Learning hypothesis spaces and dimensions through concept learning

Joseph L. Austerweil (Joseph.Austerweil@gmail.com)
Thomas L. Griffiths (Tom_Griffiths@berkeley.edu)
Department of Psychology, University of California, Berkeley, Berkeley, CA 94720-1650 USA

Abstract

Generalizing a property from a set of objects to a new object is a fundamental problem faced by the human cognitive system, and a long-standing topic of investigation in psychology. Classic analyses suggest that the probability with which people generalize a property from one stimulus to another depends on the distance between those stimuli in psychological space. This raises the question of how people identify an appropriate metric for determining the distance between novel stimuli. In particular, how do people determine if two dimensions should be treated as separable, with distance measured along each dimension independently (as in an L_1 metric), or integral, supporting Euclidean distance (as in an L_2 metric)? We build on an existing Bayesian model of generalization to show that learning a metric can be formalized as a problem of learning a hypothesis space for generalization, and that both ideal and human learners can learn appropriate hypothesis spaces for a novel domain by learning concepts expressed in that domain.

Keywords: generalization; categorization; Bayesian modeling; similarity; integral and separable dimensions

Introduction

Almost every two objects, events, or situations (or the sensory data for the *same* object at two different moments) that we encounter are unique. Despite this fact, when people (and animals) learn that one stimulus has a property, they reliably and systematically believe certain other stimuli have that property and others do not (Shepard, 1987). For example, if you learn a dark, large circle is a *gnarble*, how likely is a dark, slightly smaller circle or a dark very small circle to be a *gnarble*? This is the problem of *generalization*, which is pervasive across cognitive science. It occurs in many forms from higher-level cognition (e.g., concept learning, Tenenbaum, 2000) to linguistics (e.g., word learning, Xu & Tenenbaum, 2007) to perception (e.g., color categorization, Kay & McDaniel, 1978). How should an ideal learner generalize a property from a group of stimuli observed to have the property to other stimuli?

One of the most celebrated theoretical results of cognitive psychology provides a deceptively simple answer to this question, indicating that we should generalize a property from one object to another object when the two objects are similar, or equivalently, close in some psychological space (Shepard, 1987). However, this establishes a new problem: How should the distance between objects be measured? More formally, the problem is one of identifying a *metric* on a space, a basic challenge that also arises when using machine learning methods that rely on computing distances, such as nearest-neighbor classification (Xing, Ng, Jordan, & Russell, 2002; Davis, Kulis, Jain, Sra, & Dhillon, 2007). Cognitive psychologists have determined that people use two different kinds of metrics when forming generalizations about multidimensional stimuli: *separable* dimensions are associated with "city-block" distance or the L_1 metric, while *integral* dimensions are associated with Euclidean distance or the L_2 metric (Garner, 1974). These different metrics also have consequences beyond generalization behavior, influencing how people categorize objects varying along different dimensions (Handel & Imai, 1972) and whether people can selectively attend to each dimension (Garner & Felfoldy, 1970).

Analyses of human generalization have tended to treat the metric as a fixed property of stimuli. However, determining the appropriate metric on a psychological space is an important step towards developing an appropriate representation for the properties of novel objects. If two dimensions are separable, then those dimensions form privileged axes for representing locations in the psychological space, and it is easier to learn categories defined by rules that align with those axes (Kruschke, 1993). This is qualitatively different from an integral representation, in which there are no natural axes for representing the space. Identifying whether dimensions should be separable or integral is thus just as basic a step towards forming a representation for a novel domain as determining the number of dimensions, or the locations of each stimulus in the resulting space.

In this paper, we consider how a learner could identify the appropriate metric for representing a novel domain, comparing an ideal Bayesian learner with human judgments. The starting point for this investigation is an existing Bayesian model of generalization, introduced by Shepard (1987) and extended by Tenenbaum and Griffiths (2001). In this model, the property of interest is possessed by all stimuli within an unknown region of psychological space, and the probability of generalizing to a new stimulus is computed by summing over all candidate regions containing the new stimulus and the previous stimuli observed to have some property, weighted by the posterior probability of that region. The difference between separable and integral dimensions emerges as the result of probabilistic inference with different hypothesis spaces of regions (Shepard, 1987, 1991; Davidenko & Tenenbaum, 2001). The hypothesis spaces that produce generalization corresponding to separable and integral dimensions consist of axis-aligned and axis-indifferent regions in the space, respectively (see Figure 1). Axis-aligned regions produce stronger generalization along the axes, while axis-indifferent regions produce generalization that depends only on the Euclidean distance between stimuli.

This analysis of separable and integral dimensions lays the groundwork for our account of how people learn an appropriate metric for a novel space. Learning a metric thus becomes a matter of inferring an appropriate hypothesis space on which to base generalization. We define a hierarchical

Bayesian model that makes this inference from a set of observed concepts. We demonstrate that this model infers a city-block or Euclidean generalization metric when given axis-aligned or axis-indifferent concepts, respectively, and that people infer a hypothesis space for generalization based on the concepts they learn in a way that is consistent with this ideal observer analysis. This extends previous results by Goldstone (1994) who changed dimensions from being integral to separable via repeated training of a single concept.

The plan of the paper is as follows. The next section provides the theoretical background for our approach, summarizing the basic generalization model, revisiting some of the literature on separable and integral dimensions, and laying out our approach to hypothesis space learning. We then present a test of the predictions of this model with human learners. Finally, we conclude the paper with a discussion of our results and possible future directions.

Theoretical Framework

Our theoretical framework builds directly on the Bayesian generalization model introduced in Shepard (1987) and Tenenbaum and Griffiths (2001), so we begin by summarizing the key ideas behind this approach. We then show how this approach produces separable and integral generalization, and how it can be extended to allow an ideal learner to infer an appropriate representation for novel stimuli.

The Bayesian Generalization Model

Let $\mathcal{X}$ be the stimulus space and $\mathcal{H}$ be the hypothesis space, where $h \in \mathcal{H}$ is a hypothesis as to which objects have and do not have the property of interest (i.e., a hypothesis is a set of $x \in \mathcal{X}$). After observing that a set of stimuli $X = \{x_1, \ldots, x_n\}, x_i \in \mathcal{X}$, stimuli have some property, how should you update your belief in: (1) which property it is and (2) which other stimuli have that property? Assuming that stimuli are generated uniformly and independently under the true hypothesis at random for the property ($p(X|h) = \prod_i p(x_i|h) = |h|^{-n}$ for a hypothesis containing all stimuli in the given set; $p(X|h) = 0$ otherwise) and taking some prior over hypotheses $p(h)$, the posterior probability that a hypothesis h is the property that n given stimuli share is

$$p(h|X) = \frac{p(h)\prod_{i=1}^{n} p(x_i|h)}{\sum_{h' \in \mathcal{H}} p(h')\prod_{i=1}^{n} p(x_i|h')} \quad (1)$$

which is simply Bayes' rule. Using Equation 1, we can derive the probability of generalizing from X to some other stimulus y as the sum over the posterior probability of hypotheses containing y

$$p(y|X) = \sum_{h:y \in h} P(h|X) \quad (2)$$

which constitutes a form of *hypothesis averaging* (Robert, 2007). The predictions of the model depends intimately on the nature of the hypotheses under consideration, with different hypothesis spaces leading to different generalization patterns.

Separable and Integral Dimensions

Psychological explorations of human similarity metrics of multidimensional stimuli discovered two different ways in which people use these dimensions: separable and integral (Shepard, 1987). Separable dimensions can be interpreted independently and form natural axes for representing a space, while integral dimensions are difficult to perceive independently. The dimensional structure of stimuli affects many aspects of human information processing, including the ease of categorizing objects into groups and perceived distance between objects (Garner, 1974). For example, Garner and Felfoldy (1970) found that categorization time was facilitated for objects with integral dimensions (e.g., saturation and lightness of a color) into groups where the values of the dimensions of the objects in each group are correlated (light and desaturated vs. dark and saturated). However, there was interference for objects categorized into groups of objects where the values of the dimensions are orthogonal (light and satured vs. dark and desaturated). Conversely, there were no major differences in categorization time for these types of categorization structures when the dimensions were separable.

Dimensional structure also affects the perceived distances between objects (Shepard, 1991). The perceived distance metric for objects with separable dimensions is the "city-block" distance, also known as the L_1 metric, with the distance between two stimuli x_i and x_j being $d(x_i, x_j) = \sum_k |x_{ik} - x_{jk}|$, where k ranges over dimensions and x_k is the value of stimulus x on dimension k. The perceived distance metric for objects with integral dimensions is the Euclidean distance, or L_2 metric, with $d(x_i, x_j) = \sqrt{\sum_k (x_{ik} - x_{jk})^2}$. The use of these different distance metrics is consistent with the different properties of separable and integral dimensions: city-block distance sums the distance along each axis separately for all points in the space, while Euclidean distance is insensitive to whether a point is located along an axis, and is thus invariant to changes in the axes used to represent the space. Recent extensions of classic multidimensional scaling techniques bear out these results, and provide a way to identify whether people seem to use separable or integral dimensions in their representation of a set of stimuli (Lee, 2008).

In the Bayesian generalization model introduced in the previous section, the difference between integral and separable dimensions emerges from using two different hypothesis spaces (Shepard, 1987). Using a hypothesis space in which regions are aligned with the axes results in behavior consistent with separable dimensions, while a hypothesis space in which regions are indifferent to the axes results in behavior consistent with integral dimensions. Figure 1 shows a schematic of two such hypothesis spaces, restricted to rectangular regions in two dimensions, together with the generalization gradient for a single exemplar concept in each space.[1]

[1] We calculated the generalization gradients by sampling from the prior distribution over hypotheses for the axis-aligned and axis-indifferent hypothesis spaces, then weighting each hypothesis by the likelihood given the single exemplar $E5$. The gradients were evaluated on a discretized 9×9 grid.

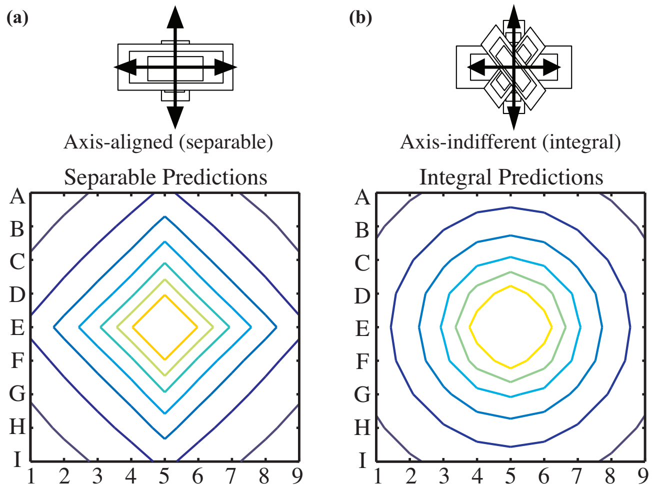

Figure 1: Hypothesis spaces and generalization gradients. (a) Axis-aligned (separable) and axis-indifferent (integral) hypothesis spaces. (b) Resulting generalization gradients for each hypothesis space given a single exemplar of a concept.

The generalization gradient resulting from the axis-aligned hypothesis space given a single exemplar of a concept decreases with distance under a city-block metric, while the gradient resulting from the axis-indifferent hypothesis space decreases with Euclidean distance. Models using the appropriate hypothesis spaces capture generalization judgments well for concept learning tasks using separable and integral dimensions for both single and multiple exemplars (Davidenko & Tenenbaum, 2001; Tenenbaum, 1999).

Learning a Hypothesis Space

The Bayesian generalization framework naturally extends to learning an appropriate hypothesis space by introducing the hypothesis space itself as a higher-level random variable in a hierarchical Bayesian model. Given an enumerable set of hypothesis spaces $\mathcal{M} = \{\mathcal{H}_1, \ldots, \mathcal{H}_M\}$, the probability that an ideal observer generalizes to a new stimulus y given a set of stimuli X have a property and a set of previously observed concepts $\mathcal{C}$ (where each concept itself is a set of stimuli) is

$$P(y|X,\mathcal{C}) = \sum_{m=1}^{M} P(y|\mathcal{H}_m, X)P(\mathcal{H}_m|\mathcal{C}, X) \qquad (3)$$

where the first term is the probability of generalizing from X to y under hypothesis space $\mathcal{H}_m$ (as specified by Equation 2), and the second term is the posterior probability of hypothesis space $\mathcal{H}_m$ given the previous concepts $\mathcal{C}$ and the observed stimuli of the current concept of interest. This posterior probability can be computed by applying Bayes' rule

$$P(\mathcal{H}_m|\mathcal{C}, X) = \frac{P(\mathcal{C}, X|\mathcal{H}_m)P(\mathcal{H}_m)}{\sum_{m=1}^{M} P(\mathcal{C}, X|\mathcal{H}_m)P(\mathcal{H}_m)} \qquad (4)$$

where $P(\mathcal{C}, X|\mathcal{H}_m)$ is the probability of observing a set of concepts $\mathcal{C}$ and the currently observed stimuli under hypothesis space $\mathcal{H}_m$ and $P(\mathcal{H}_m)$ is the prior probability of hypothesis space $\mathcal{H}_m$. The probability of concepts $\mathcal{C}$ and current stimuli X under hypothesis space $\mathcal{H}_m$ is

$$P(\mathcal{C}, X|\mathcal{H}_m) = \prod_{C \in (\mathcal{C} \cup X)} \sum_{h \in \mathcal{H}_m} P(h|\mathcal{H}_m) \prod_{x \in C} P(x|h) \qquad (5)$$

where C plays the same role as X, but for the previously observed concepts.

Intuitively, the model can be thought as being composed of m Bayesian generalization "submodels" (each with their own hypothesis space). The model's generalization judgments are made by averaging over the generalizations made by the individual submodels (given the current stimulus X) weighted by how well the submodel explains the previously and currently observed stimuli. Thus, the model "learns" to use hypothesis spaces that explain the observed concepts well.

Human Learning of Hypothesis Spaces

The model presented in the previous section predicts that a learner should be able to infer whether dimensions are integral or separable for a novel domain after seeing some examples of concepts expressed in that domain. Preliminary support for this idea is provided by the results of Goldstone (1994), who showed that teaching people a novel axis-aligned concept could affect generalization along that axis in both integral and separable spaces. However, shifting a representation all the way towards integral or separable dimensions will require learning more than one concept. To test whether human learners behaved in this way, we conducted an experiment in which we examined how the generalization judgments that people produce depend on the concepts they have learned. We used rectangles varying in width and height as our set of stimuli, and participants learned 20 concepts that were either aligned with or orthogonal to these dimensions (rectangles with the same aspect ratio or area). The key prediction was that participants observing axis-aligned concepts should show a generalization gradient consistent with a city-block metric, whereas participants observing concepts indifferent to these axes should show a generalization gradient consistent with a Euclidean metric. This prediction results from the different hypothesis spaces the two groups of participants should infer are appropriate for these domains.

Stimuli and Methods

The stimuli for this experiment were rectangles where the two manipulated dimensions were the width and height (ranging from 13 to 115 pixels in increments of approximately 25 pixels). The stimulus set is shown in Figure 2. We chose rectangles because it is easy to think of concepts on our two manipulated dimensions (same width or height) and the diagonals of the dimensions (same aspect ratio or area). Previously, Krantz and Tversky (1975) found people weakly favor using area and aspect ratio as separable dimensions (the diagonals of separable dimension space). However, people can use any of the four potential dimensions for generalization depending on the context rectangles are in. This natural flexibility

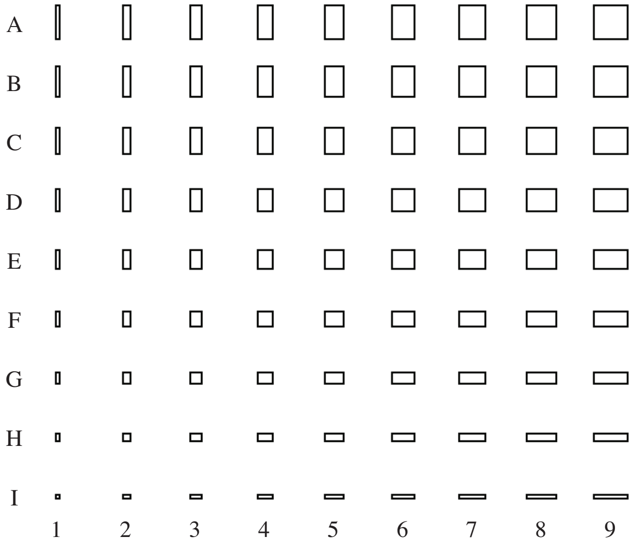

Figure 2: Stimuli used in our experiment (not to scale).

makes rectangles an ideal candidate for training participants to represent rectangles using different dimensional structures.

There were two phases to the experiment: training and test. For the training phase, there were two between-subjects conditions: the *separable* condition ($n = 15$), in which people observed axis-aligned concepts, and the *integral* condition ($n = 18$)[2] in which people observed axis-indifferent concepts. The test phase was the same for all participants. The cover story for the experiment was:

> On a small island in the Pacific Ocean, scientists found the ancient ruins of a small civilization. While excavating the ruins, they discovered objects on the doors of particular houses. They believe that the objects carry information about the people in the houses. Some of the objects the scientists found had names written under them.

Stimuli were then presented as objects with names, and people guessed what other objects would share the same name.

The 20 concepts shown to the training groups are shown in Figure 3 (each concept is a straight line picking out several points, corresponding to stimuli). The concepts for the two conditions were chosen such that each condition saw each object an equal number of times, there were two to four objects in each concept, and the concepts spanned the space of objects. The 20 concepts were presented to participants in a random order as examples of objects that were called different nonsense names randomly chosen from a standardized list. While the objects in each concept were on the screen, participants were asked whether or not they thought every object in $\{A,C,E,G,I\} \times \{1,3,5,7,9\}$ shown individually below the objects in the concept could be called that name.

The test phase of the experiment was identical to the first phase except participants' generalizations were tested for

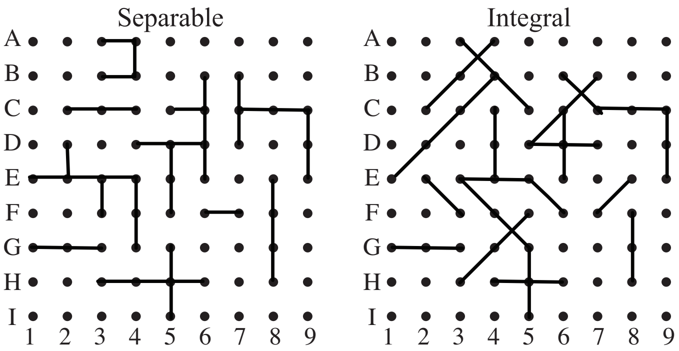

Figure 3: The 20 concepts for each training condition. Each concept is the collection of objects on a straight line on the grid. The separable concepts are axis aligned and the integral concepts are indifferent to axes.

concepts consisting of single objects ($\{B2, B8, E5, H2, H8\}$ were tested) over the total 9×9 set of objects.

Results

Figure 4 shows averaged results for single exemplar generalization for the test phase in the two conditions. The single exemplar concept results were re-aligned to $\{E,5\}$ and then averaged over the five concepts per participant and over participants. We then took the difference between the generalization gradients for the two conditions, and compared them with the difference between the generalization gradients produced by the Bayesian model. The integral group generalizes more on the diagonals and less on the axes than the separable group as predicted if the integral and separable groups used Euclidean and city-block distance metrics respectively.

To test quantitatively that the two groups learn integral and separable dimensions, we found that the integral training group generalized significantly more often on diagonals than axes (averaging over $\{C,D,F,G\} \times \{3,4,6,7\}$ vs. $C5, D5, F5, G5$, $t(32) = 3.23, p < 0.005$). Within the separable group, the generalization judgments on the axes were significantly greater than the diagonals ($t(34) = 2.66, p < 0.05$); however, the integral group did not differentiate between changes on the axes and the diagonals ($t(30) = 0.43, p = 0.43$). Interestingly, both groups of participants treated the positive diagonal ($F3, F4, G3, G4, C6, C7, D6, D7$) differently than the negative diagonal ($C3, C4, D3, D4, F6, F7, G6, G7$) ($t(34) = 2.58, p < 0.05$ for separable and $t(30) = 2.63, p < 0.05$ for integral). This replicates Krantz and Tversky (1975)'s finding that people tend to generalize rectangles based on constant aspect ratio. This is not surprising as constant aspect ratio is an important invariance of an object's projection on the retina as it changes in depth (keeping the viewpoint orientation constant) due to perspective projection (Palmer, 1999).

Finally, we calculated a mixed effect 2×2 ANOVA that corrobrates the conclusions of our other statistical tests. It identified a main effect of generalizing on the diagonal vs. the

[2] The different number of participants in each group was due to the computer crashing mid-experiment.

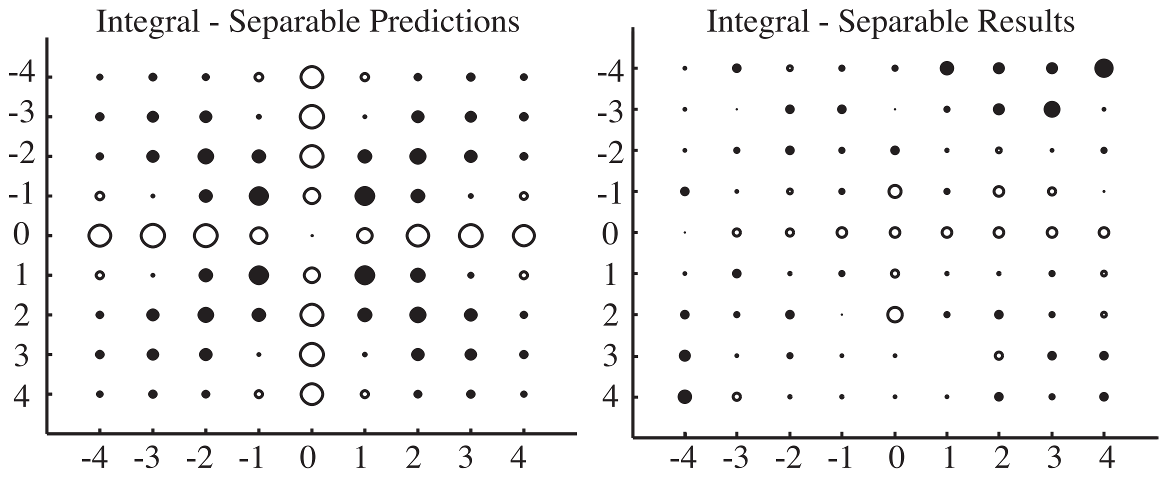

Figure 4: Predictions of the difference between the two Bayesian models formed by model averaging given the separable and integral concepts, and difference between the human generalization results from the two conditions. The results are presented as bubble plots where the size of the bubble represents the degree of generalization. Solid and open bubbles represent positive and negative values respectively. Each single exemplar concept results were re-aligned to $E5$ and then averaged over the five concepts per participant and over participants. Notice how the differences on the axes aligned with the given stimulus ($E5$) are negative and the differences on the diagonals are positive.

axes ($F(1,32) = 44.258, p < 0.001$) and an interaction between generalizing on the diagonal vs. the axes and the training group ($F(1,32) = 10.453, p < 0.005$). This suggests that in the future we should include a hypothesis space into our hierarchy that includes regions varying on the axes and the positive diagonal (but not the negative diagonal).

Discussion

Generalization is an essential problem that basically every cognitive system needs to solve in virtually every domain. Previous analyses of the generalization problem (Shepard, 1987; Tenenbaum & Griffiths, 2001) indicated how an ideal learner should act assuming that an appropriate representation of the stimuli and hypothesis space for generalizations is known. However, how people arrive at a representation and hypothesis space has been left as an open question. As it seems unlikely that people would be born with the appropriate representation and hypothesis space for all possible domains, people need to be able to infer this information from their observations of the properties of stimuli. Using the problem of learning a metric as an example, our analysis shows how an ideal learner would go about inferring such hypothesis spaces, and our experimental results suggest that people do so in a way that is consistent with this model.

To our knowledge, our results provide the first behavioral evidence that people can learn whether stimuli should be represented with separable or integral dimensions. Our results also provide compelling support for the idea that the difference between separable and integral dimensions can be thought of as the result of different hypothesis spaces for generalization, building on (Shepard, 1987, 1991; Davidenko & Tenenbaum, 2001). In future work, it would be interesting to further test this account of separable and integral dimensions by exploring if after training participants show other consequences of having separable or integral dimensions, such as classification and attentional effects. Additionally, this would address a potential confound that the training affects the attention participants pay to each dimension. Fortunately, our larger conclusion that people use the concepts they are given to learn the appropriate hypothesis space for a domain holds regardless of the potential confound (as this conclusion is agnostic to the exact mechanism affecting generalization).

One attractive aspect of this analysis (over using a different solution, like model selection) is that it provides a way to explain why the empirical literature suggests that integrality has been found to be a fuzzy rather than a binary distinction (Garner, 1974). Such fuzzy boundaries emerge as a consequence of Bayesian inference when there is uncertainty to which hypothesis space is appropriate for generalization. We would predict that the "integrality" of natural dimensions are a consequence of how real world objects are categorized along those dimensions. For example, the reason why the saturation and brightness of a color are integral is because in our environment we do not make distinctions between colors at different saturations and brightnesses. "Light" green is a typical color word; however, "saturated" green is an esoteric word, reserved only for artists, designers, and perceptual psychologists. In fact, Goldstone (1994) and Burns and Shepp (1988) found that these dimensions are separable in people who regularly distinguish between the two (color experts and participants trained to distinguish between the two), which implies that they have concepts aligned with the axes of brightness and saturation.

Another important implication of our results is that humans learn the metric appropriate for generalization in a particular domain from the concepts they observe. It would be interesting to compare how metric learning algorithms developed in machine learning (e.g., Xing et al., 2002; Davis et al., 2007) compare to human metric learning on this task, and after learning other types of concepts. This could pave the way towards new machine learning algorithms that automatically infer dimensions intuitive to people from a given set of concepts. Dimensionality reduction techniques like multi-dimensional scaling and principal component analysis are some of the most widely used tools for scientific data analysis, but only produce the equivalent of integral dimensions. An algorithm that determines whether a space is better represented by separable or integral dimensions, and produces interpretable separable dimensions, would be a valuable addition to any data analysis toolkit.

Though Bayesian models have become very popular and successful at explaining different cognitive phenomena (Chater, Tenenbaum, & Yuille, 2006), the hypothesis spaces used in the models are handpicked by the modeler and usually specific to the particular investigated phenomenon. This leaves open the question of how people choose the hypotheses for a set of observed stimuli. Our framework presents an answer to this problem – a hypothesis space is used for a set of observed stimuli depending on how well it explains the observed stimuli and its prior probability. We provide behavioral evidence for our framework in the case study of learning whether or not two dimensions should be separable or integral. Futhermore, this introduces an interesting equivalence between learning the structure of dimensions used to represent stimuli and the set of candidate hypotheses for generalization, which we plan to investigate in future research.

Acknowledgments. We thank Rob Goldstone, Stephen Palmer, Karen Schloss, Tania Lombrozo, and the Berkeley Computational Cognitive Science Lab for insightful discussions, and Shubin Li, Brian Tang, and David Belford for help with experiment construction, running participants, and data analysis and four anonymous reviewers and Nick Chater for their comments on a previous draft of the paper. This work was supported by grant FA9550-07-1-0351 from the Air Force Office of Scientific Research.

References

Burns, B., & Shepp, B. E. (1988). Dimensional interactions and the structure of psychological space: The representation of hue, saturation, and brightness. *Perception and Psychophysics*, *43*, 494-507.

Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Special issue on “Probabilistic models of cognition”. *Trends in Cognitive Sciences*, *10*(7), 287-344.

Davidenko, N., & Tenenbaum, J. B. (2001). Concept generalization in separable and integral stimulus spaces. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society.* Mahwah, NJ.

Davis, J. V., Kulis, B., Jain, P., Sra, S., & Dhillon, I. S. (2007). Information-theoretic metric learning. In *Proceedings of the 24th International Conference on Machine Learning.* Corvallis, OR.

Garner, W. R. (1974). *The Processing of Information and Structure*. Maryland: Erlbaum.

Garner, W. R., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, *1*, 225-241.

Goldstone, R. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, *2*(123), 178-200.

Handel, S., & Imai, S. (1972). The free classification of analyzable and unanalyzable stimuli. *Perception & Psychophysics*, *12*, 108-116.

Kay, P., & McDaniel, C. K. (1978). The linguistic significance of the meanings of basic color terms. *Language*, *54*, 610-646.

Krantz, D. H., & Tversky, A. (1975). Similarity of rectangles: An analysis of subjective dimensions. *Journal of Mathematical Psychology*, *12*, 4-34.

Kruschke, J. K. (1993). Human category learning: Implications for backpropagation models. *Connection Science*, *5*, 3-36.

Lee, M. D. (2008). Three case studies in the Bayesian analysis of cognitive models. *Psychonomic Bulletin and Review*, *15*(1), 1-15.

Palmer, S. E. (1999). *Vision Science*. Cambridge, MA: MIT Press.

Robert, C. P. (2007). *The Bayesian choice: A Decision-theoretic Motivation*. New York: Springer.

Shepard, R. N. (1987). Towards a universal law of generalization for psychological science. *Science*, *237*, 1317-1323.

Shepard, R. N. (1991). Integrality versus separability of stimulus dimensions: from an early convergence of evidence to a proposed theoretical basis. In *The Perception of Structure: Essays in Honor of Wendell R. Garner* (p. 53-71). Washington, DC: American Psychological Association.

Tenenbaum, J. B. (1999). Bayesian modeling of human concept learning. In M. S. Kearns, S. A. Solla, & D. A. Cohn (Eds.), *Advances in Neural Information Processing Systems 11* (p. 59-65). Cambridge, MA: MIT Press.

Tenenbaum, J. B. (2000). Rules and similarity in concept learning. In S. A. Solla, T. K. Leen, & K.-R. Muller (Eds.), *Advances in Neural Information Processing Systems 12.* Cambridge, MA: MIT Press.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences*, *24*, 629-641.

Xing, E. P., Ng, A. Y., Jordan, M. I., & Russell, S. (2002). Distance metric learning, with application to clustering with side-information. In *Advances in Neural Information Processing Systems* (Vol. 12). Cambridge, MA: MIT Press.

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