Category Learning: Comparison of computational and human methods

Turner Bohlen Troy Astorino

January 1, 2013

1 Introduction

Rational analyses of human cognition seek to explain and quantify human behavior and thought processes under the assumption that they are an optimal adaptation to the constraints of the environment. Anderson [1] argues that categorization is a basic function of human cognitive processes, and that Bayesian statistical inference is a theoretically motivated and effective model for human categorization. Sanborn, Griffiths, and Navarro [2] further investigate Bayesian algorithms for category learning, finding a single-particle particle filter to be most effective, and most similar to human behavior.

To extend upon the previous work concerning particle filters and category learning, this experiment analyzes each individual move made by a human while sorting data, rather than simply analyzing the end result. A single end sort can be achieved by a number of paths exponential in the size of the largest category. Analyzing the process step by step allows access to this vast amount of missed information, assuming it can be analyzed in some useful way. In this way, move-by-move analysis is an exponentially tougher test for human inference models.

2 Experimental Methods

The task that trial subjects were faced with was designed to facilitate comparison between the subjects' decisions and those that would be made by a particle filter. Modeling sequential, online category learning was useful both because that was the type of categorization algorithm developed in previous work, and because it has a straightforward manifestation in a constrained human task. Trial subjects were presented with stimuli images, sequentially, and asked to assign each image into a group. Once the image was placed in a group, the subject was not allowed to switch the image to another group. The instructions given to each subject can be viewed in the Appendix, Figure 6.

The stimuli used were 100 by 100 pixel images. A sample stimulus can be viewed in Figure 1. Each pixel in an image is a shade of gray between white and black, with 256 possible grayscale values. This type of stimuli is convenient because it is directly interpretable for the inference algorithm: each pixel-block is a feature, and each feature can take one of 256 possible values. This type of stimuli was also considered attractive because of the

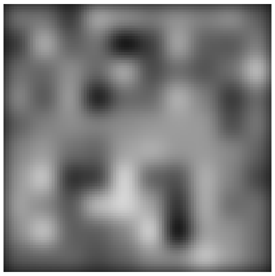


Figure 1: A sample stimulus to be categorized. The stimuli were generated from 10 by 10 matrices of values between 0 and 255, where the value determined the grayscale value. The images were smoothed.

difficulty for the human subject of categorizing the images. It was thought that this difficulty would compel the subjects to rely on subconscious processes for categorization. In retrospect, as is addressed in Section 5, the high dimensionality of the images may have made it too difficult for the brain to categorize the images effectively, and may have lessened the interpretive power of the results. Figure 2 shows how the interface would appear for a mostly completed trial.

Workers from Mechanical Turk were hired to be trial subjects. They were presented with a website on which to conduct the trial: each worker was first shown the instruction page in Figure 6, and then were brought to the page on which the trial was conducted, show in Figure 2 All the actions performed by the subject were collected in a database to be used for later analysis.

3 Particle Filter

The algorithm implemented was based on the single-particle particle filter described by Sanborn et al[2]. Justification for the use of a particle filter to perform probabilistic inference for sequential clustering can be found in that paper, in addition to a development of the models used to build this particle. Presented below is only a brief description of the underlying probability distributions used, in addition to the modifications required to tailor the particle filter to this experiment.

The posterior probability that a stimulus is assigned to a group is proportional to a prior probability multiplied by a likelihood, as is always the case with Bayesian inference. The prior probability encodes a preference

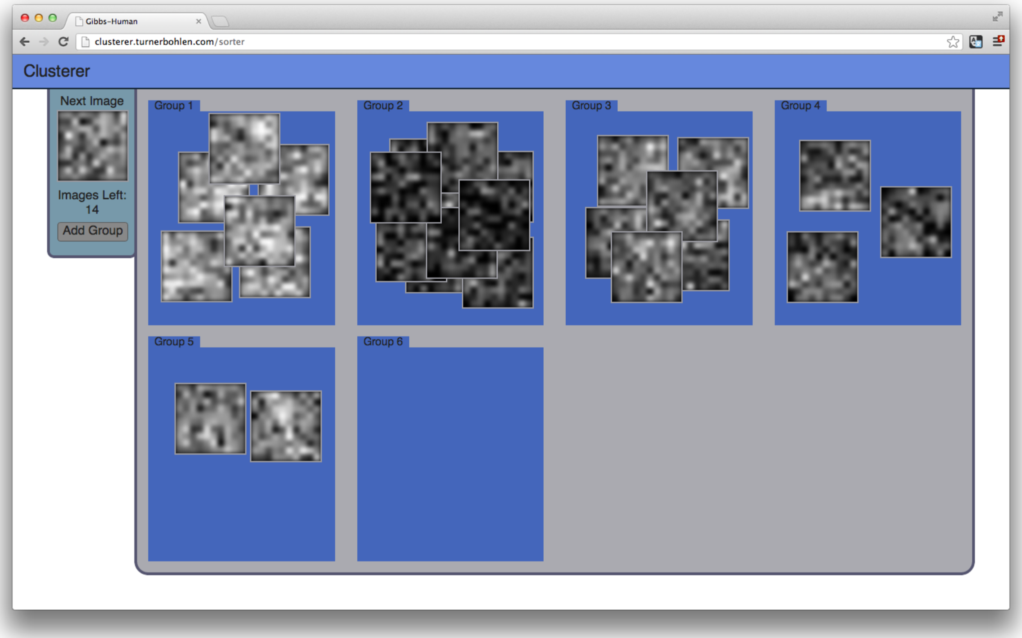


Figure 2: Interface through which a trial was completed. Images are dragged from the 'Next Image' box into one of the groups. Once an image was placed in a group, it could not be switched into another group. The subject was limited to creating 8 groups.

over group sizes: relatively how large groups should be, when new groups should be created, etc. This prior must protect against overfitting, which in this context of category learning would be creating a new category for each slightly different stimulus. The likelihood function encodes the probability that the stimulus is drawn from the same cluster that produced the stimuli already in the cluster. Using the primarily the same notation as the Sanborn et al. paper, the prior takes the form, this proportionality is represented by

$$P(z_i = k|\mathbf{X}_i = \mathbf{x}_i, \mathbf{z}_{i-1}) \propto P(\mathbf{X}_i = \mathbf{x}_i|z_i = k, \mathbf{z}_{i-1})P(z_i = k|\mathbf{z}_{i-1}) \tag{1}$$

where $z_i = k$ notates assigning the i^{th} stimulus to group k, $\mathbf{z}_{i-1}$ refers to the assignment of groups that the previous $i-1$ stimuli have gone through, and M_k is the number of stimuli in group k after the previous $i-1$ stimuli had been sorted. In Equation 1, $P(\mathbf{X}_i = \mathbf{x}_i|z_i = k, \mathbf{z}_{i-1})$ represents the likelihood and $P(z_i = k|\mathbf{z}_{i-1})$ represents the prior.

A Dirichlet process models the prior distribution over the probability that any given input stimuli will be grouped with a given cluster, whether that is one of the existing clusters or would be a new cluster. This results in the following form for the prior probability:

$$P(z_i = k|\mathbf{z}_{i-1}) = \begin{cases} \frac{M_k}{i-1+\alpha} & M_k > 0 \\ \frac{\alpha}{i-1+\alpha} & M_k = 0 \end{cases} \tag{2}$$

This says the probability that the i^{th} stimulus is placed in group k is proportional to the number of stimuli already in group k, or to a parameter of the Dirichlet process, α, if group k would be a new group. α is the dispersion parameter of the Dirichlet process; the larger α, the larger the probability that a stimulus will be assigned to a new group. The value used for this parameter, as well as the values used for other parameters of the particle filter, can be found in Figure 3.

The likelihood model for a stimulus being in a given group assumes that each feature in a group follows a Gaussian distribution. The prior on the variance of this Gaussian is modeled as an inverse χ^2 distribution, and the prior on the mean is modeled as another Gaussian. These priors result in the likelihood function over each feature having the form of a Student's t distribution with a_i degrees of freedom.

$$X_{i,d}|z_i = k, \mathbf{z}_{i-1} \sim t_{a_i}\left(\mu_i, \sigma_i^2\left(1 + \frac{1}{\lambda_i}\right)\right) \tag{3}$$

where

$$
\begin{aligned}
\lambda_i &= \lambda_0 + M_k && (4)\\
a_i &= a_0 + M_k && (5)\\
\mu_i &= \frac{\lambda_0 \mu_0 + M_k \bar{x}}{\lambda_0 + M_k} && (6)\\
\sigma_i^2 &= \frac{a_0 \sigma_0^2 + (n-1)s^2 + \frac{\lambda_0 M_k}{\lambda_0 + M_k}(\mu_0 - \bar{x})^2}{a_0 + M_k} && (7)\\
& && (8)
\end{aligned}
$$

$X_{i,d}$ is a random variable for feature d of the i^{th} stimulus. In Equation 3, $X_{i,d}$ is conditioned on the i^{th} being assigned to group k and the group assignments of the previous $i-1$ stimuli. M_k is again the number of elements in group k, but in this instance assuming that the i^{th} stimulus has been added to group k. The prior mean is μ_0, and the prior variance is σ_0^2, and the confidences in the prior mean and prior variance are λ_0 and a_0, respectively. μ_0 is set to midpoint of the potential values for each feature, and σ_0 is set to be $1/8^{th}$ the range of the potential values.

The input were 100 by 100 pixel images, with each pixel taking on a grayscale value between 0 and 255. Each feature was treated as an independent feature. Because the range for each feature was limited and discrete, the Student's t distribution was discretized and renormalized along the valid range of the feature each time a feature likelihood value was calculated.

In order to match the methodology used by Sanborn et. al, the features are treated as being independently distributed, so the likelihood for the entire stimulus is simply a product over all the feature likelihoods:

$$P(\mathbf{X}_i = \mathbf{x}_i | z_i = k, \mathbf{z}_{i-1}) = \prod_d P(X_{i,d} = x_{i,d} | z_i = k, \mathbf{z}_{i-1}) \qquad (9)$$

Additionally, the particle filter was only allowed to create 8 groups, to match this limitation that was placed on human subjects. The restriction was enacted by not allowing the particle filter to consider placing a stimulus into a new group after 8 groups had already been created.

Parameter	Value
α	30
μ_0	127.5
λ_0	0.5
σ_0	32
a_0	2.0

Table 1: The parameters used for the particle filter. α is the dispersion parameter for the Dirichlet process prior. μ_0 and σ_0 are the mean and standard deviation of the prior Gaussian distribution over each feature, and λ_0 and a_0 are the confidence in the prior mean and the confidence in the prior variance, respectively.

4 Results

62 mechanical-turk trial results were analyzed, giving a total of 2480 moves. For each move, the probability assigned by the particle filter to the categorization decision made by the subject was recorded. The mean of these 2480 log probabilities was an unimpressive -1379.256, and the median was -29.612.

The three histograms in figures 3, 4, and 5, though, show that 48.9% of the log probabilities were greater than -0.25 and 51.5% were greater than -1.0. 1216 moves saw the particle filter inference algorithm assigning probabilities (*not* log probabilites) greater than 0.5 to the categorization selected by the human, indicating that the particle filter model used in this experiment agreed with the human data at least 49.032% of the time. The average move involved a choice between 4.929 groups, making a 20.286% success rate equivalent to chance, and showing that the particle filter performed far better than chance.

5 Discussion

The results of this experiment indicated promise for Baysian models, and, in particle, particle filter models, of human cognition, which was to be expected considering the work by Anderson, Sanborn, Griffiths, and Navarro supporting this belief. That said, the 49.032% success rate was very promising but hardly conclusive evidence that a particle filter model could accurately model

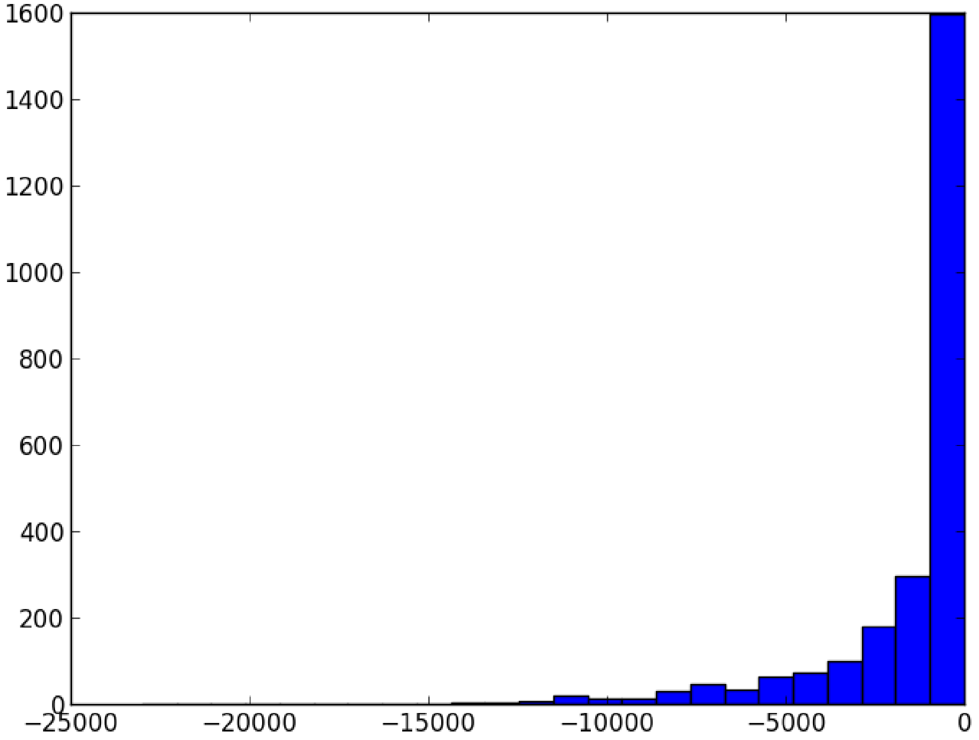


Figure 3: A histogram showing the log probability of the particle filter making the same move as the human for each of the 2480 human moves.

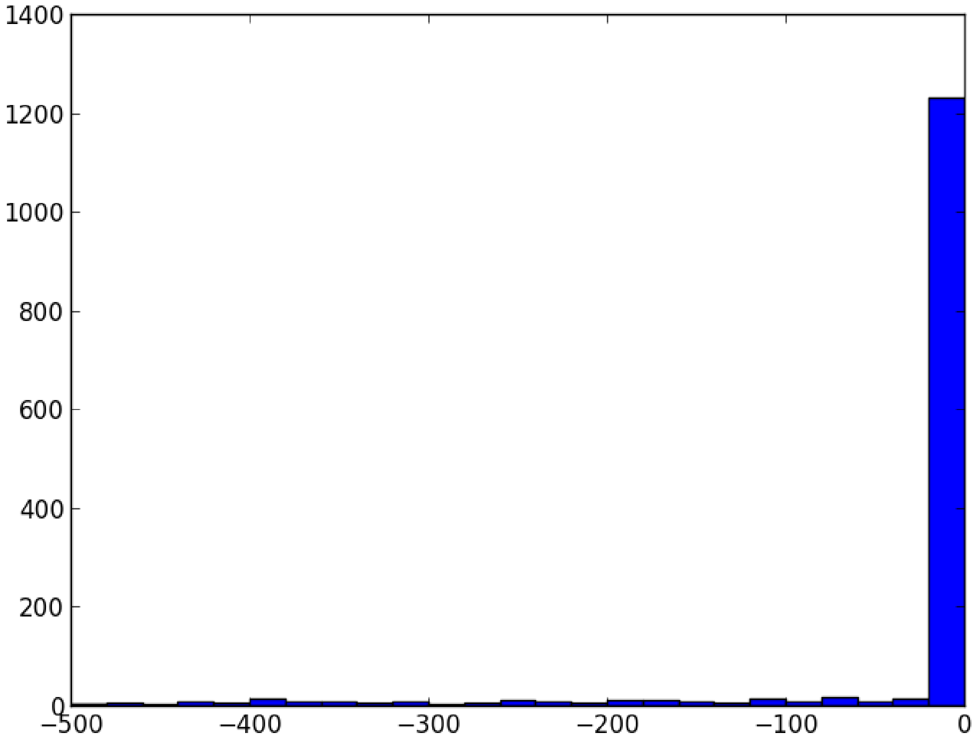


Figure 4: A histogram showing the log probability of the particle filter making the same move as the human for each of the moves for which this value was greater than −500. This is a zoomed in view, so to speak, of 3.

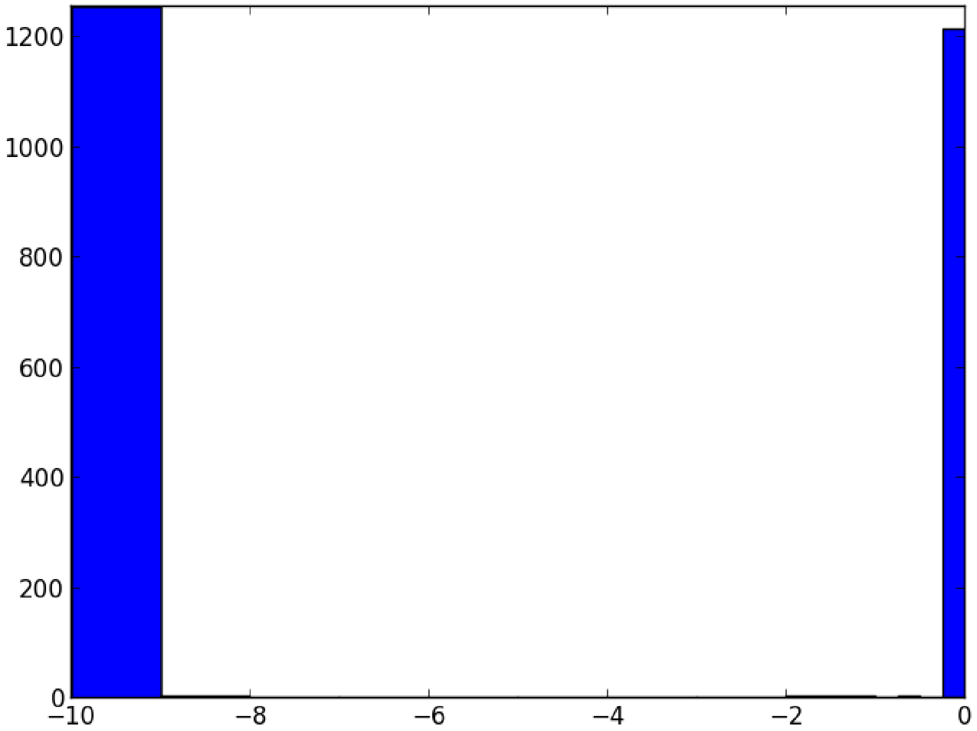


Figure 5: A histogram showing the log probability of the particle filter making the same move as the human for each of the 2480 human moves, but with all moves for which this value was less than −9 grouped into the lowest probability bin.

the human cognitive process. Further analysis of the data presented here, as well as a number of refined studies suggested below would be necessary to strengthen this claim.

The error in the particle filter selection process, when compared to human data, could be the result of a number of errors or incorrect assumptions. The particle filter treated each pixel of the 100 by 100 pixel images as a feature, presenting two potential problems. First, this extremely high dimensionality may have been the reason for the tendency of the filter to assign a probability of nearly 1 on its top choice, and probability of nearly 0 to all other choices.

More worryingly, such a treatment of the image does not fit well with a human's tendency to extract features, even from such relatively featureless images. In future this could be remedied by performing a feature extraction or simply including in the likelihood model a covariance between dimensions. Future experiments could also use more structured images which would be more easily interpreted by the subjects, and, therefore, potentially result in better data.

The particle filter likelihood calculation involved the discretization of a Student's t distribution via discrete sampling of a continuous Student's t distribution followed by a normalization of these samples. This method may not be an acceptable mapping of the continuous function to a discrete domain. Additionally, the lack of any quantification of the 8-group cap in the prior probability distribution could have caused error.

Another possible source of error, which should be considered in all experiments using web interfaces and paid participants, was the interface itself. The fact that images layered on top of one another when added to the same group may have skewed the human categorization behavior and perception of the groups. A future experiment that presents half of the subjects with the average of all image in a group, and all others with each individual image in a group and analyzed the resulting categorization data for biases could help determine if this is a valid concern. Other subtle interface decisions, such as the fixed area of the group boxes presented on the web interface could have further biased the data and should be considered carefully in future experiments. Additionally, a subset of the samples generated from Mechanical Turk should always be checked to make sure the subjects did not create ”junk” trials simply to win their pay.

Despite these many concerns, this study did show that comparing human categorization to a Baysian categorization model on a move by move basis is both feasible and worthwhile. The particle filter took approximately twelve hours to analyze all 2480 individual moves on a 2.2Ghz Intel Core i7 processor, and required around 100MB of memory to do so. Move-by-move analysis does indeed offer a tougher text for Baysian categorization models, as mentioned in the introduction, and so further use of this technique seems prudent.

One final, although quite interesting, potential direction for further experiments involves investigating the parameters used in the model. A machine inference algorithm could be decisively shown to exclude some logic that a human categorizer is using if the parameters that result in the best fit between a human's decision and the algorithm's decision vary significantly from move to move. Such analysis seems valuable to pursue in future experiments. Phrased another way, significant variation in the best fit parameters indicates that the humans had to apply additional logic to modify his internal model of the system, and so the inference model in question is missing some human reasoning. Such analysis seems valuable to pursue in future experiments.

6 Conclusion

This experiment was designed to extend the previous work concerning particle filters and category learning by Anderson, Sanborn, Griffeths, and Navarro by comparing not just the results of Baysian machine categorization with human categorization, but each move made during the categorization process. Such an analysis has the potential to offer much stronger evidence for or against human categorization models

68 subjects participated in the 40-image categorization test, and the particle filter inference algorithm used agreed with the human choices 49.032% of the time. Chance would have resulted in a 20.286% agreement rate.

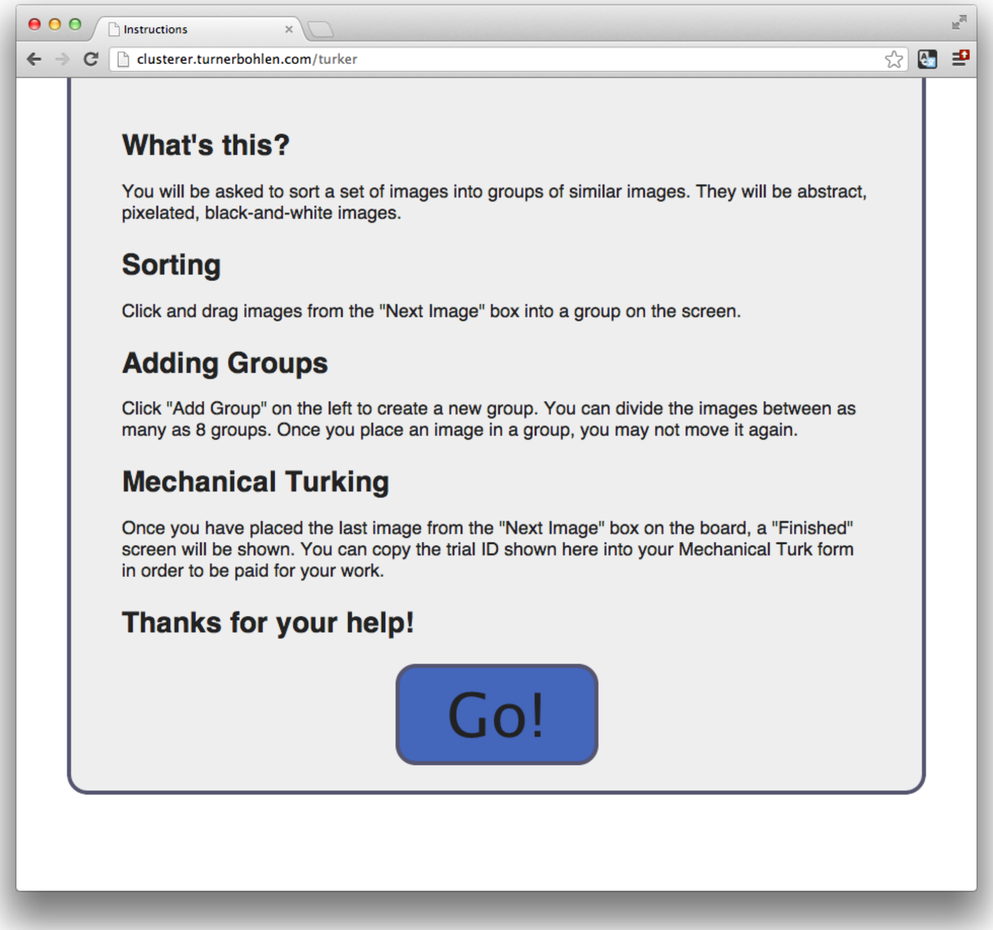
Although not conclusive, the results do support the existing evidence that Baysian models, and, in particular, particle filters are good potential models for human categorization. Potentially more importantly, this study functioned as a proof-of-concept for such move-by-move analysis and the future use of such analyses, in experiments like those suggested in Section 5 or in very different work, seems a good investment of research time and effort.

References

[1] J.R. Anderson. The adaptive nature of human categorization. *Psychological Review*, 98(3):409, 1991.

[2] A.N. Sanborn, T.L. Griffiths, and D.J. Navarro. Rational approximations to rational models: alternative algorithms for category learning. *Psychological Review*, 117(4):1144, 2010.

A Appendix



What's this?

You will be asked to sort a set of images into groups of similar images. They will be abstract, pixelated, black-and-white images.

Sorting

Click and drag images from the "Next Image" box into a group on the screen.

Adding Groups

Click "Add Group" on the left to create a new group. You can divide the images between as many as 8 groups. Once you place an image in a group, you may not move it again.

Mechanical Turking

Once you have placed the last image from the "Next Image" box on the board, a "Finished" screen will be shown. You can copy the trial ID shown here into your Mechanical Turk form in order to be paid for your work.

Thanks for your help!

Figure 6: The instructions presented to the Mechanical Turk worker before starting a trial.