Effect of Glyph Design on Probabilistic Categorization Accuracy

Aditeya Pandey*
Northeastern University

Peter J. Bex†
Northeastern University

Michelle A. Borkin‡
Northeastern University

Abstract

Categorization tasks are common in everyday life, from sorting objects to a doctor diagnosing a patient's disease. In many categorization tasks, classification information is visually represented. Past work in psychology and information visualization has shown that anthropomorphic representations of data can aid in the quick understanding and recall of information. Categorization tasks can utilize this phenomenon and visually represent multidimensional binary classification information (e.g., symptom present/absent in a medical diagnosis) with anthropomorphic glyphs. However it remains to be investigated if anthropomorphic visualizations continue to be beneficial when conveying abstract information that is not directly related to parts of the human body. We study the effects of anthropomorphic and abstract glyph designs on the accuracy of abstract probabilistic categorization tasks. In our within-subject evaluation, 480 participants categorized two of four different glyph visualizations each of which encode 3 abstract probabilistic features. We hypothesized that if visual representation affects accuracy then anthropomorphic glyphs would lead to higher categorization accuracy. However, contrary to our hypothesis, subjects were significantly more accurate at categorization with the most abstract glyph design.

Index Terms: Human-centered computing—Visualization—Visualization techniques—Data Glyphs; Human-centered computing—Visualization—Visualization design and evaluation methods

1 Introduction

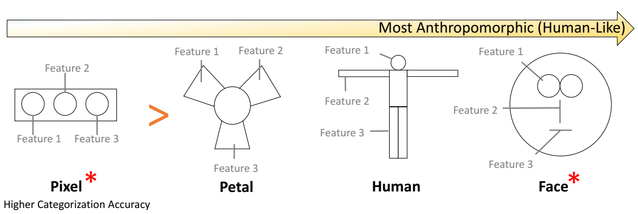

Figure 1: The glyphs evaluated in our study encode three variables as features. In our study we found that the most abstract glyph ('Pixel') was significantly(*) more accurate than the 'Face' glyph in a probabilistic categorization task.

Categorization is the classification of objects based on their features. In many categorization tasks, features of an object can occur in association with more than one category. For example, in a medical diagnosis a physician needs to decide whether a patient's symptom of having a headache is due to a tumor or simply exhaustion. Categorization tasks in which object features are associated with categories probabilistically are called probabilistic categorization tasks. Probabilistic categorization has been extensively studied for the development and testing of formal models of learning and memory (e.g., [1, 5]). In categorization studies the use of visual features are common but their design is usually not justified nor considered a factor in the experiment. In a human centered task like categorization, use of visual stimuli without knowledge of their perceptual effects, or whether visual encodings effect categorization task performance, may give rise to erroneous results. We aim to fill this gap by studying the effect of visual representation on categorization accuracy.

Probablistic categorization requires synthesis of many dimensions of data to determine its category. Glyphs [2] are designed to combine multiple data dimensions and represent them as an object. Therefore, data glyphs are an effective method to visually communicate categorization data. One subset of glyphs that are particularly of interest to categorization tasks, as humans are especially skilled at differentiating human faces [6], are schematic representations of anthropomorphic (human-like) objects such as Chernoff Faces [4]. Anthropomorphic representations may have different shapes to represent individual features, to ensure similarity with natural objects. Additionally, work on the memorability of natural images [7] and data visualizations [3] has demonstrated that the inclusion of natural images results in improved memorability of the figures and visualizations. Consequently, we **hypothesize**:

A human-like or anthropomorphic glyph will aid in learning and recall of the categorization rule resulting in a higher categorization accuracy.

To test our hypothesis, we evaluated the effectiveness of anthropomorphic glyphs as compared to abstract glyphs (see Fig. 1). Two of the glyphs were of abstract design (Pixel and Petal) and two of the glyphs were human-like (Human and Face) so that we could observe whether there was a positive benefit to the more anthropomorphic glyphs.

2 Experiment

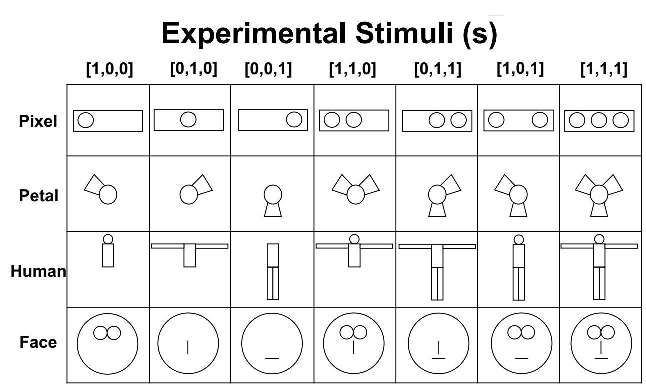

Figure 2: Columns represent all the unique permutations of a three dimensional binary featured dataset. Rows show the corresponding visual encoding of the features with the four glyph designs.

Experiment Data and Stimuli: In probabilistic categorization a stimulus is made up of one or more features. In our study, we use three dimensional features (see Fig. 1), and all the features are of binary data type, i.e., they can take 0/1 value. Three dimensional binary features can have 2^3 unique permutations. Out of the 2^3 permutations, the stimulus representation where no features are present ([0,0,0]) provides no information to categorize a concept and can be visually confusing. Consequently, the feature with no stimulus is removed allowing $2^3 - 1$ total permutations as shown in Fig. 2. We provide a neutral premise and wording to the categorization task. For example, in Fig. 1, all of the features are marked as 1, 2, or 3.

*e-mail:pandey.ad@husky.neu.edu
†e-mail: p.bex@northeastern.edu
‡e-mail: m.borkin@northeastern.edu

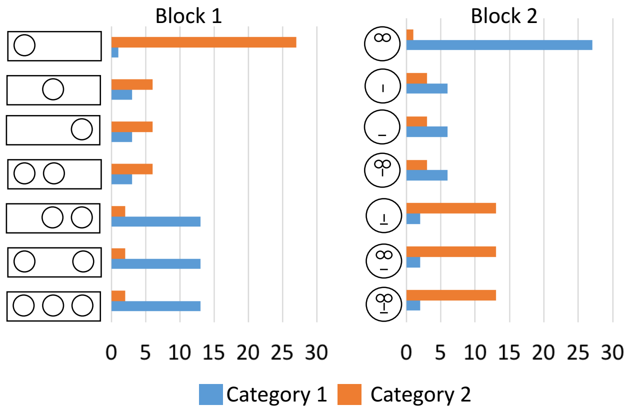

Figure 3: Frequency Distribution in Categorization Task: In a categorization task, probability is communicated as frequencies. For example, in (Block1) stimulus [1,1,1] appears 29 times in the training block of 100 samples, and it appears 27 times with Category 2 and 2 times with Category 1.

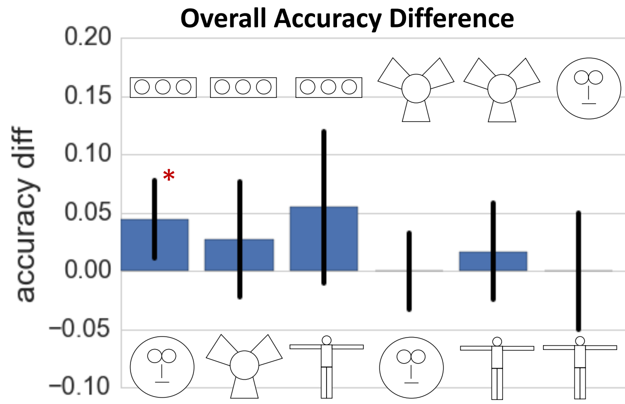

Figure 4: Summary of the average differences in accuracy between each pair of glyphs evaluated in the study. On the y-axis, positive ratios denote that glyphs on the top of the chart had greater accuracy, and visa versa for negative. For each glyph comparison, the 99.9% C.I. is plotted, and asterisks (*) denote Bonferroni-corrected significance in accuracy of one stimulus over other.

Design and Procedure: We use a within-subject experiment design to study if the same person has different categorization ability with different glyphs. In each within-subject study, subjects complete two categorization tasks. For example, in Fig. 3 a participant completes a categorization task with Pixel Glyph (Block 1) and Face Glyph (Block 2). Each task consisted of the following steps: training instruction, training, testing instruction, testing and optional strategy and demographic survey. We train subjects probabilities on a trial-by-trial basis using a frequency distribution. In the training block we show stimuli and the category label. In the testing step, the same set of glyphs were repeated in a random order without the category label, and participants had to choose the likely category based on the training phase. For the strategy survey, we ask participants to explain the strategy rule they adopted in the task.

Participants: A total of 480 study participants (mean age=35, and gender participation of 48.7% women and 51.3% men) were recruited through Amazon's Mechanical Turk. The experiment was conducted in a single 15 minute session. Study participants were monetarily compensated and received $2.00 for their participation.

3 Results

Our experiment comprises 6 pairwise comparisons of categorization accuracy for each of the 4 glyphs. The results are summarized in Fig. 4. Contrary to our hypothesis, we found participants were significantly more accurate with abstract than anthropomorphic glyphs. The ***Pixel glyph visual encoding generated the most precise categorization performance and led to statistically significantly higher accuracy than the Face glyph***.

Post-Hoc Strategy Analysis: A categorization strategy is a rule people use to predict the category for an object. Gluck et. al. [5] define three strategies that participants employ in categorization tasks:

1. Multi-Cue Strategy: People perform inclusive categorization in which they use all the features to read the visual stimulus.

2. Singleton Strategy: People learn one 'primary' stimulus and based on this stimulus guess categories for other stimuli depending on how similar or different they are from the primary.

3. Single-Cue Strategy: People categorize on the basis of presence or absence of a single feature in the stimulus.

Gluck [5] also showed that participants who used a Multi-Cue strategy had the highest categorization accuracy, and those who used the Single-Cue had the lowest accuracy [5]. The result of our analysis indicates that participants used an optimal categorization strategy with abstract glyphs and a sub-optimal strategy with anthropomorphic glyphs. With anthropomorphic glyphs, users learn categorization rules with an inherent bias towards visually salient features, which leads to a non-optimal category rule. We found that with the Face glyph a significant number of participants constructed a categorization rule, e.g., if eyes are present as a feature it belongs to one category and otherwise not (Single-Cue strategy). In contrast, people pay equal attention to all features and use Multi-Cue strategy for categorization with the Pixel glyph, but not the Face glyph.

4 Conclusion

Probabilistic categorization tasks are common in everyday life. We found that the visual encoding of probabilistic categorization data as glyphs can affect human performance for learning and performing categorization tasks. To find an effective data representation, we evaluated four glyph designs ranging from abstract to anthropomorphic. We hypothesized that glyphs, which are more human-like, would lead to higher categorization accuracy. Contrary to our hypothesis, our results show that abstract glyphs lead to higher categorization accuracy. Through a post-hoc analysis of quantitative and qualitative experimental data, we learned that human-like glyphs introduce biases as people relate differently to anatomically salient features. Based on these results we propose if the categorization task requires equal attention for all features, it is essential that glyph designers use an encoding in which all features are equally perceptually salient.

References

[1] A. R. Aron, D. Shohamy, J. Clark, C. Myers, M. A. Gluck, and R. A. Poldrack. Human midbrain sensitivity to cognitive feedback and uncertainty during classification learning. *Journal of neurophysiology*.

[2] R. Borgo, J. Kehrer, D. H. S. Chung, E. Maguire, R. S. Laramee, H. Hauser, M. Ward, and M. Chen. Glyph-based Visualization: Foundations, Design Guidelines, Techniques and Applications. In *Eurographics 2013 - State of the Art Reports*. The Eurographics Association, 2013.

[3] M. A. Borkin, A. A. Vo, Z. Bylinskii, P. Isola, S. Sunkavalli, A. Oliva, and H. Pfister. What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2306–2315, Dec. 2013.

[4] H. Chernoff. The use of faces to represent points in k-dimensional space graphically. *Journal of the American Statistical Association*, 68(342):361–368, 1973.

[5] M. A. Gluck, D. Shohamy, and C. Myers. How do people solve the "weather prediction" task?: Individual variability in strategies for probabilistic category learning. *Learning & Memory*, 9(6):408–418, 2002.

[6] D. Homa, B. Haver, and T. Schwartz. Perceptibility of schematic face stimuli: Evidence for a perceptual gestalt. *Memory & Cognition*, 4(2):176–185, 1976.

[7] P. Isola, D. Parikh, A. Torralba, and A. Oliva. Understanding the intrinsic memorability of images. In J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, and K. Q. Weinberger, eds., *Advances in Neural Information Processing Systems 24*, pp. 2429–2437. Curran Associates, Inc., 2011.