Concept learning as motor program induction: A large-scale empirical study

Brenden M. Lake
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology

Ruslan Salakhutdinov
Department of Statistics
University of Toronto

Joshua B. Tenenbaum
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology

Abstract

Human concept learning is particularly impressive in two respects: the internal structure of concepts can be representationally rich, and yet the very same concepts can also be learned from just a few examples. Several decades of research have dramatically advanced our understanding of these two aspects of concepts. While the richness and speed of concept learning are most often studied in isolation, the power of human concepts may be best explained through their synthesis. This paper presents a large-scale empirical study of one-shot concept learning, suggesting that rich generative knowledge in the form of a motor program can be induced from just a single example of a novel concept. Participants were asked to draw novel handwritten characters given a reference form, and we recorded the motor data used for production. Multiple drawers of the same character not only produced visually similar drawings, but they also showed a striking correspondence in their strokes, as measured by their number, shape, order, and direction. This suggests that participants can infer a rich motor-based concept from a single example. We also show that the motor programs induced by individual subjects provide a powerful basis for one-shot classification, yielding far higher accuracy than state-of-the-art pattern recognition methods based on just the visual form.

Keywords: concept learning; one-shot learning; structured representations; program induction

The power of human thought derives from the power of our concepts. With the concept "car," we can classify or even imagine new instances, infer missing or occluded parts, parse an object into its main components (wheels, windows, etc.), reason about a familiar thing in an unfamiliar situation (a car underwater), and even create new compositions of concepts (a car-plane). These abilities to generalize flexibly, to go beyond the data given, suggest that human concepts must be representationally rich. Yet it is remarkable how little data is required to learn a new concept. From just one or a handful of examples, a child can learn a new word and use it appropriately (Carey & Bartlett, 1978; Markman, 1989; Bloom, 2000; Xu & Tenenbaum, 2007). Likewise, after seeing a single "Segway" or "iPad," an adult can grasp the meaning of the word, an ability called "one-shot learning." A central challenge is thus to explain these two remarkable capacities: what kinds of representations can support such flexible generalizations, and what kinds of learning mechanisms can acquire a new concept so quickly? The greater puzzle is putting them together: how can such flexible representations be learned from only one or a few examples?

Over the last couple of decades, the cognitive science of concepts has divided into different traditions, focused largely on either the richness of concepts or on learning from sparse data. In contrast to the simple representations popular in early cognitive models (e.g., prototypes; Rosch, Simpson, & Miller, 1976) or conventional machine learning (e.g., support vector machines), one tradition has worked to develop more structured representations that can generalize in deeper and more flexible ways. Concepts have been characterized in terms of "intuitive theories," which are mental explanations that underly a concept (e.g., Murphy & Medin, 1985), or "structural description" models, which are compositional representations based on parts and relations (e.g., Winston, 1975; Hummel & Biederman, 1992). In the latter framework, the concept "Segway" might be represented as two wheels *connected by* a platform, which *supports* a motor, etc. Most recently, research in AI and cognitive science has emphasized rich *generative* representations. Concepts like "house" can vary in both the number and configuration of their parts (windows, doors, balconies, etc.), much like the variable syntactic structure of language. This has lead researchers to model objects and scenes using generative grammars (Wang et al., 2006; Savova, Jakel, & Tenenbaum, 2009; Zhu, Chen, & Yuille, 2009) or programs (Stuhlmuller, Tenenbaum, & Goodman, 2010).

A different tradition has focused more on rapid learning and less on conceptual richness. People can acquire a concept from as little as one positive example, contrasting with early work in psychology and standard machine learning that has focused on learning from many positive and negative examples. Bayesian analyses have shown how one-shot learning can be explained with appropriately constrained hypothesis spaces and priors (Shepard, 1987; Tenenbaum & Griffiths, 2001), but where do these constraints come from? For simple prototype-based representations of concepts, rapid generalization can occur by just sharpening particular dimensions or features, as described in theories of attentional learning (Smith, Jones, Landau, Gershkoff-Stowe, & Samuelson, 2002) and overhypotheses in hierarchical Bayesian models (Kemp, Perfors, & Tenenbaum, 2007). From this perspective, prior experience with various object concepts may highlight the most relevant dimensions for whole classes of concepts, like the "shape bias" in learning object names (as opposed to a "color" or "material bias"). It is also possible to learn new features over the course of learning the concepts (Schyns, Goldstone, & Thibaut, 1998), and recent work has combined dimensional sharpening with sophisticated methods for feature learning (Salakhutdinov, Tenenbaum, & Torralba, 2011).

Despite these different avenues of progress, we are still far from a satisfying unified account. The models that explain how people learn to perform one-shot learning are restricted to the simplest prototype- or feature-based representations; they have not been developed for more sophisticated representations of concepts such as structural descriptions, grammars, or programs. There are also reasons to suspect that these richer representations would be difficult if not impossible to learn from very sparse data. In linguistics, for instance, grammar induction is typically studied in the limit as the number of examples goes to infinity; why should we expect learning a grammar that describes instances of houses, or cars, to be possible from just one example? Theoretical arguments (e.g., the bias/variance tradeoff; Geman, Bienenstock, & Doursat, 1992) imply that representationally rich concepts should generally require more data to learn, not less. The work of Winston (1975) and Lovett, Dehghani, and Forbus (2007) might be the closest to human-level concept learning, where they learned relational schemata for simplified notions of "arches," "houses," "stoves," and "fireplaces" from short sequences of examples. But a fully human-like, one-shot learning ability was beyond their scope.

Even with these gaps in our understanding, we believe that the power of human concepts will be best explained by bringing these two traditions together. By doing so, we hope to explore the extent to which people can learn representationally rich concepts from very sparse data, and we also hope to explain this ability in computational terms. These are the long-term goals of our work. Here we take a first step with a large-scale empirical study of one-shot concept learning, using a domain of handwritten characters from the world's alphabets (see Figure 1). These objects are not nearly as complex as many object concepts such as "house," "dog," or "Segway," but they still offer a vast number of novel, high-dimensional, and cognitively natural categories with important relational structure. They are much richer than the highly oversimplified artificial stimuli used in previous laboratory studies of one-shot learning (Feldman, 1997; Kemp & Jern, 2009). Yet characters are still simple enough for us to hope that tractable computational models can represent all the structure people see in them – unlike natural images.

What is the right structural representation for these simple visual concepts? The generative process for any handwritten character is a motor program, which is a set of instructions, in the mind of the drawer, that can be sent to the motor effectors such as an arm or a hand. These programs are complex compositions of pen strokes (the "parts" or the "subroutines" of the program), which might vary in their number, order, and style across drawers. Despite these various degrees of freedom, human drawing is noted for its regularity, which has been likened to a "grammar of action" (Goodnow & Levine, 1973). Thus it seems fruitful to explore a generative approach based on motor programs, especially since people have the generative capacity for drawing. There are also well-developed, feature-based alternatives from psychology (Grainger, Rey, & Dufau, 2008) and machine learning, especially "deep learning" models which have achieved some of the best results on handwritten digit classification (0, 1, 2, ..., 9) (e.g., Salakhutdinov & Hinton, 2009). Thus it will be important to compare multiple computational approaches, with the goal of better understanding the psychological processes and also improving one-shot learning in machines.

To begin exploring these questions, we ran a large-scale online study where participants drew novel character concepts after seeing just a single example. We refer to this task as "one-shot category production," drawing inspiration from numerous studies that have used the generation of category exemplars as a window into conceptual representation (e.g., Battig & Montague, 1969; Rosch et al., 1976; Feldman, 1997). We see one-shot category production as a special case of "one-shot learning," which includes classification and other types of generalization from just one example. Our large-scale study produced about 32,000 images of characters across a set of 1,600 concepts, and the on-line drawing trajectories were recorded for each image. From the production data, we analyzed the extent to which people can infer a robust motor program representation from a single example. We also compared humans and multiple computational approaches on a one-shot classification task, using methods based on either the motor data or just the visual forms.

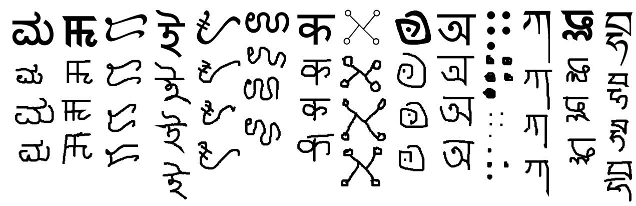

Figure 1: The top row shows example characters from our dataset, in the original printed form. Below are three example drawings from participants.

Category production experiment

The 1,600 character concepts were collected from 50 alphabets, including current or historic scripts (e.g., Bengali, Sanskrit, and Tagalog) and invented scripts for purposes like sci-fi novels. The characters were taken from www.omniglot.com in printed fonts, and several originals and their subsequently drawn images are shown in Fig. 1. This dataset was previously used to compare models of one-shot classification (Lake, Salakhutdinov, Gross, & Tenenbaum, 2011).

The drawing experiment was run through Amazon Mechanical Turk, and participants were asked to draw at least one entire alphabet. For each template image, they were asked to "draw each character as accurately as you can." An alphabet's printed characters were displayed in rows on a webpage, with an associated drawing pad below each image. Participants could draw by holding down a mouse button and moving the mouse, and we also included "forward," "back," and "clear" buttons. Some participants made minor image adjustments with small mouse movements, and we tried to mitigate this inconsistency by excluding strokes that were very short in time and space from the analysis.

The structure of the motor programs

When people perceive a new character, in what sense do they infer a new concept? While this mental representation might be just a bundle of features, the concept might also include richer structure in the space of motor programs. To investigate this possibility, we analyzed how multiple drawers produced a particular concept during the drawing task. We rea-

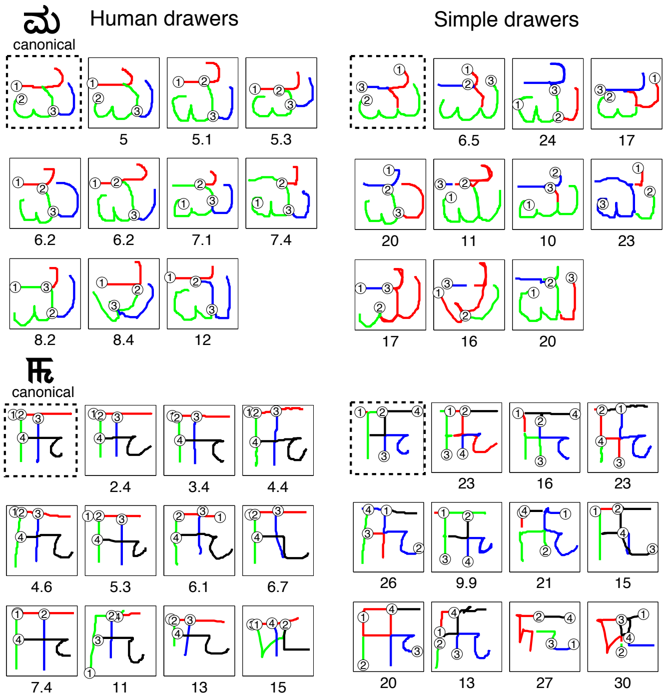

Figure 2: For two concepts (out of the 1600 total), each box shows the motor data produced by human drawers (left) or simple drawers (right). "Canonical" drawers are in the dotted boxes, and their distances (Eq. 1) to the other examples are the numbers below each frame. Stroke color shows correspondence to the canonical, circles indicate the beginning of each stroke, and numbers inside circles denote stroke order.

soned that in order to do this task, participants must infer a novel motor program, which will be reflected in the time course of drawing. Consistency in the structure of these drawings would provide evidence for two interlinked claims: people seem to grasp the same underlying concept from one example, and this concept includes a highly structured generative program. To measure consistency for a particular character, we quantitatively analyzed the number, shape, direction, and order of the parts (strokes) in the motor data.

The number of parts

This analysis (and subsequent ones) used just 20 of the alphabets in the dataset, excluding the six most common as determined by Google hits. The remaining alphabets were needed to train the alternative models in the later classification experiment. The simplest statistic to analyze was the number of parts. For each character, we investigated whether the drawers clustered around a common number of parts (the mode number across participants). Aggregating across each drawing in the dataset, the histogram in Fig. 3A (red) shows the absolute difference between the actual number of strokes and the mode number of strokes from all of the drawings of that character. Although this distribution is guaranteed to peak at zero, a strikingly large percentage of drawers used exactly the modal number (66%). As a control, a null dataset was created by replacing each number of strokes by a uniform draw (1 to 6 here, but other values are similar). This distribution was not nearly as peaked around the mode (Fig. 3A blue).

The shape of the parts

The parse of a character into parts (strokes) is at the core of each drawing. When people look at a new concept, do they perceive the same parts? This is difficult to analyze, since the number and length of the strokes can differ between images. A similarity measure should also be invariant to the order and direction of the strokes. Despite these challenges, we found that it was possible to analyze consistency in the shape of the strokes, and we discuss our method in the section below.

Shape-based distance in motor space. Since most drawers (66%) used the modal number of strokes, we restrict this and subsequent analyses to only these modal drawings. With this simplification, the strokes in two images can be matched in correspondence (one-to-one and onto). Our approach also matches the sub-structure within two strokes, finding an alignment between the points in the two trajectories (onto but not one-to-one). Given an optimal matching at both levels, the overall shape distance is roughly the mean distance between all of the aligned trajectory points. Before computing distance, characters were also transformed to be translation and scale invariant.[1] Examples of the distance are illustrated in Fig. 2, where the number below each drawing is the distance to the drawing in the dotted box.

The details of the distance measure are as follows. Consider two drawings $S_1, ..., S_k$ and $R_1, ..., R_k$ with k strokes each. Each stroke is a sequence of positions $S_i = [S_{i1}, ..., S_{in}]$ with arbitrary length, where $S_{ij} \in \mathbb{R}^2$. The overall distance between the characters is defined as

$$\min_{\pi} \frac{1}{k} \sum_{i=1}^{k} \min \left[dtw(S_i, R_{\pi(i)}),\ dtw(S_i, F(R_{\pi(i)})) \right], \quad (1)$$

where $\pi(\cdot)$ is a permutation on the stroke indices $1, ..., k$ (a bijective function from the set $\{1, ..., k\}$ to $\{1, ..., k\}$), and the flip function $F(S_i) = [S_{in}, ..., S_{i1}]$ reverses the stroke direction to provide direction invariance. The distance $dtw(\cdot, \cdot)$ between two trajectories is calculated by Dynamic Time Warping (DTW; Sakoe & Chiba, 1978), which fits a non-linear warp such that each point in one trajectory is aligned with a point in the other. The DTW distance is then the mean Euclidean distance across all pairs of aligned points.

The simple drawer model. Upon visual inspection of the stroke matches $\pi(\cdot)$ chosen by the outer minimization in Eq. 1, there is a striking consistency across drawers in the inferred parts for a character. We show two characters in Fig. 2, where color denotes the stroke matches (left panels). The plots for the entire dataset are available online.[2] While this qualitative correspondence may reflect richly structured motor processes, there could be a more simplistic explanation. The consistency could be a consequence of selection bias, since we selected drawers that used the modal number of strokes,

[1] This transformation subtracts the center of gravity and rescales, such that the range of the largest dimension is 105.

[2] http://web.mit.edu/brenden/www/consistency.html

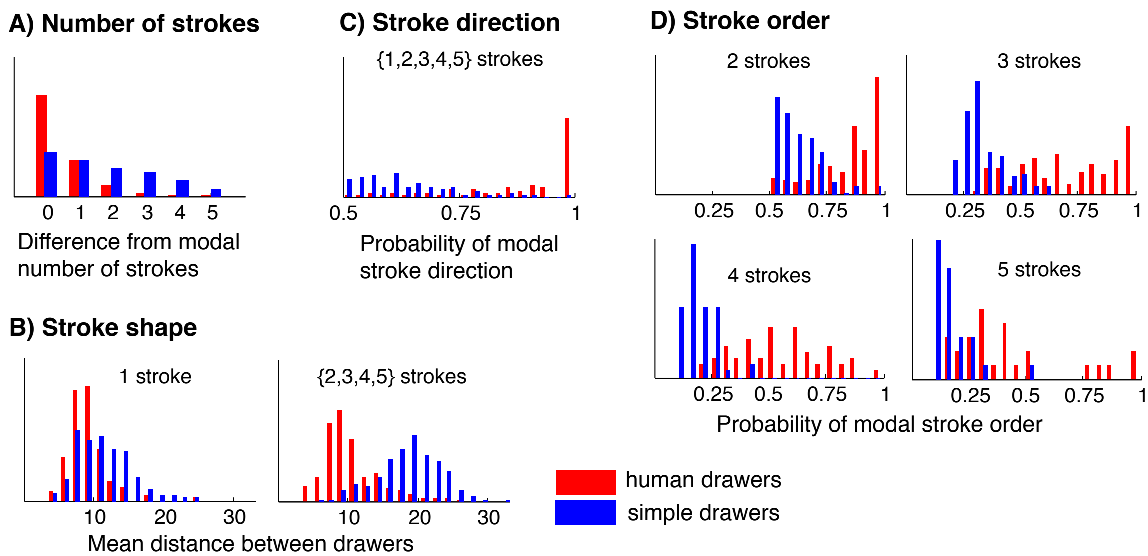

Figure 3: A histogram analysis of the consistency in the motor data, comparing human drawers (red) with a parallel dataset of simple drawers (blue) designed as a null hypothesis. Humans are strikingly consistent across a range of statistics compared to the simple model. As labeled, some histograms pool data from characters with different numbers of strokes (e.g., {2,3} includes 2- and 3-stroke characters).

and there will be fewer degrees of freedom available to a k-stroke drawer for any given k. In the special case of k disjoint segments (like in Braille), there may only be one production option. To explore the degrees of freedom and to provide a baseline for the observed consistency, we devised a "simple drawer" model that is likely to mimic human drawers when the space is highly constrained, but otherwise it more freely explores the potential motor space.

The simple drawer is given access to the same set of points a real drawer traversed in the motor data, but without the sequential information. It then tries to draw the same character as efficiently as possible using the same number of strokes. It must visit every point exactly once, while minimizing the distance traveled while ink is flowing. Given a real drawing with strokes $S_1, ..., S_k$, the simple drawer's interpretation $Q_1, ..., Q_k$ is defined by the problem

$$\underset{Q_1,...,Q_k}{\text{argmin}} \sum_{i=1}^{k} \sum_{j=1}^{|Q_i|-1} ||Q_{ij} - Q_{i(j+1)}||_2, \qquad (2)$$

where $|\cdot|$ is the number of points in the stroke sequene, and $||\cdot||_2$ is Euclidean distance. Each point S_{ij} in the original drawing is equal to exactly one point Q_{ab} in the new drawing. This formulation encourages smooth strokes, but it also leads to creative parses (Fig. 2 right panels), in part because there are multiple optima. A drawback of the model is that it sometimes draws paths where no ink exists. To reduce this problem, the simpler drawer is not allowed to travel large distances between adjacent points, where the upper bound is the maximal adjacent distance in the corresponding real drawing. For optimization, we can reformulate the problem as the well-known traveling salesman problem (TSP) by inserting k cost-free "points" to indicate the stroke breaks. Inspired by efficient approximate solvers for the TSP problem, we optimized using simulated annealing with alternating Metropolis-Hasting node swaps and Gibbs sampling (Rubinstein & Kroese, 2008).

Results. The simple drawer was used to re-sketch each image, creating an entire parallel dataset for comparative analysis. The shape-based consistency of a character is the mean distance (Eq. 1) between each pair of drawings of that character. Fig. 3B shows histograms of this consistency measure for the human drawers (red) and the simple drawers (blue). The aggregate histogram (right) for characters with two to five strokes shows a large difference in the consistency of the parts. The histogram for characters with one stroke (left) shows a closer correspondence between participants and the simple drawer, due to the limited degrees of freedom.[3] These results suggest that people inferred motor programs that were based on a characteristic set of strokes.

The direction of the parts

Do different drawers infer the same stroke directions? For each character, a single canonical drawer was chosen to minimize the sum shape-based distance across all other drawers of that character (Eq. 1). Example canonical drawers are shown in the dotted boxes in Fig. 2 (left). For each person's drawing compared to the canonical drawing, the chosen value of the inner minimization in Eq. 1 indicates whether each stroke, or that stroke in reverse direction ($F(\cdot)$), is a better match to the corresponding stroke in the canonical drawer. Aggregating across each stroke in the dataset, Fig. 3C (red) displays the proportion of times the modal stroke direction was picked, using the canonical drawer as the reference point. The dataset

[3] Some single stroke characters can still be drawn in a number of ways, such as choosing the starting location of an "O." People tend to start at the top, while the simpler drawer is agnostic.

of simple drawers (blue) provides a direction-agnostic baseline. By comparison, people's inferred programs clearly have preferred directions.

The order of the parts

Is stroke order also consistent across drawers? As in the analysis of direction, and the canonical drawers were used as the reference points, from which stroke order was defined. For any person's drawing compared to the canonical drawing of that character, the chosen permutation $\pi(\cdot)$ from the outer minimization in Eq. 1 defines a relative ordering of the strokes. Aggregating across each drawing, Fig. 3D shows the proportion of times the modal stroke order was picked. Like the other statistics, stroke order was also highly consistent across characters. Unsurprisingly, consistency was less pronounced as the number strokes increased.

One-shot classification

The previous analyses suggest that people can infer rich motor-based concepts from just a single example. If the same perceptual inferences occurred in the context of categorization, would these representations prove useful for one-shot classification? We investigated this question by using the motor data to classify characters by type, based on the shape-based distance measure (Eq. 1). The model received 20 random characters with just one example each. Test examples (2 per class) were classified as the best fitting category. All 20 categories used the same (modal) number of strokes. This classification task was repeated 20 times with different characters, and the mean percent correct is shown in Fig. 4.

We used several baselines for comparison. The simplest method picked the closest image in pixel space, using Euclidean distance. We also tested Deep Boltzmann Machines (DBMs; Salakhutdinov & Hinton, 2009) as a representative feature-based model. DBMs learn a hierarchy of distributed feature representations for the raw pixels, without using a priori knowledge about the geometry of images. DBMs have obtained state-of-the-art performance on handwritten digit recognition when trained with thousands of digits, and we pre-trained it on the 30 alphabets that were not used for classification (about 19,000 images). For one-shot classification, new items were represented in feature space and classified based on cosine similarity across all hidden layers (two with 1000 units each). We also tested a model that infers latent stroke-like parts from the raw images (Lake et al., 2011), as well as the simple drawing model, which uses the same motor data but without the strong structural consistency.

Performance was measured across a range of different numbers of strokes (Fig. 4). Chance performance is 5% correct, and pixel distance performed at 20% correct on average ("pixels" in Fig. 4). Next was the DBM at 37% ("features"), the inferred parts model at 48%, and then the simple drawer at 50%. The real stroke data was far better than all of the other methods, with an average performance of 83% correct. We also tried to include stroke order and direction information in the classification cost function, but performance did

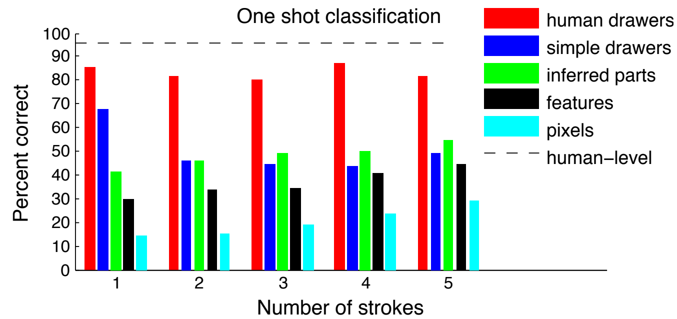

Figure 4: Classification performance based on one example of 20 different characters. Test instances were compared to each class, and the best match was selected.

not improve significantly. Finally, human one-shot classification performance was 96%, as measured behaviorally in a 20-way classification task ("human-level" in Fig. 4; see footnote for experimental setup).[4] Overall, the motor data was by far the most effective means for one-shot classification.

Discussion

Our category production experiment produced over 32,000 images of handwritten characters. Each of the roughly 1,600 characters was drawn by 20 different participants, and we found a strong correspondence in the structure of their inferred motor programs. On the whole, the number, shape, order, and direction of the parts (strokes) was highly consistent across participants. Also the motor data provided a powerful basis for one-shot classification. These results suggest that when people look at a new character, they can infer a richly structured motor program. This motor program is capable of both synthesizing new examples and classifying new instances with high discriminative accuracy.

How can these motor programs be learned from just one example? This ability clearly depends on prior experience, but how does this translate into constraints on the formation of these programs? There are various possibilities. Prior knowledge might come in the form of shared sub-programs or shared strokes, like our preliminary model in Lake et al. (2011). From their general writing and drawing experience, people might learn sub-routines like "circles, diagonal lines, or S-shapes," and then they could parse novel characters into this rich set of parts. But prior knowledge could come in many other forms, including more general constraints and biases (learned or otherwise) in human drawing and motor capabilities. Researchers have found a number of rules that usefully characterize drawing: start drawing at the top-left, draw horizontal strokes left-to-right, draw vertical strokes top-to-bottom, and minimize the number of strokes (Goodnow & Levine, 1973; Van Sommers, 1984). In a preliminary analysis, we have observed strong versions of these effects in our dataset of natural alphabets. Thus, it is possible that

[4]This study was run on Amazon Mechanical Turk with 15 participants and 50 trials. Each trial consisted of a single test image, and participants were asked to pick one of the 20 other images that looked the most similar. This was the same task that the models performed, except that characters with different numbers of strokes were intermixed and a different set of alphabets was used.

some of the richness of these newly acquired concepts (including shape, direction, and order) is a consequence of relatively simple, low-level principles. But it is also unclear how these directives should be combined when they conflict, or how they might interact with other forms of prior knowledge. Computational models are well-suited to help answer these questions, and we hope that future work will clarify how prior knowledge can support such rapid program induction.

Finally, although our work has focused on handwritten characters, we expect that similar phenomena and computational accounts are relevant more broadly. Characters share interesting structure with other kinds of symbols used for communication, including spoken words and gestures. Characters are produced by a sequence of strokes, and likewise, spoken words are produced as a sequence of phonemes. Characters, spoken words, and gestures are also "embodied," since the mind and body can both generate and perceive concepts in these domains (e.g., Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Freyd, 1983). All of these concepts must also be learnable from one or a few examples, in the context of efficient communication and social learning. One-shot program induction may also be possible in learning very different kinds of natural concepts, such as trees or ferns that have distinctive branching patterns and unique leaf shapes. One-shot learning could be possible here for a different reason: not because of the strong priors imposed by motor constraints or previous learning, but because a single example is highly complex and contains extensive repeated structure. We hope that future work will explore a fuller range of rich representations for concepts, while explaining how these same concepts can be learned from just one or a few examples.

Acknowledgements We gratefully acknowledge Jason Gross for developing the experimental interface and for collecting the data. We also thank John McCoy for helpful comments, and Dan Ellis for making his DTW code available.

References

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Jounral of Experimental Psychology Monograph*, *80*(3).

Bloom, P. (2000). *How Children Learn the Meanings of Words*. Cambridge, MA: MIT Press.

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development*, *15*, 17–29.

Feldman, J. (1997). The structure of perceptual categories. *Journal of Mathematical Psychology*, *41*, 145–170.

Freyd, J. (1983). Representing the dynamics of a static form. *Memory and Cognition*, *11*(4), 342–346.

Geman, S., Bienenstock, E., & Doursat, R. (1992). Neural Networks and the Bias/Variance Dilemma. *Neural Computation*, *4*, 1–58.

Goodnow, J. J., & Levine, R. A. (1973). "The Grammar of Action": Sequence and syntax in childen's copying. *Cognitive Psychology*, *4*(1), 82–98.

Grainger, J., Rey, A., & Dufau, S. (2008). Letter perception: from pixels to pandemonium. *Trends in Cognitive Sciences*, *12*(10), 381–387.

Hummel, J. E., & Biederman, I. (1992, July). Dynamic binding in a neural network for shape recognition. *Psychological Review*, *99*(3), 480–517.

Kemp, C., & Jern, A. (2009). Abstraction and relational learning. In *Advances in Neural Information Processing Systems 22.*

Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, *10*(3), 307–321.

Lake, B. M., Salakhutdinov, R., Gross, J., & Tenenbaum, J. B. (2011). One shot learning of simple visual concepts. In *Proceedings of the 33rd Annual Cognitive Science Conference.*

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431–461.

Lovett, A., Dehghani, M., & Forbus, K. (2007). Incremental Learning of Perceptual Categories for Open-Domain Sketch Recognition Kenneth Forbus Comparisons and Generalization. In *Proceedings of the International Joint Conference on Artificial Intelligence.*

Markman, E. M. (1989). *Categorization and Naming in Children*. Cambridge, MA: MIT Press.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, *92*(3), 289–316.

Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(4), 491–502.

Rubinstein, R. Y., & Kroese, D. P. (2008). *Simulation and the Monte Carlo method* (Second ed.). Hoboken, New Jersey: John Wiley & Sons.

Sakoe, H., & Chiba, S. (1978, February). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *26*(1), 43–49.

Salakhutdinov, R., & Hinton, G. E. (2009). Deep Boltzmann Machines. In *12th International Conference on Artificial Intelligence and Statistics (AISTATS).*

Salakhutdinov, R., Tenenbaum, J. B., & Torralba, A. (2011). Hierarchical deep models for one-shot learning. In *Neural Information Processing Systems (NIPS).*

Savova, V., Jakel, F., & Tenenbaum, J. B. (2009). Grammar-based object representations in a scene parsing task. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society.*

Schyns, P. G., Goldstone, R. L., & Thibaut, J.-P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences*, *21*, 1–54.

Shepard, R. N. (1987). Toward a Universal Law of Generalization for Psychological Science. *Science*, *237*(4820), 1317–1323.

Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object Name Learning Provides On-the-Job Training for Attention. *Psychological Science*, *13*, 13–19.

Stuhlmuller, A., Tenenbaum, J. B., & Goodman, N. D. (2010). Learning Structured Generative Concepts. In *Proceedings of the Thirty-Second Annual Conference of the Cognitive Science Society.*

Tenenbaum, J. B., & Griffiths, T. L. (2001, August). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences*, *24*(4), 629–40.

Van Sommers, P. (1984). *Drawing and Cognition*. Cambridge University Press.

Wang, W., Pollak, I., Wong, T.-S., Bouman, C., Harper, M., & Siskind, J. (2006, October). Hierarchical Stochastic Image Grammars for Classification and Segmentation. *IEEE Transactions on Image Processing*, *15*(10), 3033–3052.

Winston, P. H. (1975). Learning structural descriptions from examples. In P. H. Winston (Ed.), *The psychology of computer vision.* New York: McGraw-Hill.

Xu, F., & Tenenbaum, J. B. (2007). Word Learning as Bayesian Inference. *Psychological Review*, *114*(2), 245–272.

Zhu, L., Chen, Y., & Yuille, A. (2009, January). Unsupervised learning of Probabilistic Grammar-Markov Models for object categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*(1), 114–28.