The cluster representation of knowledge can appear as if knowledge is hierarchically structured.

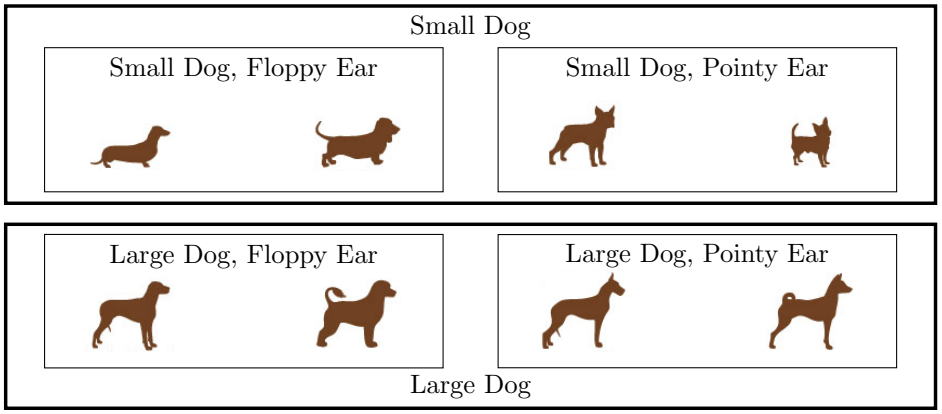
Takao Noguchi, Brad Love
Department of Experimental Psychology, University College London

Abstract

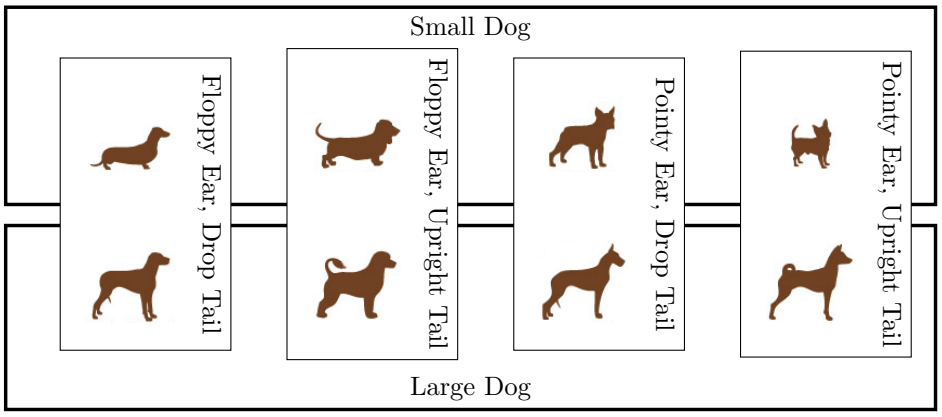
Environments surrounding people are often hierarchically structured: a single object usually fits into a series of progressively more general categories (e.g., a terrier, dog, and animal). Given the prevalence of this hierarchical structure, previous studies often assumed that knowledge representation is also hierarchical, although empirical evidence does not provide a clear support. In this contribution, we demonstrate that neither the environment nor knowledge representations need to be hierarchically structured for people to prefer the hierarchical structure: the hierarchical structure enables people to make the most accurate inference and should be preferred, even when knowledge representation is flat. This optimality is most pronounced when knowledge is represented with clusters of observed objects. Further, the cluster representation naturally leads to the effects of learning order, which explain mechanisms behind the basic level advantage and the age of acquisition effect.

Even informal observation of everyday categorization reveals that many objects fit into a number of categories. A single object might be called a terrier, dog, mammal or animal. This hierarchical structure of categories — a sequence of progressively larger categories — has been suggested as a universal property of category structure across cultures (Atran, 1998; Berlin, 1992). Given the wide-spread adaptation of the hierarchical structure, it is tempting to assume that knowledge is hierarchically represented. This assumption of hierarchical representation, indeed, often underlies theoretical propositions (e.g., Tenenbaum, Kemp, Griffiths, & Goodman, 2011). In this study, however, we demonstrate that the hierarchical structure should be preferred even when knowledge representation is flat, without the depth represented in the hierarchical structure.

Word count: 2276 (excluding the appendix).
Last updated: March 14, 2016.



(a) Hierarchical structure, where dogs share the maximum number of feature values at both levels, and categories at the specific level are nested within categories at the general level.



(b) Non-hierarchical structure, where dogs share the maximum number of feature values at each level, but categories at the specific level are orthogonal to categories at the general level.

Figure 1. Example structures with dogs. A rectangle encloses dogs in a category, with a thicker rectangle indicating a category at the general level.

Hierarchical structure of categories

To begin, let us illustrate the hierarchical structure. An example hierarchical structure is illustrated in Figure 1 (a). For brevity, we assume discrete features. Then formally, the hierarchical structure satisfies the two characteristics: maximization of feature information and the inclusion relation. First, the hierarchical structure maximizes the information which objects, X, provide about a category y:

$$I\left(y;\, X\right) = H\left(X\right) - H\left(X|\, y\right), \tag{1}$$

which is mutual information (denoted as I) between category labels and objects. This feature information is maximum when the conditional entropy, denoted as $H\left(X|\, y\right)$, is minimum. This is achieved when all the objects in a category share values on the largest possible number of features.

The hierarchical structure also satisfies the inclusion relation, that all the objects in one category at the specific level fit into the same category at the general level (see Figure 1 for examples).

Previous studies have demonstrated preference for the hierarchical structure. When asked to categorize objects in any manner they prefer, for example, people are very likely to provide hierarchical structure of categories (e.g., Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). When categorizing the eight dogs in Figure 1, for example, people are more likely to produce a structure similar to Figure 1 (a) than (b).

Given the wide-spread preference for hierarchical structure of categories, it has been argued that knowledge is hierarchically represented (e.g., Markman, 1989; Markman & Callanan, 1984). Markman (1989), for example, argues that people learn to hierarchically represent knowledge as they accumulate knowledge (see also Inhelder & Piaget, 1964; Vygotsky, 1962).

This assumption of hierarchical representation has been underlying theoretical propositions in cognitive science. For example, a mathematical model has been proposed to explain how people could learn to hierarchically represent knowledge (Kemp & Tenenbaum, 2008), and an inference drawn with this model correlates with inference people make (Kemp & Tenenbaum, 2009). The assumption of hierarchical representation also underlies theories in many other domains within cognitive science: for example, inference (Osherson, Smith, Wilkie, López, & Shafir, 1990), memory (Bower, Clark, Lesgold, & Winzenz, 1969; Glass & Holyoak, 1975), reasoning (Collins & Michalski, 1989; Shastri & Ajjanagadde, 1993), and word learning (Xu & Tenenbaum, 2007).

The assumption of hierarchical representation, however, is not well supported with empirical evidence (see Murphy & Lassaline, 1997, for review). Sloman (1998), for example, report that when reasoning and making inferences with the hierarchical structure, people often neglect the inclusion relation. Thus despite the preference for the hierarchical structure, an inference people make does not appear to follow from the hierarchical knowledge representation.

In this study, we step back and consider the possibility that preference for hierarchy is not an indication of hierarchical representation. In particular, we show that even when models of category learning do not assume the hierarchical structure, the hierarchical structure should still be preferred. Then, we demonstrate that these models also predict how

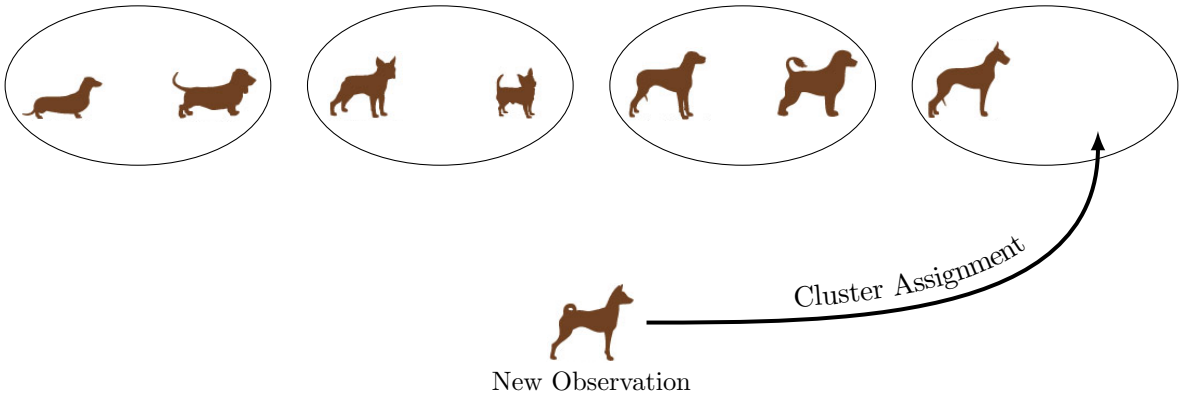


Figure 2. Example clusters. An ellipse encloses dogs in a cluster. A new observation is assigned to a cluster which contains most similar dogs.

early learning shapes knowledge representation and influences inference accuracy later in life.

Models of human cognition

Unlike the theoretical propositions discussed above, many cognitive models of category learning do not assume the hierarchical representation. Here, we discuss two of the dominant representations: cluster and exemplar representations. The cluster representation assumes that people represent knowledge as clusters of the observed objects (e.g., Anderson, 1991; Love, Medin, & Gureckis, 2004; Sanborn, Griffiths, & Navarro, 2010).

Every time a new object is observed, the new object is assigned to an existing cluster with similar objects (see Figure 2 for an example). If none of the existing clusters contains sufficiently similar objects, a new cluster is created to represent the new object. Then using the knowledge represented as clusters, an inference is drawn from the observed objects in the cluster. If a new object is assigned to a cluster predominantly with terriers, for example, the new object is considered to be a terrier.

Under the hierarchical structure as discussed above, the feature information is at the maximum, and objects in the same category share values on as many features as possible. Since objects which share values tend to be assigned to the same cluster, the cluster structure mirrors the category structure at one level. The cluster structure illustrated in Figure 2, for example, mirrors the category structure at the specific level in Figure 1 (a): each cluster contains dogs with the same ear (pointy or floppy) and of the same size (small or large). Then as the objects in the same cluster fit into the same category, an inference on the category labels tends to be more accurate.

In contrast, a non-hierarchical category structure cannot be mirrored in the cluster structure. With the structure in Figure 1 (b), for example, if the cluster structure mirrors the category structure at the specific level, objects in the same cluster fit into different categories at the general level, and then, an inference cannot be accurate at the general level. The only way to allow an accurate inference with the non-hierarchical category structure is to assign only one object to each cluster, which is often termed the exemplar

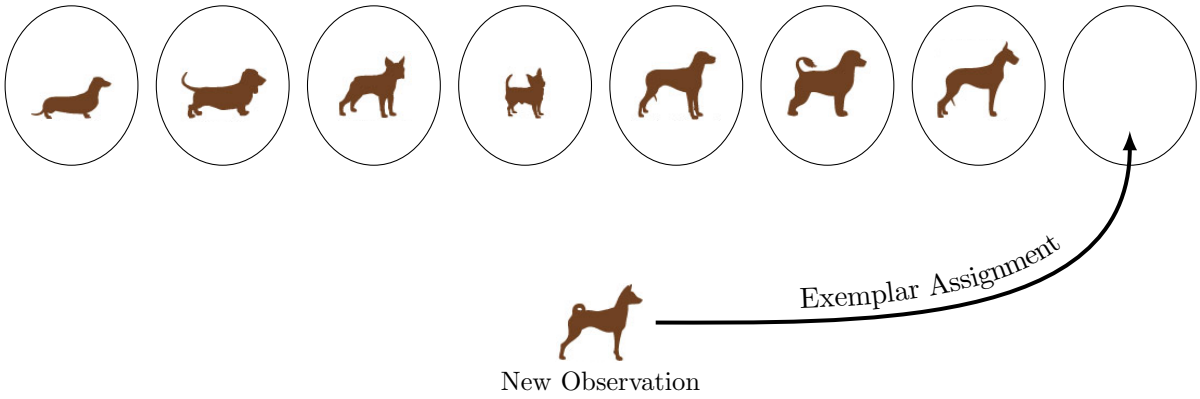


Figure 3. Exemplar representation, where each object is assigned to its own cluster.

representation (Nosofsky, 1986, 1991) (see Figure 3 for an illustration).

With the exemplar representation, each cluster contains a single object, and an inference is based on one of the most similar among the observed objects. If an object is most similar to an observed object labeled as a terrier, for example, the object is inferred to be a terrier. Thus, an inference is more accurate when similar objects fit into the same category, which corresponds to when the feature information is maximized. As the cluster structure cannot mirror the category structure, however, the exemplar representation is insensitive to the inclusion relation: whether objects in a category at the specific level fit into the same category at the general level[1][2].

Therefore, the hierarchical structure is expected to be optimal, such that a human learner is most likely to make an accurate inference across multiple levels of categories. This optimality, however, depends on knowledge representation. With the cluster representation, both the feature information and the inclusion relation improve the inference accuracy. With the exemplar representation, however, only the feature information improves the inference accuracy. To demonstrate this, we ran a simulation study reported in the next section.

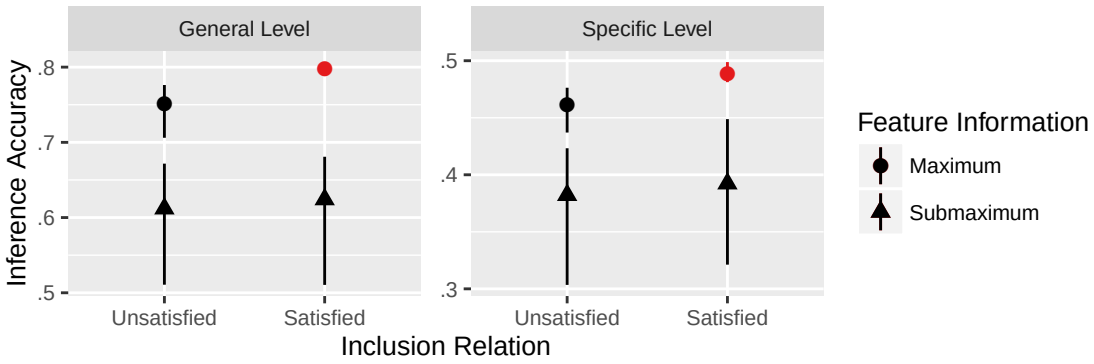
Preference for the hierarchical structure

We took the eight dogs with the three features in Figure 1 and tested all the possible category structures. For brevity, we constrained the category structures, such that the general level has two categories with four dogs each, and the specific level contains four categories with two dogs each.

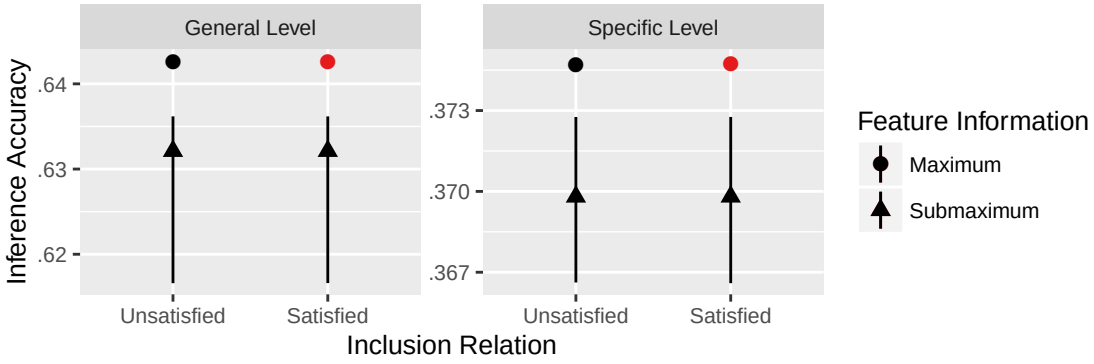
For each category structure, the model is trained for 10 blocks. Each block involves 16 trials, the eight objects with a category label at either general or specific level, in a random order. After the training blocks, the average inference accuracy was calculated for each level of categories. This accuracy was further mean-averaged across the 10^4 simulations for

[1]We assume that people do not infer a category label at another level than being asked: when making an inference on the general category, people do not infer the specific category (see also, Nosofsky, 2015).

[2]The insensitivity to the inclusion relation reflects people's neglect of the inclusion relation in making an inference, as discussed above. The insensitivity to the inclusion relation, however, can stem from the cluster representation, which we discuss in relation to influences of learning orders later in this article.



(a) Cluster representation. An inference is most accurate at both levels, when the feature information is maximum and the inclusion relation is satisfied.



(b) Exemplar representation. An inference is more accurate when the feature information is maximum but the accuracy is not affected by the inclusion relation.

Figure 4. The average inference accuracy for each category structure. The levels of categories are shown in columns: the left panels illustrate the general level, and the right panels illustrate the specific level. A dot represents mean, and error bar is the empirical interval. Red summarizes hierarchical structures, and black summarizes non-hierarchical structures.

each category structure and summarized in Figure 4 (Please refer to Appendix A for more details of the models and the simulation).

Figure 4 shows that with the cluster representation, the inference is more accurate with the feature information and the inclusion relation: The highest accuracy is achieved when the feature information is maximum and the inclusion relation is satisfied. With the exemplar representation, in contrast, the inference is more accurate only with the feature information. Here, the highest accuracy is achieved with the maximum feature information but the inclusion relation does not appear to have an impact.

Thus, the simulation results show that the hierarchical structure is optimal given human cognition. This optimality is more pronounced with the cluster representation than with the exemplar representation.

Advantage of certain level in the hierarchical structure

The above simulation proves that knowledge does not have to be hierarchically represented for the hierarchical structure to be preferred. This demonstration may appear in contradiction to previous findings which concern the hierarchical representation: in particular, the basic level advantage (Mervis & Rosch, 1981; Rosch et al., 1976) and the age of acquisition effects (Gerhand & Barry, 1998; Morrison & Ellis, 1995).

The basic level advantage documents that an inference people make is generally more accurate and faster at the basic level (e.g., dog) than at a more general (e.g., mammal) or more specific level (e.g., terrier). This basic level advantage has been explained with the assumption that knowledge is hierarchically represented.

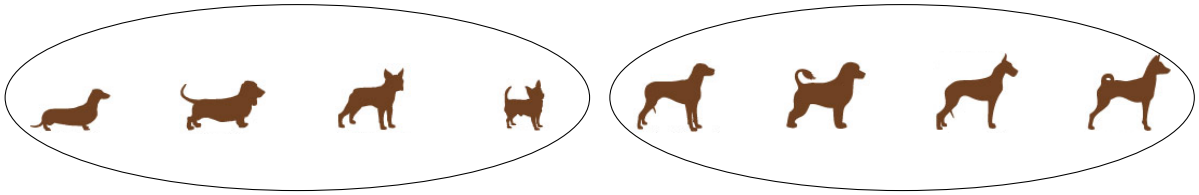
Jolicoeur, Gluck, and Kosslyn (1984), for example, argue that knowledge is accessed at the basic level at first and knowledge at other levels is subsequently accessed through the basic level. This proposition is not well supported by empirical studies. Clinical patients with semantic dementia, for example, can make an inference at the general level but cannot at the basic level (Hodges, Graham, & Patterson, 1995), suggesting that the patients can access to knowledge at the general level without accessing knowledge at the basic level. To resolve this contradiction, Rogers and Patterson (2007) argue that the basic level advantage can be observed even when knowledge at the general level is accessed at first.

Here again, we take a step back and consider the possibility that knowledge is not hierarchically represented. In this vein, key findings are from developmental studies, which report that categories at the basic level are learned earlier in the childhood (Berlin, Breedlove, & Raven, 1973; Brown, 1958; Horton & Markman, 1980; Mervis & Crisafi, 1982). Thus, the basic level advantage appears associated with learning order: an inference tends to be more accurate at a category level learned at first. Similarly, the age of acquisition effect documents that words which are acquired earlier in childhood are processed more accurately than words that are acquired later in life.

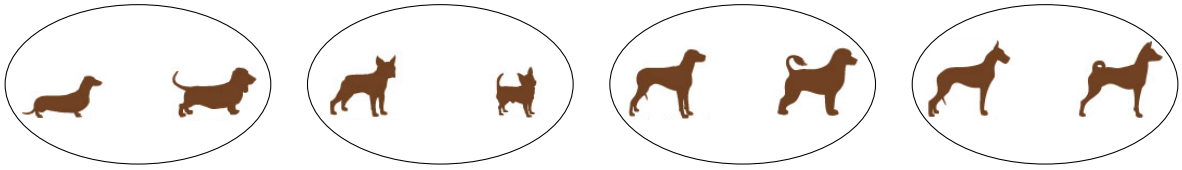
These effects of learning order are naturally borne out of the cluster representation. Figure 5 (a) illustrates a typical knowledge representation after learning the categories at the general level. As the cluster assignment of an object depends on its category label as well as its features, objects from the same category tend to be perceived similar and fit into the same cluster. As a result, the cluster structure tends to mirror the category structure learned at first.

Subsequent learning at another category level builds upon this initial cluster structure, and hence, the cluster structure cannot follow as closely the category structure learned later. Consequently, an inference tends to be more accurate at the category level learned at first. With the knowledge represented as in Figure 5 (a), for example, an inference at the general level may be accurate, but an inference at the specific level may not be as accurate. This is because each cluster contains dogs from multiple categories at the specific level.

When a cluster contains objects from multiple categories, an inference drawn with the cluster can appear to neglect the inclusion relation (e.g., Sloman, 1998). When knowledge about the dogs is represented as in Figure 5 (a), for example, all the dogs in one category at the general level can be accurately inferred to be small or large. This is because all the dogs in one cluster fit into the same category at the general level, and all the dogs in one cluster are of same size.



(a) When categories at the general level are learned first, the cluster structure mimics the category structure at the general level, allowing an accurate inference at the general level.



(b) When categories at the specific level are learned first, the cluster structure mimics the category structure at the specific level, allowing an accurate inference at the specific level.

Figure 5. Illustrative cluster structures for each learning order. An ellipse encloses dogs in a cluster.

When knowledge is represented as in Figure 5 (b), however, it becomes less clear whether all the dogs in one category at the general level are of same size. To infer this, the category of a dog has to be inferred at the general level first, and then, dogs from the same category at the general level have to be searched through all the clusters. This extra steps of inference and search may introduce error in inference, resulting in apparent neglect of the inclusion relation.

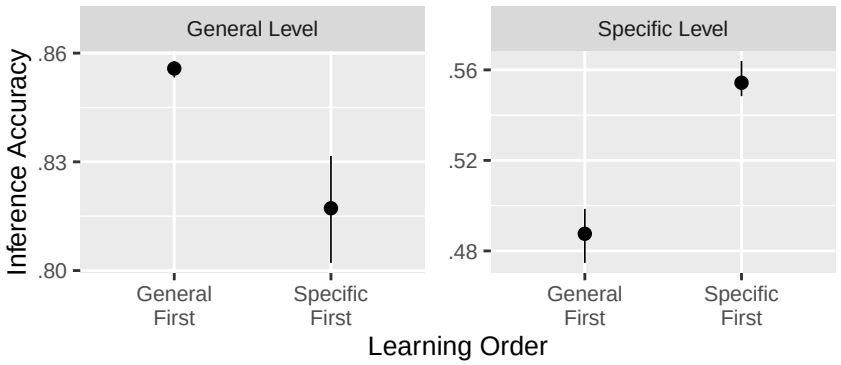
These influences of learning order make a contrast to the exemplar representation. With the exemplar representation, each object is assigned to its own cluster, regardless of learning order. Thus, the exemplar representation does not explain the influences of learning order. We demonstrate this contrast between the cluster and exemplar representations in the next section.

Learning order and knowledge representation

Here, we tested the hierarchical structures with two possible learning orders: general-first and specific-first. A training block in this simulation presents the eight objects in a random order with a category label at either general or specific level. Then, the learning order of general-first involves 10 training blocks with the general level first, followed by 10 training blocks with the specific level. This order is reversed for the learning order of specific-first.

Figure 6 illustrates the average inference accuracy for each learning order. This figure shows that with the cluster representation, an inference tends to be more accurate at the level learned at first. The left panel in Figure 6 (a), for example, shows that when the categories at the general level are learned at first, an inference is more accurate at the general level than when the same categories are learned later. With the exemplar representation, in contrast, the inference accuracy does not appear much influenced by the learning order.

These results show that with the cluster representation, learning is built upon the



(a) Cluster representation. An inference tends to be more accurate at the level learned at first.

(b) Exemplar representation. An inference tends to be more accurate when the specific level is learned at first, but the effect appears to be minor. The difference is less than .01 at both levels.

Figure 6. The average inference accuracy for each learning order. The levels of categories are shown in columns: the left panels illustrate the general level, and the right panels illustrate the specific level. A dot represents mean, and error bar is the empirical interval.

existing representation. The cluster representation tends to mirror the category structure learned at first, and this cluster structure carries over to the subsequent learning. As a result, an inference tends to be more accurate at the level learned first.

Conclusion

Although we do not argue against the possibility that knowledge is hierarchically represented, we have demonstrated that how people categorize objects may not necessarily correspond to the way knowledge is represented. In particular, we showed that the preference for the hierarchical structure is not necessarily due to the hierarchical representation. Even with the cluster representation, which does not distinguish general and specific levels, the hierarchical structure is optimal and should be preferred. The cluster representation also explains two major empirical findings in relation to the hierarchical structure of categories. Here, we have shown that the category structure learned at first shapes knowledge

representation and an inference accuracy later in life.

References

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*, 409–429.

Atran, S. (1998). Folk biology and the anthropology of science: cognitive universals and cultural particulars. *Behavioral and Brain Sciences*, *21*, 547–609.

Berlin, B. (1992). *Ethnobiological classification: principles of categorization of plants and animals in traditional societies*. New Jersey: Princeton University Press.

Berlin, B., Breedlove, D. E., & Raven, P. H. (1973). General principles of classification and nomenclature in folk biology. *American Anthropologist*, *75*, 214–242.

Bower, G. H., Clark, M. C., Lesgold, A. M., & Winzenz, D. (1969). Hierarchical retrieval schemes in recall of categorized word lists. *Journal of Verbal Learning and Verbal Behavior*, *8*, 323–343.

Brown, R. (1958). How shall a thing be called? *Psychological Review*, *65*, 14–21.

Collins, A. & Michalski, R. (1989). The logic of plausible reasoning: a core theory. *Cognitive Science*, *13*, 1–49.

Gerhand, S. & Barry, C. (1998). Word frequency effects in oral reading are not merely age-of-acquisition effects in disguise. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *24*, 267–283.

Glass, A. L. & Holyoak, K. J. (1975). Alternative conceptions of semantic theory. *Cognition*, *3*, 313–339.

Hodges, J. R., Graham, N., & Patterson, K. (1995). Charting the progression in semantic dementia: implications for the organisation of semantic memory. *Memory*, *3*, 463–495.

Horton, M. S. & Markman, E. M. (1980). Developmental differences in the acquisition of basic and superordinate categories. *Child Development*, *51*, 708–719.

Inhelder, B. & Piaget, J. (1964). *The early growth of logic in the child : classification and seriation*. London: Routledge and Kegan Paul.

Jolicoeur, P., Gluck, A., & Kosslyn, S. M. (1984). Pictures and names: making the connection. *Cognitive Psychology*, *16*, 243–275.

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

Kemp, C. & Tenenbaum, J. B. (2009). Structured statistical models of inductive reasoning. *Psychological Review*, *116*, 20–58.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, *111*, 309–332.

Markman, E. M. (1989). *Categorization and naming in children*. Massachusetts: MIT Press.

Markman, E. M. & Callanan, M. A. (1984). An analysis of hierarchical classification. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 2, pp. 325–365). New Jersey: Lawrence Erlbaum Associates, Inc.

Mervis, C. B. & Crisafi, M. A. (1982). Order of acquisition of subordinate-, basic-, and superordinate-level categories. *Child Development*, *53*, 258–266.

Mervis, C. B. & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology*, *32*, 89–115.

Morrison, C. M. & Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *21*, 116–133.

Murphy, G. L. & Lassaline, M. E. (1997). Hierarchical structure in concepts and the basic level of categorization. In K. Lamberts & D. Shanks (Eds.), *Knowledge, concepts and categories* (pp. 93–132). East Sussex: Psychology Press.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.

Nosofsky, R. M. (1991). Relation between the rational model and the context model of categorization. *Psychological Science*, *2*, 416–421.

Nosofsky, R. M. (2015). An exemplar-model account of feature inference from uncertain categorizations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 1929–1941.

Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological Review*, *97*, 185–200.

Rogers, T. T. & Patterson, K. (2007). Object categorization: reversals and explanations of the basic-level advantage. *Journal of Experimental Psychology: General*, *136*, 451–469.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439.

Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2010). Rational approximations to rational models: alternative algorithms for category learning. *Psychological Review*, *117*, 1144–1167.

Shastri, L. & Ajjanagadde, V. (1993). From simple associations to systematic reasoning: a connectionist representation of rules, variables and dynamic bindings using temporal synchrony. *Behavioral and Brain Sciences*, *16*, 417–451.

Sloman, S. A. (1998). Categorical inference is not a tree: the myth of inheritance hierarchies. *Cognitive Psychology*, *35*, 1–33.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: statistics, structure, and abstraction. *Science*, *331*, 1279–1285.

Vygotsky, L. (1962). *Thought and language*. Massachusetts: MIT Prss.

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

Appendix
Details of simulation

Here, we describe the model of category learning, which we used in the simulation. The model was proposed by Anderson (1991) and subsequently extended by Sanborn et al. (2010). Here, The exemplar and cluster representations are simulated with the same model but with different parameter values.

Suppose a learner has observed $n-1$ objects $\{x_1, x_2, \ldots, x_{n-1}\}$ with corresponding category labels $\{y_1, y_2, \ldots, y_{n-1}\}$. Each of these objects fits into a cluster. The cluster label for the ith object is denoted as z_i.

Drawing an inference

Then, the probability that the nth object fits into category w is expressed as follows:

$$\begin{aligned} p(y_n = w \mid x_n) &= \sum_{k \in \mathbb{Z}} p(z_n = k \mid x_n)\, p(y_n = w \mid z_n = k) \\ &= \sum_{k \in \mathbb{Z}} \frac{p(z_n = k)\, p(x_n \mid z_n = k)}{p(x_n)}\, p(y_n = w \mid z_n = k) \\ &= \sum_{k \in \mathbb{Z}} \frac{p(z_n = k)\, p(x_n \mid z_n = k)}{\sum_{s \in \mathbb{Z}}\ p(z_n = s)\, p(x_n \mid z_n = s)}\, p(y_n = w \mid z_n = k). \end{aligned} \tag{2}$$

Here, $\mathbb{Z}$ is a set of all the possible clusters to which the nth object can be assigned. The three terms in Equation 2 are described below in turn.

First, the probability that the nth object fits into cluster k is given by:

$$p(z_n = k) = \begin{cases} \dfrac{c\, m_k}{(1-c) + c\,(n-1)} & \text{if} \quad m_k > 0 \\ \\ \dfrac{(1-c)}{(1-c) + c\,(n-1)} & \text{if} \quad m_k = 0 \end{cases} \tag{3}$$

where c is a parameter called the coupling probability and m_k is the number of objects already assigned to cluster k.

Following Anderson (1991) and Sanborn et al. (2010), we assume that an object has independent dimensions given cluster. Therefore,

$$p(x_n \mid z_n = k) = \prod_{d \in D} p(x_{n,d} \mid z_n = k), \tag{4}$$

where D is a set of dimensions in which an object is described. The above term is computed with

$$p(x_{n,d} = v \mid z = k) = \frac{B_{v,d} + \beta_c}{m_k + J_d\, \beta_c}, \tag{5}$$

where $B_{v,d}$ is the number of objects in cluster k with value of v on dimension d, and J_d is the number of values which an object can take on dimension d. Sensitivity parameter β_c determines knowledge representation, as discussed below.

Similarly, the probability that the nth object has category label w, given a cluster, is given by:

$$p(y_n = w \mid z = k) = \frac{B_w + \beta_l}{B_{\cdot} + J\, \beta_l}, \tag{6}$$

where B_w is the number of observed objects with category label w in cluster k, $B_{\cdot}$ is the number of object in cluster k, and J is the number of category labels. Also, sensitivity parameter β_l determines knowledge representation, as discussed below.

Learning

The learning is equivalent to assigning an object into a cluster. The probability that an object is assigned to cluster k is computed as

$$p(z_n = k \mid x_n, y_n) \propto p(z_n = k)\, p(x_n \mid z_n = k)\, p(y_n \mid z_n = k). \tag{7}$$

This is computed with Equations 3, 4 and 6. Additionally, this cluster assignment is conducted using the sequential Monte Carlo with one particle, which produces behavior much like human (Sanborn et al., 2010).

Knowledge representation

With the above specification, knowledge representation is determined by values for sensitivity parameter β. For the cluster representation, we used $\beta_f = 1.0$ and $\beta_l = 0.5$. These values are among the best fitting values to human performance (Sanborn et al., 2010).

For the exemplar representation, we used small values of β ($\beta_f = 0.001$ and $\beta_l = 0.001$) during the training, ensuring that a cluster can only contain identical objects. Inference based on such small β is, however, often deterministic, and to allow more probabilistic inference, we used a larger value for β ($\beta_f = 1.0$ and $\beta_l = 5.0$) during the testing.

For both representations, the coupling probability c is at 0.5 during the training blocks and at 1.0 during the testing blocks. These coupling probabilities prevent a new cluster from being created during the testing and ensure that an inference is not made with random guessing.