Identifying representations of categories of discrete items using Markov chain Monte Carlo with People

Anne S. Hsu (anne.hsu@ucl.ac.uk)
Department of Cognitive, Perceptual, and Brain Sciences, University College London, London WC1H 0AP UK

Jay B. Martin (jbmartin@nyu.edu)
Department of Psychology, New York University, New York, NY 10003 USA

Adam N. Sanborn (A.N.Sanborn@warwick.ac.uk)
Department of Psychology, University of Warwick, Coventry CV4 7AL UK

Thomas L. Griffiths (tom_griffiths@berkeley.edu)
Department of Psychology, University of California, Berkeley, CA 94720 USA

Abstract

Identifying the structure of mental representations is a basic problem for cognitive science. We present a method for identifying people's representations of categories that are defined over a set of discrete items, such as a collection of images. This method builds on previous work using Markov chain Monte Carlo algorithms as the basis for designing behavioral experiments, and we thus call it discrete Markov chain Monte Carlo with People (d-MCMCP). We illustrate how this approach can be used to identify the structure of visual categories using real images drawn from large databases.

Keywords: category representation; Markov chain Monte Carlo; image databases

Introduction

Humans outperform the most sophisticated computers in their ability to process complex stimuli, such as recognizing faces or comprehending ambiguous linguistic input. These abilities are facilitated by organizing stimuli into categories. People's representations of categories directly affect their behavior: Recognizing scenes, parsing language, and making decisions, for example, are all influenced by people's category representations. Therefore, understanding the structure of these representations is an important goal for cognitive science. Most research on computational models of categorization has tended to use artificial stimuli, because such stimuli lend themselves to controlled experiments and yield results which are easily quantified (e.g., Nosofsky, 1986; Ashby, 1992; Nosofsky, 1987). However, the stimuli constituting real-life categories – such as images or words – are often characterized by complex features that vary along a large number of dimensions that are hard to quantify. In this paper, we present a method for estimating the structure of categories using an arbitrary discrete set of stimuli, making it possible to investigate real-life categories using complex stimuli such as images drawn from large online databases.

Many computational models of categorization can be interpreted as representing a category as a probability distribution over stimuli (Ashby & Alfonso-Reese, 1995). For example, a category c might be represented by the probability distribution over images x associated with that category, $p(x|c)$. Using this insight, new experimental methods have been developed for estimating these subjective probability distributions. These methods are based on implementing Markov chain Monte Carlo (MCMC) algorithms, which are widely used in computer science and statistics for sampling from complex probability distributions. The Markov chain Monte Carlo with People (MCMCP) method (Sanborn & Griffiths, 2008; Sanborn, Griffiths, & Shiffrin, 2010) adapts MCMC algorithms so as to sample from subjective probability distributions, such as the distributions over stimuli associated with categories. The MCMCP method has been used to estimate the structure of categories defined on continuous, easily parameterized stimuli, such as stick-figure animals and basic fruit shapes (Sanborn et al., 2010) or computer-generated faces (McDuff, 2010; Martin, Griffiths, & Sanborn, 2012).

While the introduction of MCMCP made it easier to explore complex, high-dimensional representations, the original method could only be used with stimuli that vary along a fixed set of parameterized dimensions. This is a serious limitation for exploring real-life categories. For example, it is difficult to quantify the difference between faces with genuine smiles vs. disengenuous smiles. Here, we present an extension of MCMCP that removes this limitation. Our method, which we call discrete Markov chain Monte Carlo with People (d-MCMCP), allows estimation of probability distributions over arbitrary discrete sets of stimuli. This supports the exploration of categories relating to real-life stimuli such as photographic images, every-day objects, real commercial products, and linguistic materials such as documents and words. Because we no longer need to explicitly parameterize the stimuli being examined, d-MCMCP allows us to exploit the vast array of natural stimuli available from the internet.

The outline of this paper is as follows. The next section introduces the key ideas behind MCMCP. We then present our extension of this method to discrete sets of stimuli. The remainder of the paper focuses on two experiments that demonstrate the utility of this method. The first experiment explores the categories of *happy* and *sad* faces using photographic images, allowing us to compare against previous results obtained using the original MCMCP algorithm applied to parameterized images of faces (Martin et al., 2012). The second experiment explores people's concepts of the seasons *Spring*, *Summer*, *Autumn* and *Winter* using a set of 4000 images drawn from online databases.

Markov chain Monte Carlo with People

Markov chain Monte Carlo algorithms are a class of methods for generating samples from complex probability distributions by constructing Markov chains that converge to those distributions over time (see Gilks, Richardson, & Spiegelhalter, 1996). If we want to draw a sample from the probability distribution $p(x)$, we define a Markov chain such that the stationary distribution of that chain is $p(x)$, and sample a sequence of states from that chain. If the sequence is long enough, the states of the chain can be treated similarly to samples from $p(x)$. The Metropolis algorithm (Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953) is one of the most popular methods for constructing such a Markov chain. The sequence of states is initialized with an arbitrary value, x. The next value in the sequence is generated via a two step process. First, a candidate for the next value, x', is chosen by sampling from an arbitrary proposal distribution conditioned on x that is specified by the designer of the algorithm, $q(x';x)$. Second, a decision is made as to whether that proposed value will be accepted using a valid acceptance function which is a function of the relative probability of x and x' under the target distribution $p(x)$. While the original Metropolis algorithm used a different acceptance function, an example of a valid acceptance function is the Barker function (Barker, 1965) which specifies the acceptance probability to be

$$A(x^*;x) = \frac{p(x')}{p(x')+p(x)} \quad (1)$$

and defines a Markov chain that converges to $p(x)$ provided $q(x';x)$ is symmetric, with $q(x';x) = q(x;x')$.

The Markov chain Monte Carlo with People method uses the idea that categories can be represented as probability distributions over stimuli (Ashby & Alfonso-Reese, 1995). The distribution over stimuli x for category c, $p(x|c)$ indicates the degree to which a stimulus item x is perceived to represent a given category c. In theory, the simplest approach to measuring human categories would be to ask people to rate the degree of category membership for all possible stimuli. However, this has two serious limitations. First, categories span such a large number of possible items that collecting individual ratings of each are not practical. Second, a question such as "How good is this example of a *happy face*?" is difficult to answer, and there is no obvious scale to use for the answer. A solution to the second limitation would be to ask people to make pairwise judgments, i.e. "Which example is a better example of a *happy face*?". However, this only exacerbates the first limitation because the number of judgments required for all possible pairs of n items is now on the order of n^2.

Markov chain Monte Carlo with People addresses the challenge of estimating the distribution associated with a category by constructing a Markov chain that produces samples from that distribution. The method is based on a correspondence between human choice behavior and the Barker acceptance function. If a task can be constructed in which people are offered a choice between x and x' and choose x' with probability

$$P_{\text{choice}}(x';x|c) = \frac{p(x'|c)}{p(x'|c)+p(x|c)} \quad (2)$$

then this provides a valid acceptance function for a Markov chain that will converge to $p(x|c)$. Equation 2 has a long history as a model of human choice probabilities, where it is known as the Luce choice rule or the ratio rule (Luce, 1963; Shepard, 1957). This rule has been shown to provide a good fit to human data when people choose between two stimuli based on a particular property (Bradley, 1954; Clarke, 1957; Hopkins, 1954). The Luce choice rule has also been used to convert psychological response magnitudes into response probabilities in many models of cognition (Nosofsky, 1986; Ashby, 1992; Nosofsky, 1987; McClelland & Elman, 1986).

Based on this correspondence, the MCMCP method implements the Metropolis algorithm, using people's choices to determine which proposals are accepted (Sanborn et al., 2010). In a standard experiment, people would be asked to make a series of pairwise decisions in which they are asked to choose the best category member from two proposed stimuli. The stimuli that are presented in each decision correspond to the values x and x' in the Metropolis algorithm, and the choices that people make determine which proposals are accepted. With enough decisions, MCMCP will converge to samples from the probability distribution associated with that category, and individual stimuli will be encountered with probability given by $p(x|c)$. The proposal distribution can be selected by the experimenter, provided it is symmetric in the way required by the Barker acceptance rule.

Estimating categories for discrete sets of stimuli

The MCMCP method requires defining a proposal distribution $q(x';x)$ for choosing the next stimulus to present on each trial based on the current stimulus. When stimuli are described by a fixed set of parameters, this is easy – previous work has used Gaussian or uniform distributions to generate proposals for continuous parameters, and a multinomial distribution can be used to propose new values for discrete features. However, real-life categories are not made up of easily parameterized items: Real-life categories apply to stimuli that are difficult to parameterize such as real objects, images, sounds, and words. The lack of parameterization makes it unclear how to propose a reasonable stimulus based on the current stimulus, which is central to the MCMCP algorithm's efficiency. Hence, in order for MCMCP to measure representations of a wide range of real-life categories, we need to a method for making reasonable proposals when exploring stimuli that are not easily parameterized. In this section, we introduce such a method, which we call discrete Markov Chain Monte Carlo with People (d-MCMCP).

The d-MCMCP procedure adds three steps to MCMCP. The first step is to create a database of stimulus items over which the probability distribution associated with a category is to be estimated. The second step is computing a rough measure of the similarity between all possible item pairs, giving a symmetric similarity matrix, S. The similarity metric can be chosen as appropriate for the domain, and need only provide a heuristic guide to the perceived similarity of human participants. For example, similarity between color histograms can used to quantify similarity for color images. The third step is constructing a graph of the stimulus items based on these similarities. A random walk on this graph is then used to define the proposal distribution used in MCMCP.

A key assumption in using the Barker acceptance function is that the proposals must be symmetric. That is the probability of choosing a proposal value given a current value would be the same if the proposal and current values were reversed. In order for this to be true of a random walk on a graph, the edges must be symmetric (ie. the walk can traverse an edge in each direction), and each node in the graph must have the same degree (ie. each node must have the same number of neighbors). Just choosing the b nearest neighbors (as given by the similarity metric) for each node does not suffice, because while node i might be one of b nearest neighbors to node j the reverse is not does not have to be true. As a result nodes will have different degrees. Taking the union or intersection of edges resulting from considering the nearest neighbors of each item will also result in unequal degrees.

To address this problem, we instead construct the graph that maximizes the similarity along edges while keeping the degree of each node constant. Formally, we want to find

$$\arg\max_G \sum_{ij} G_{ij} S_{ij} \quad \text{s.t.} \quad \sum_j G_{ij} = b; \quad G_{ii} = 0; \quad G_{ij} = G_{ji}$$

where G is the adjacency matrix of the graph, with $G_{ij} = 1$ if there is an edge from i to j and $G_{ij} = 0$ otherwise. This is an instance of the *maximum weight b-matching* problem (Papadimitriou & Steiglitz, 1998). Exact algorithms exist for solving this problem, such as the *blossom* algorithm (Edmonds, 1965), but these are impractical for large-scale applications. Consequently, we use an approximate algorithm based on max-product message passing to find a b-matching (Jebara & Shchogolev, 2006).

Given a graph on stimuli that is a b-matching, proposals for the d-MCMCP algorithm can be made in a variety of ways. The selected proposal method is held constant throughout the experiment (as is standard in MCMC and MCMCP). The most straightforward proposal method is to choose a proposal uniformly from all b neighbors, where the value of b is chosen at the experimenter's discretion. A second method is to make a geometric proposal. Here, the proposal is generated iteratively using a number of steps, n_{geom}, that is chosen from a geometric distribution with a fixed parameter. A random walk of length n_{geom} is then performed, choosing the next node uniformly from the b neighbors of the most recent one. The node at the end of the random walk is the proposal. For all proposal methods it is prudent to allow for some small probability of choosing uniformly from all possible stimulus items to allow the algorithm to move between local maxima.

Experiment 1: *happy* and *sad* faces

As a first test of the d-MCMCP method, we examined the categories of *happy* and *sad* faces using a database of images of real faces. Previous work had applied the original MCMCP method to estimating these categories using parameterized face stimuli, where a continuous space was derived from eigenfaces computed from the same set of images (Martin et al., 2012). Use of the same image database allows direct comparison of the results of d-MCMCP and MCMCP with a matched stimulus set, and ratings of the emotional content of the resulting faces provide a way to evaluate these results.

Method

Participants. A total of 60 undergraduates participated in exchange for course credit.

Stimuli. A database of 1271 images of faces was assembled from the California Facial Expression (CAFE) database, a collection of 1280 normalized 40 × 64 pixel gray-scale portraits containing 64 individuals (Dailey, Cottrell, & Reilly, 2001), expressing approximately eight distinct "FACS-correct" emotions, which are classified according to the taxonomy of the Facial Action Coding System (Ekman & Friesen, 1978).

Procedure. Face images were convolved with Gabor filters at 8 scales and 5 orientations. Principal Components Analysis (PCA) was then applied to the whole set of convolved images and the Euclidean distance between the top 50 components was used as the similarity metric for defining the matrix S. Two graphs G were produced using the approximate b-matching algorithm from Jebara and Shchogolev (2006), one with $b = 6$ and one with $b = 16$. This algorithm gives an approximate solution to the b-matching problem, so there was still some minor variation in the degree of individual nodes. Our empirical evaluation of the performance of the d-MCMCP procedure will thus also help to indicate whether this residual variation affects the results. There is no guarantee that a maximal b-matching is connected, so we used the largest connected component as the basis for the d-MCMCP procedure. The largest connected component contained 1216 images with $b = 6$ and all 1271 images with $b = 16$.

We compared three different methods for defining the proposal distributions. For all three proposal methods, we allowed for a 10% chance of proposing a jump to a node chosen uniformly at random. The three methods for choosing the remaining proposals were the uniform random walk on the graph with $b = 6$ (U6), the uniform random walk on the graph with $b = 16$ (U16), and the geometric proposal with $n_{\text{geom}} = 0.5$ on the graph with $b = 6$ (G6).

Participants were randomly assigned to proposal-type conditions. Trials were presented on three different computers, one for each proposal type. Each participant completed trials corresponding to four d-MCMCP chains (two for *happy*

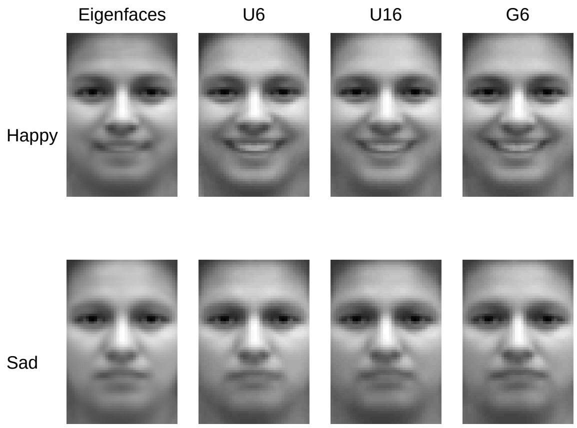

Figure 1: Results of comparing MCMCP using eigenfaces and d-MCMCP with a variety of proposal methods on the same set of face stimuli. Average faces for each type of proposal. Averages are taken across all trials and all four chains corresponding to *happy* and *sad*.

faces, two for *sad* faces). There were 100 trials for each of the four chains. On a given trial, the participant decided which of a pair of faces was either more *happy* or more *sad*. Twelve trials in the beginning were offered as practice, which were not included in the analysis. There were also 40 catch trials with face pairs for which the more *happy* or *sad* face was clearly obvious (in this case, we used the emotion designations in the CAFE database to select faces that should clearly be happy or sad). Thus each participant responded to $100 \times 4 + 12 + 40 = 452$ trials, which took approximately 25 minutes. The responses were linked in chains of ten participants each: The last trial of each of the four chains was passed along to the next participant as his/her first non-practice trial to form a linked chain of 1000 trials. Participants who did not correctly answer at least 27 catch trials ($p < .01$ under random guessing) were not included in the results, or added into a chain. We collected two chains of 10 participants for each proposal type, corresponding to four *happy* and *sad* chains with 1000 trials in each chain.

Results

The images selected on each trial were averaged together to produce the average faces shown in Figure 1. All three proposal methods produced mean faces that appeared reasonably consistent with the target emotions. Also included in Figure 1 are the results reported in (Martin et al., 2012) using MCMCP in a parameterized space based on the eigenfaces derived from the image database we used for d-MCMCP. Qualitatively, the results from d-MCMCP are at least as good and perhaps better than those produced using eigenfaces.

To quantify the performance of the different variants of the algorithm, we conducted a follow-up experiment in which a group of 40 participants recruited via Amazon Mechanical

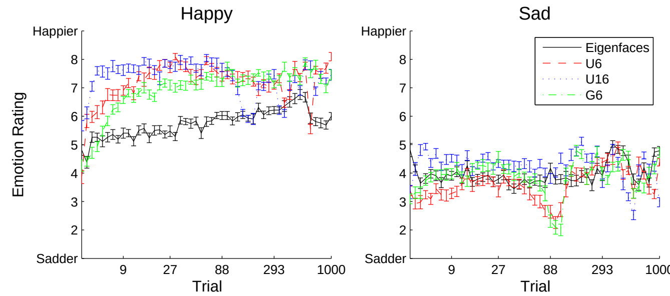

Figure 2: Happiness ratings for average faces for three types of d-MCMCP proposals as well as original MCMCP method as a function of trial number (error bars show one standard error). Averages are taken across the 50 most recent trials (or starting from the first trial for trials less than 50) and across all four chains corresponding to the same emotion (*happy* or *sad*). Also included are face ratings for the results of a previous MCMCP experiment that used eigenfaces derived from the same image database (Martin et al., 2012).

Turk provided ratings of the emotions exhibited by faces derived from our chains. For each proposal type (and for the chains based on eigenfaces used in Martin et al. (2012)), cumulative average faces were computed for each of 40 logarithmically spaced numbers of trials, averaging across all four chains that corresponded to each emotion. For trial numbers greater than 50 images were averaged only over the 50 most recent trials, meaning that no more than 200 faces contributed to any single image. Participants rated the emotion exhibited by each of these mean faces on a scale from 1-9, where 1 indicates "very sad" and 9 indicates "very happy". All participants rated all faces, and received $1 in compensation for their time.

The results of our follow-up experiment are shown in Figure 2. The d-MCMCP method results in statistically significantly higher ratings for faces derived from *happy* chains regardless of proposal type, perhaps as a consequence of being able to explore a larger space of faces than the eigenface method. Results for *sad* chains are more comparable. There are no systematic differences between the different proposal types, although the U16 proposal appeas to produce happier faces faster than the other two proposals. For both *happy* and *sad* chains there is some variation in the emotion ratings of mean faces over time, consistent with the idea that MCMCP should be exploring the distribution of faces associated with the category (and possibly moving between modes of that distribution) rather than finding the most extreme instance of that category.

Experiment 2: Seasonal images

Our first experiment indicated that d-MCMCP produced comparable or better performance to MCMCP when applied to a set of stimuli where both methods could be used. In our second experiment, we used d-MCMCP to explore categories defined on a set of stimuli for which there is no simple parametric representation. Specifically, we explored the categories of images associated with the seasons *Spring*, *Summer*, *Autumn*, and *Winter*, using 4000 images obtained from online image databases. By applying the d-MCMCP procedure to these stimuli, we can identify high probability images and compute informative aggregate statistics for each category, allowing us to answer questions such as what distribution of colors is associated with each season.

Method

Participants. A total of 90 participants were recruited using Amazon Mechanical Turk. Each participant was paid $1 for completing the 25 minute experiment.

Stimuli. A set of 4000 colored season-related images was assembled by searching for public domain web images using the phrases "spring season", "summer season", "autumn season", and "winter season" in Google Image Search and on Flickr.com. The top 500 results for searches on Google and Flickr for each season were downloaded using Bulk Image Downloader. All images were resized so that the maximum dimension was 250 pixels, while preserving the original ratio of image height to width.

Procedure. The similarity between all possible image pairs (7998000 pairs for 4000 images) was quantified using both the Basic Color Histogram (BCH) descriptor (Griffin, 2006) and the Scale-Invariant Feature Transform (SIFT; Lowe, 1999). BCH classifies and counts pixels as belonging to one or other of the eleven basic colors (black, white, grey, red, orange, yellow, green, blue, purple, pink, and brown). SIFT applies local filters to transform images into collections of local feature vectors which are invariant to scaling, rotation and translation of the image. Similarity results over all pairs of images for both methods were normalized to have unit variance and then added together, thus yielding a similarity measure which combined results of both BCH and SIFT. The similarity between all pairs was represented as a similarity matrix which was fed into the *b*-matching algorithm. A graph was found using $b = 5$, which was the smallest value such that all 4000 images remained fully connected. We used a proposal distribution corresponding to a uniform random walk on this graph.

Each participant made pairwise choices between images by answering questions such as *Which image is more representative of Spring?*. There were 100 trials for each of four chains, one for each season. There were also 12 practice trials, and 40 catch trials for which one image of the pair obviously corresponded to a particular season (as judged by the experimenter). Thus each participant completed 452 trials. Participants who did not at least get 27 catch trials correct were not included in the chains or analysis. We collected data by linking three sets of 10 participants forming three chains of 1000 trials for each of the four seasons.

Results

The top ten images that were chosen most often over all three chains for each season are shown in Figure 3. Clearly, the images are very indicative of each season. Figure 4 (a) shows, as a function of the number of trials, the L1 distance between 11-bin color histograms calculated for cumulative images, both between chains for the same season and between chains corresponding to different seasons. Within-chain distance decreases over time, and is typically lower than the similarity between chains, supporting the idea that chains are converging towards different parts of the space of images. Figure 4 (b) shows a simple example of the kind of statistical analyses that can be done on the resulting samples. The color histograms for the different seasons are quite different from one another, and each correspond to a palette that seems consistent with our intuitive representation of each season.

Conclusion

We have presented a new method for estimating the structure of people's mental representation of categories, showing that it produces performance that is comparable to existing methods, and can be used with rich sets of complex stimuli such as images derived from online databases. By extending the MCMCP algorithm so that it can be applied to any arbitrary set of stimuli, our d-MCMCP method makes it possible to measure people's representations of a broader range of natural categories, and in a greater variety of real-world settings. Using our approach, MCMCP algorithms can be applied to large databases which contain discrete items, such as images or text. This has the potential to lead to significant advances for cognitive scientists interested in studying categories in a way that goes beyond simple parameterized stimuli. The results of such an investigation are likely to be valuable to machine learning and computer vision researchers interested in training systems to produce and improve on human performance in categorizing images and other complex stimuli. Conducting experiments using d-MCMCP on a large scale will allow us to build up a catalogue of human category representations, taking a step towards understanding how those categories are formed.

Acknowledgments. This work was supported by grant number IIS-0845410 from the National Science Foundation.

References

Ashby, F. G. (1992). *Multidimensional models of perception and cognition*. Hillsdale, NJ: Erlbaum.

Ashby, F. G., & Alfonso-Reese, L. A. (1995). Categorization as probability density estimation. *Journal of Mathematical Psychology*, *39*, 216-233.

Barker, A. A. (1965). Monte Carlo calculations of the radial distribution functions for a proton-electron plasma. *Australian Journal of Physics*, *18*, 119-133.

Bradley, R. A. (1954). Incomplete block rank analysis: On the appropriateness of the model of a method of paired comparisons. *Biometrics*, *10*, 375-390.

Clarke, F. R. (1957). Constant-ratio rule for confusion matrices in speech communication. *The Journal of the Acoustical Society of America*, *29*, 715-720.

Dailey, M., Cottrell, G., & Reilly, J. (2001). *California facial expressions (CAFE)*. University of California, San Diego: Unpublished digital images.

Edmonds, J. (1965). Paths, trees and flowers. *Canadian Journal of Mathematics*, *17*, 449-467.

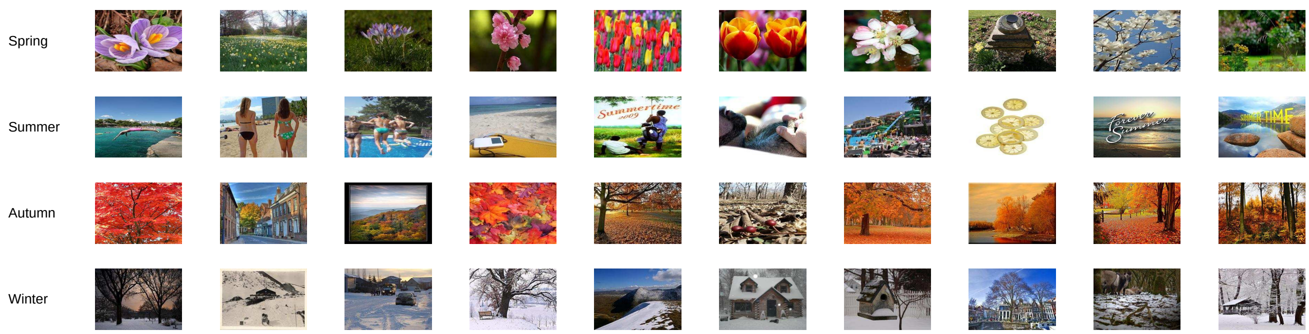

Figure 3: Top ten most popular images over all chains for each season, decreasing in popularity from left to right.

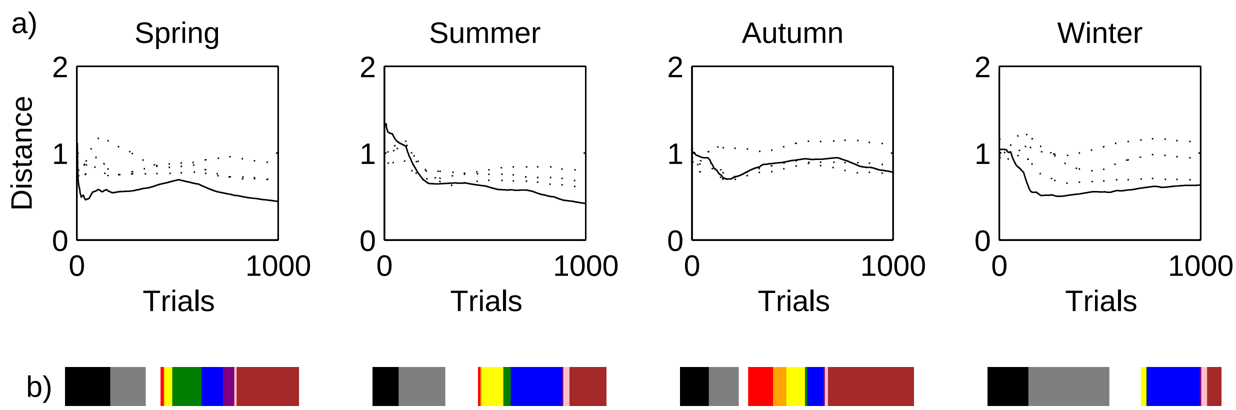

Figure 4: 11-bin color histograms were calculated for all cumulative images in all three chains as a function of the number of trials. a) average L1 distance between the cumulative histogram of a single chain and the other two chains which correspond to the same season (solid line) or the other three chains which correspond to a different season (one dotted line for each other season). b) Color histograms of all images, averaged over all chains for each season (Griffin, 2006).

Ekman, P., & Friesen, W. (1978). *Facial action coding system: A technique for the measurement of facial movement*. Palo Alto, CA: Consulting Psychologists Press.

Gilks, W., Richardson, S., & Spiegelhalter, D. J. (Eds.). (1996). *Markov chain Monte Carlo in practice*. Suffolk: Chapman and Hall.

Griffin, L. D. (2006). The basic colour categories are optimal for classification. *Journal of the Royal Society Interface*, *3*, 71-85.

Hopkins, J. W. (1954). Incomplete block rank analysis: Some taste test results. *Biometrics*, *10*, 391-399.

Jebara, T., & Shchogolev, V. (2006). B-matching for spectral clustering. In *Proceedings of the European Conference on Machine Learning, (ECML).*

Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Proceedings of the International Conference on Computer Vision (ICCV).*

Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology, volume 1* (p. 103-190). New York and London: John Wiley and Sons, Inc.

Martin, J. B., Griffiths, T. L., & Sanborn, A. N. (2012). Testing the efficiency of Markov Chain Monte Carlo with people using facial affect categories. *Cognitive Science*, *36*, 150-162.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1-86.

McDuff, D. (2010). A human-markov chain monte carlo method for investigating facial expression categorization. In *Proceedings of the 10th International Conference on Cognitive Modeling.*

Metropolis, A. W., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, *21*, 1087-1092.

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

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 87-108.

Papadimitriou, C., & Steiglitz, K. (1998). *Combinatorial Optimization: Algorithms and Complexity*. New York: Dover.

Sanborn, A. N., & Griffiths, T. L. (2008). Markov Chain Monte Carlo with people. In *Advances in Neural Information Processing Systems 20.*

Sanborn, A. N., Griffiths, T. L., & Shiffrin, R. (2010). Uncovering mental representations with Markov Chain Monte Carlo. *Cognitive Psychology*, *60*, 63-106.

Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, *22*, 325-345.