives data for less than half of the unique uncles in the trees. When you relax the sampling assumption to uniform, the model does learn UNCLE and having the correct hypothesis in the space alters the time scale of the shift (to around 30 data points).

[3]This pattern holds if the data distribution is uniform or becomes more peaked—i.e., a Zipfian exponent of 0, 1 or 2.



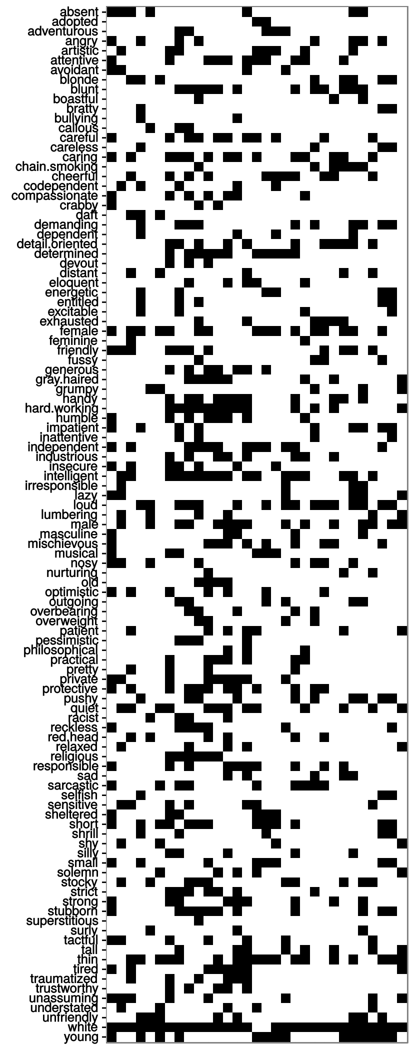


Figure 2: Feature matrix (adjective by family member) supplied by Informant One.

Before seeing data	
X (i.e., the speaker)	−0.0861777
male(X)	−4.7775256
complement(X)	−4.7775256
After seeing 3 data points	
outgoing	−19.69045
nosy	−20.49538
small	−21.56817
One data point after shift	
parents(parents(X))	−67.18689
outgoing	−67.31635
female(parents(parents(X)))	−68.14575

Figure 3: Best hypotheses at three different time points and their log posterior probability for Informant One learning GRANDMA.

Figure 4: For each tree, the top panel displays the posterior probability of using a characteristic (solid line) or a defining (dashed line) hypothesis as a function of the amount of data observed. The bottom panel displays the posterior weighted F_1 score conditioned on hypothesis type (characteristic as solid line, defining as dashed line) as a function of data.

in the prior) are substantially more likely in explaining the data and thus come to dominate the posterior. Put simply, the characteristic-to-defining shift can be a by-product of data-driven learning.

There are two interesting implications/predictions of our model. First, our model predicts that the ideal learner will shift from characteristic to defining hypotheses even when she is capable of using abstraction from the outset of learning. This suggests that characteristic hypotheses may be useful, and that the observation that children accept and generate characteristic hypotheses at a young age does not preclude their ability to use abstraction or generate logical/defining hypotheses. Second, our model predicts that if there is a characteristic-to-defining shift, the relevant characteristic features should not capture a concept as well as defining features capture the concept; however, in order for a characteristic-to-defining shift to occur, the characteristic features must be informative to a certain degree. If characteristic features are completely uninformative, defining hypotheses should dominate across all amounts of data.

In our initial stab at the problem, we have made several simplifications. For one, the grammar generated hypotheses to be purely characteristic or purely defining. This simplification is reasonable given how adults would extend a kinship term. For example, if you meet a friend's family for the first time at a neighborhood BBQ, you would presumably extend the term *uncle* to their parent's male siblings and not the neighbors, who might share several characteristic features with your friend's uncles. This is not to say that competent adult speakers do not maintain characteristic information about kinship terms (e.g., grandmothers are typically nice, old ladies). In the same vein, our characteristic and defining features did not share the same formalism (i.e., feature matrices vs. graph-theoretical functions). A future version of the model should permit characteristic and defining primitives within the same hypothesis and possibly within the same formalism (e.g., a feature matrix containing both characteristic and defining features). This model should also be extended beyond the kinship domain. Lastly, the model is sensitive to the structure of the PCFG in determining the prior. Further research should characterise the robustness of the model to variation in the prior.

Conclusion

In summary, the widely observed characteristic-to-defining shift falls out naturally from a rational data-driven process. Our simulations show that a data-driven inference mechanism (1) demonstrates a characteristic-to-defining shift in the task of concept learning without positing a change in cognitive representations or processes and (2) succeeds at learning most kinship words from a data distribution based on natural language statistics. We find that an ideal learner will demonstrate a shift even when more accurate abstract/logical representations are possible from the onset of learning provided that characteristic features are informative but imperfect in their ability to capture the underlying concept. While we address the problem of concept learning within the kinship domain, the model framework can be extended to explain concept learning across multiple domains using different representational formalisms.

References

Benson, N. J., & Anglin, J. M. (1987). The child's knowledge of english kin terms. *First Language*, *7*(19), 41–66.

Bruner, J. S., Olver, R. R., Greenfield, P. M., et al. (1966). *Studies in cognitive growth.* Wiley.

Chi, M. T. (1985). Interactive roles of knowledge and strategies in the development of organized sorting and recall. *Thinking and learning skills*, *2*, 457–483.

Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive science*, *5*(2), 121–152.

Danziger, K. (1957). The child's understanding of kinship terms: A study in the development of relational concepts. *The Journal of genetic psychology*, *91*(2), 213–232.

Feldman, J. (2000). Minimization of boolean complexity in human concept learning. *Nature*, *407*(6804), 630–633.

Haviland, S. E., & Clark, E. V. (1974). this man's father is my father's son: a study of the acquisition of english kin terms. *Journal of Child Language*, *1*(01), 23–47.

Imai, M., Gentner, D., & Uchida, N. (1994). Children's theories of word meaning: The role of shape similarity in early acquisition. *Cognitive Development*, *9*(1), 45–75.

Katz, Y., Goodman, N. D., Kersting, K., Kemp, C., & Tenenbaum, J. B. (2008). Modeling semantic cognition as logical dimensionality reduction. In *Proceedings of thirtieth annual meeting of the cognitive science society.*

Keil, F. C. (1983). On the emergence of semantic and conceptual distinctions. *Journal of Experimental Psychology: General*, *112*(3), 357.

Keil, F. C., & Batterman, N. (1984). A characteristic-to-defining shift in the development of word meaning. *Journal of Verbal Learning and Verbal Behavior*, *23*(2), 221–236.

Kemler, D. G. (1983). Exploring and reexploring issues of integrality, perceptual sensitivity, and dimensional salience. *Journal of Experimental Child Psychology*, *36*(3), 365–379.

Kemp, C., & Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science*, *336*(6084), 1049–1054.

Kemp, C., & Tenenbaum, J. B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences*, *105*(31), 10687–10692.

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive development*, *3*(3), 299–321.

Marr, D. (1982). *Vision: A computational approach.* Freeman & Co., San Francisco.

Mollica, F., & Piantadosi, S. T. (2015). Towards semantically rich and recursive word learning models. In *Proceedings of the 37th annual meeting of the cognitive science society* (pp. 1607–1612).

Piaget, J., & Inhelder, B. (1969). *The psychology of the child.* Basic Books.

Piantadosi, S. T. (2014). *LOTlib: Learning and Inference in the Language of Thought.* available from https://github.com/piantado/LOTlib.

Shultz, T. R., Thivierge, J.-P., & Laurin, K. (2008). Acquisition of concepts with characteristic and defining features. *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, 531–536.

Weisleder, A., & Fernald, A. (2013). Talking to children matters early language experience strengthens processing and builds vocabulary. *Psychological Science*, *24*(11), 2143–2152.

Werner, H. (1948). *Comparative psychology of mental development.* Follett Pub. Co.

Xu, F. (2007). Rational statistical inference and cognitive development. *The innate mind: Foundations and the future*, *3*, 199–215.