Effect of Anthropomorphic Glyph Design on the Accuracy of Categorization Tasks

Aditeya Pandey
pandey.ad@northeastern.edu
Northeastern University
Boston, MA, USA

Peter Bex
p.bex@northeastern.edu
Northeastern University
Boston, MA, USA

Michelle A. Borkin
m.borkin@northeastern.edu
Northeastern University
Boston, MA, USA

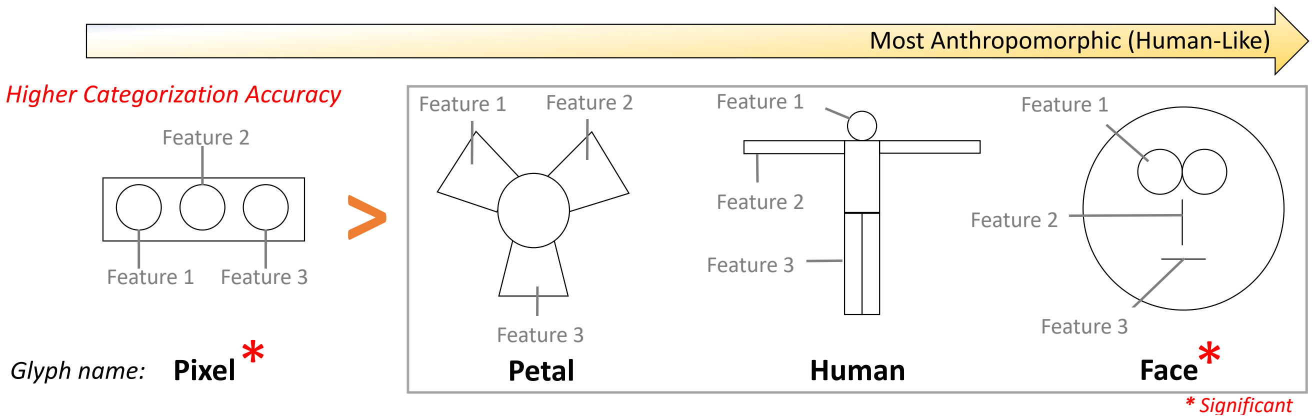


Figure 1: From left to right, glyph designs with an increasing level of anthropomorphic characteristics. In our study we found strong evidence that the least anthropomorphic glyph ('Pixel') was more accurate than the 'Petal,' 'Human,' and 'Face' glyphs for a probabilistic categorization task.

ABSTRACT

Data glyphs continue to gain popularity for information communication. However, the cognition and perception theory of glyphs is largely unknown for many tasks including "categorization". Categorization tasks are common in everyday life from sorting objects to a doctor diagnosing a patient's disease. However it is unknown how glyph designs, specifically anthropomorphic human-like representations which in prior visualization research have demonstrated improved information recall, affect accuracy in a categorization task. To better understand how people comprehend and perceive glyphs for categorization, including anthropomorphic representations, we conducted a crowdsourced experiment to evaluate whether more human-like glyphs would lead to higher categorization accuracy. Contrary to our hypothesis, we found evidence that subjects are more accurate with a less anthropomorphic glyph. A posthoc analysis also reveals that anthropomorphic glyphs introduce biases due to their anatomically salient features. Based on these results we propose design guidelines for glyphs used in categorization tasks. The supplemental material of this paper available is on https://osf.io/3bgcv/.

CHI '22 Extended Abstracts, April 29-May 5, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9156-6/22/04…$15.00
https://doi.org/10.1145/3491101.3519748

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**.

KEYWORDS

Data Glyph, Probabilistic Categorization Task, Information Visualization, Quantitative Studies

ACM Reference Format:
Aditeya Pandey, Peter Bex, and Michelle A. Borkin. 2022. Effect of Anthropomorphic Glyph Design on the Accuracy of Categorization Tasks. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '22 Extended Abstracts), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3491101.3519748

1 INTRODUCTION

Data glyphs [3] are a visualization encoding which combines multiple data dimensions and represents them as an object. For example, Fig. 2(Left) depicts diverse aspects of regional well-being using the star glyph [32], and Fig. 2(Right) shows state-wise election results as a Chernoff face with different features of the face summarizing election results [33]. The design of data glyphs offer creative freedom to visualization practitioners leading to charts beyond basic encodings such as bar and line charts [6] (Fig. 2). Consequently, over the past few years, data glyphs have become a popular technique for creating bespoke visualizations [32, 33]. While glyphs continue to grow in popularity and applications, the factors that contribute to their perception and cognition still remain largely unknown and not sufficiently studied in the visualization community [13]. With a deeper understanding of how people interpret and understand glyph encodings, this new theory will enable the creation of more effective glyph representations including optimal designs for specific tasks or use-cases. Therefore, in this paper, we take a step in this direction to generate empirically-driven design guidelines for data glyphs.

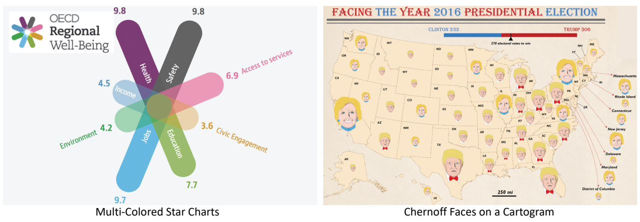


Figure 2: Left: A multi-colored star glyph depicting different aspects related to regional well-being [32].Right: Chernoff faces overlaid on a cartogram to show the results of 2016 presidential election [33].

One specific type of task not previously studied in the context of data glyph design is "categorization." In a **categorization task**, a person classifies objects based on their **features**. For example, a physician performs a categorization task when they diagnose a patient as sick or healthy based on their symptoms as features in a medical diagnosis [29]. In machine learning, categorization tasks can be helpful to categorize data into clusters based on their dimensions [20]. *Data glyphs are a particularly appropriate and effective method to visually communicate categorization data because categorization tasks require the synthesis of these many dimensions of data to determine its category.* Categorization tasks are also usually **probabilistic**. For example, in the medical diagnosis case, if a patient exhibits the symptom of having a headache there are different probabilities that the symptom may be indicative of them suffering from fatigue versus a tumor. Formally, *probabilistic categorization tasks* are defined as tasks in which object features are associated with categories probabilistically [11]. In this paper we focus on probabilistic categorization tasks because of their broad applicability as well as lack of prior work and unknown theory pertaining to its associated glyph design.

One subset of glyphs that are of particular relevance to categorization tasks are schematic representations of anthropomorphic (human-like) objects such as Chernoff Faces [9]. Humans are especially skilled at differentiating faces because they enter memory through two distinct pathways per the dual-coding theory: visual and verbal, thus enabling easy recall [17, 27]. Additionally, work on the memorability of natural images [18] and data visualizations [4, 5] has demonstrated that the inclusion of natural images and human-recognizable pictures results in improved memorability and recall of figures and visualizations. But, how does a human-like glyph representation impact the human perception and cognition of the encoded data? Can a human-like encoding, through memorability and salience, improve categorization task accuracy? Consequently, we **hypothesize**: *A human-like or anthropomorphic glyph will aid in learning and recall of a categorization rule resulting in a higher categorization accuracy.*

To test our hypothesis, we conducted a within-subject study with 480 participants on Amazon's Mechanical Turk. Each participant completed a probabilistic categorization task with two of four different glyph designs (Fig. 1) and we observed whether there was a positive benefit to the more anthropomorphic glyphs. Contrary to our hypothesis, we found strong evidence that the non-anthropomorphic glyph visual encoding was more accurate than the anthropomorphic glyphs. Additionally, participants felt less confident with anthropomorphic glyphs to complete the categorization task. To gain further insight into our study results, we conducted a posthoc data analysis. The results of the analysis indicate with anthropomorphic glyphs, participants learn categorization rules with a bias towards visually salient features and tend to ignore data represented as non-salient features in the categorization tasks. For example, in a human face glyph a participant will focus more on the data encoded with "eyes" mark.

Contributions: We present the first empirical study to measure the effect of glyph design on probabilistic categorization accuracy. The results of the study demonstrate that glyph representation can affect task completion accuracy. Through a posthoc data analysis, we explain how categorization strategy differs between glyphs based on the particular design. With our results as premise, we also recommend specific glyph design idioms for categorization tasks.

2 RELATED WORK

Data representation in categorization tasks: Categorization tasks in which object features are associated with categories probabilistically are called probabilistic categorization tasks [11]. For example in our medical diagnosis case a headache is a feature (symptom) related to both diagnoses of a brain tumor and fatigue. However, in general a headache is more likely caused by fatigue and less likely a result of brain tumor. Most real world categorization tasks are probabilistic and require selection among non-deterministic categories [11]. As a result, probabilistic category learning has been extensively studied for the development and testing of formal models of learning and memory (e.g., [1, 14, 15, 25]). In category learning studies, researchers use a visual encoding to represent the features. For instance, in a weather prediction task [14, 15], authors present each feature with a card that had a particular geometric pattern. Aron et al. [1] used a potato head glyph (similar to the anthropomorphic glyphs in this paper) with features mapped on a hat, eyeglasses, mustache, and bow tie, and the glyph had no direct relation to the task. Other studies focused instead on the task and used abstract glyphs, e.g., Shepard et al. [30] used shape, size, and color to encode features. Per this related work, the use of visual features is common in probabilistic categorization, but their design is rarely justified. Therefore, in our study we design four glyphs inspired by existing categorization studies and then study how glyph design affects categorization task performance.

Categorization Strategies: In a categorization experiment, the strategy used by a participant to predict the category of a visual stimulus, which represents a real-world categorization object, is called a categorization strategy [2, 15, 25, 28]. Here we will discuss a recently published set of strategies by Gluck et. al. [15]. Gluck's strategies were derived from trends observed in subject responses in probabilistic categorization tasks:

1. Multi-Cue Strategy: People perform inclusive categorization in which they use all the features to read the 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 et al. [15] showed that the Multi-Cue strategy had the highest categorization accuracy, and the Single-Cue had the lowest accuracy [15]. In Sec. 6, we discuss these categorization strategies in the context of our study's results. In the evaluation of glyph designs, categorization strategies can explain how participants perceive glyphs, which can be useful for the identification of appropriate visual encodings for categorization tasks. These categorization strategies are explained in more detail in the accompanying Supplemental Material.

Glyph Visualization: Glyphs are visual objects used to represent multidimensional datasets, and they map one or more attributes of the data onto one or more of an object's visual marks. A classic example of a glyph visualization is the Chernoff Face[9], where features of the face like eyes, nose, and mouth are used to represent a dataset. The glyph design space is large, and their usability is widely studied [13]. However, despite the large design space of glyphs, their evaluation remains an open topic. For example, few studies have investigated the differences between anthropomorphic glyphs such as faces versus non-anthropomorphic representations for visualization tasks [7, 22, 24, 26, 34]. However, these studies do not focus on a probabilistic categorization task. Moreover, Fuchs et al. [13] conducted a meta-analysis of these studies and concluded that the results from these evaluations are contradictory and efficacy of a glyph design is dependent on the data and the task. Therefore, in our work, we evaluate glyphs for categorization tasks and contribute further to understanding differences between anthropomorphic and non-anthropomorphic glyph designs.

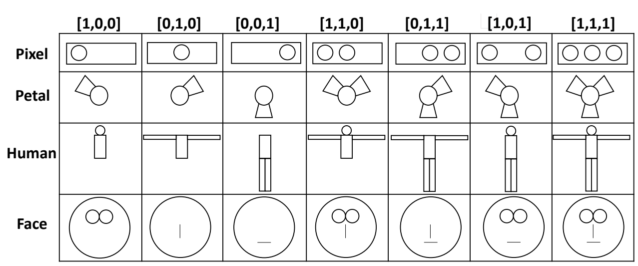


Figure 3: Columns represent all the unique permutations of a three dimensional binary featured dataset. Rows show the corresponding visual encoding of the features with the four glyph designs.

3 MULTIDIMENSIONAL STIMULI AS GLYPHS

In categorization tasks, 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 take a 0/1 value. The visual encoding maps the binary value (0 or 1) to the absence or presence of a feature (Fig. 3). This is common in medical diagnoses, where doctors use a simple binary conditional logic when dealing with symptoms [21]. For example, the logic may look like: if a headache(*symptom*) then fatigue(*diagnosis*). 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. 3. In our visual encoding, each feature is tied to a visual mark of the glyph.

The overarching motive of our study is to compare anthropomorphic and non-anthropomorphic glyphs because these two classes of glyph designs are commonly used in the probabilistic categorization literature (Sec. 2). According to our hypothesis, anthropomorphic glyphs will be more conducive for probabilistic categorization tasks. Below, we explain the glyph encodings used in our study.

Pixel: The abstract Pixel glyph uses point (circle) marks which resemble three pixels. This glyph includes a rectangle as a frame of reference around the circles. These circles are placed equidistant from each other arranged horizontally linear. This design uses the position of the circle to encode data, i.e., one feature per circle.

Petal: The abstract Petal glyph utilizes a radial layout with shape marks ("petals") around a common center point. Our petal glyph is similar to the flower [8, 23] and star glyphs [12, 22], as both of them show features in a radial layout and use position and length to encode data on the individual marks.

Human: The anthropomorphic human glyph is of a human body form in which each data dimension is encoded to a anatomical part. The three features in our study were mapped to the head, the arms, and the legs. The torso serves as a visual reference point, and it is visible in conjunction with all features. An advantage of the Human glyph is it's visual saliency in which each data encoding mark is substantially different.

Face: The anthropomorphic Face glyph encodes each data dimension to an anatomical facial feature. The three features in our study were mapped to the eyes, the nose, and the mouth. For the face stimuli, we avoided using previously published Chernoff [9] and Kabulov [19] face glyphs because their facial features represent emotion. For instance, the mouth/lips of Chernoff can show a frown or smile, based on how data is encoded. Emotional expression may affect the perception of face glyph positively or negatively. We avoid such effects by the use of a neutral Face glyph design.

4 EXPERIMENT DESIGN AND METHODS

In this section we present our experimental design and methods used to measure and compare the performance of human subjects across different glyphs for a probabilistic categorization task. As we are interested in measuring the performance difference between different glyph representations, our independent variable in the categorization task is the glyph design and our dependent variable is the categorization accuracy. This study uses an abstract categorization task. For example, in Fig. 1, all of the features are marked as 1, 2, or 3 and not, for example in the Face glyph, "eyes", "nose", and "mouth". We also provide a neutral premise and wording to the categorization task "Your task will be to look at each figure and decide which family it's from" instead of "Your task will be to look at symptoms and decide if a patient is sick or healthy". The rationale for choosing an abstract task versus a domain specific task was to ensure that results can be generalized for categorization domains and domain specific confounds are not introduced in the study. The

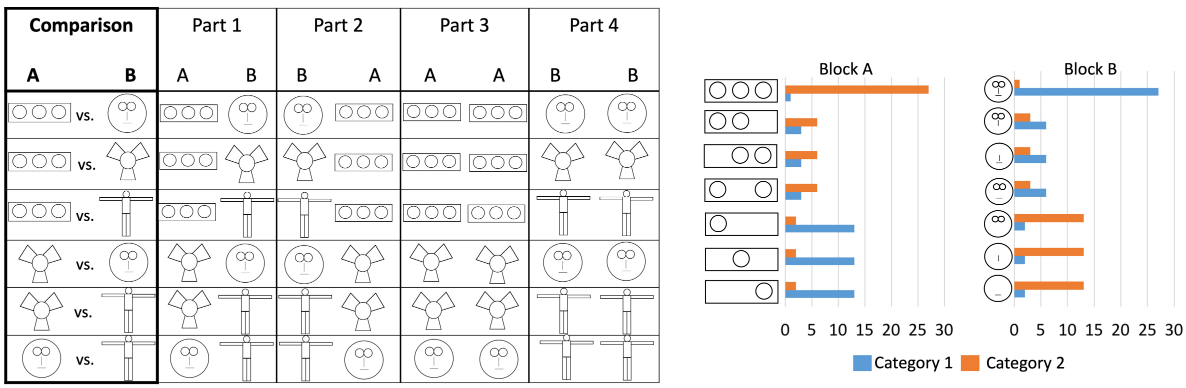


Figure 4: Left: Within-Subject Study Design. Right: In a categorization task, probability is communicated as frequencies. For example, in (Block A) stimulus [1,1,1] appears 29 times in the training block of 100 samples, and it appears 27 times with Category 2 and 2 times with Category 1.

wording used in the experiment is included in the Supplemental Material.

Participants: A total of 480 study participants (mean age=35; 48.7% women and 51.3% men) were recruited through Amazon's Mechanical Turk. All participants were based in the United States of America and had a 95% or greater approval rating. Before starting the experiment, participants indicated their informed consent in accordance with the guidelines of our organization's IRB and our approved study protocol. The experiment was conducted in a single 15 minute session. Study participants were monetarily compensated and received $2.00 for their participation.

4.1 Within-Subject Study Design

We used a within-subject experiment design in order to observe if the same person has different categorization abilities with different glyphs. Our study design consists of six comparisons as shown in Fig. 4(Left) (Comparison) with each part broken down into four separate within-subject studies, Fig. 4(Left) (Parts 1 - 4). Parts 1 and 2 of the comparison were designed to balance the effect of presentation order. *Parts 3 and 4 had participants repeat the experiment with the same glyph design but with different probabilities. These Parts 3 and 4 aim to counter carryover effects of learning and fatigue in a within-subject study.* In a categorization study setup it is possible for participants to learn an optimal categorization strategy with the first glyph and subsequently use the improved strategy with the second glyph. In some cases fatigue may occur and participants may lose motivation affecting their accuracy in the second task. We carefully and accurately take these carryover effects into account in the final analysis (Sec. 5).

In each within-subject study, subjects complete two categorization tasks. For example, in Fig. 4(Left) (Row 1, Part 1) a participant completes a categorization task with Pixel Glyph (Block A) and Face Glyph (Block B). We designed Blocks A and B to have different probability structures of equal difficulty to prevent subjects from directly mapping probability structures from Block A to Block B. Frequency distribution obtained from this probability structure is shown in Fig. 4(Right). More details on the calculation of the frequency distribution are available in the Supplemental Material.

4.2 Procedure

For each within-subject study, Fig. 4(Left) (Parts 1-4), participants had to complete two categorization tasks. Each task consisted of:

1. Training Instructions: In the training block instructions, study participants were introduced to a glyph representation and told that they would learn how to assign each glyph to one of two families. Most importantly, participants were instructed A) that each glyph could occur in each family but would be more common in one family than the other, and B) that approximately half of the figures were from "Family 1" and half from "Family 2". Participants were instructed that they would be shown a set of glyphs with their respective family labels and that the participant should try to learn which glyphs occur in each family.

2. Training Phase: In this phase, participants saw 100 samples randomly drawn from the frequency distribution shown in Fig. 4(Right). For each sample, stimulus and category label were visible. The glyph was displayed with the true label, and the screen remained visible until the subject advanced to the next glyph instance.

3. Testing Instructions: This step consisted of a single screen instructing the participant that they would next view a series of stimuli (all of the same glyph encoding) and categorize them. Participants were asked to press on their keyboard the number "1" for Family 1 and "2" for Family 2 to pick the correct category.

4. Testing Phase: Subjects completed 100 trials of the testing phase. In the testing block, the glyph was visible as in training; however, the correct categorization Family label was unmarked. In this way, participants could not continue learning the probability distribution and enable us to evaluate how well they learned how to categorize features. On each trial, we recorded participant response. After the testing phase, participants completed a NASA-TLX [16] based survey to measure cognitive workload measures associated with the categorization task.

At the end of the experiment, participants were asked to voluntarily provide their age, gender, and any patterns they noticed or strategies they used to complete the task. All categorization experiment procedures and stimuli are provided in Supplemental Material.

5 RESULTS

Accuracy Differences: Our experiment is comprised of six pairwise comparisons (Fig 4 (Left)) of categorization accuracy for each of the four glyphs (Pixel, Petal, Human, and Face). In each comparison, there are four parts with two categorization tasks (Fig. 4(Right)). In a categorization task, the accuracy is calculated as the proportion of correct responses, i.e., the number of trials on which the subject correctly categorized a stimulus divided by the total number of trials. Next, we take a median of accuracy across all subjects who performed the same categorization task to obtain accuracy per glyph design. Finally, the median accuracies are subtracted to find the accuracy difference between glyph designs. The pairwise accuracy difference between glyph designs are summarized in Fig. 5(Left). A stepwise calculation of the accuracy difference along with the data analysis code is included in the Supplemental Material.

Significance Test: From the results in Fig. 5(Left), we can see that accuracy with the Pixel glyph is higher than that of the other three glyphs and that the Petal glyph has slightly higher accuracy than the Human glyph. To test for the significance of these differences, we derived the null hypothesis distribution for each glyph empirically by bootstrapping [10]. Bootstrapping allows us to estimate the variability in distribution more confidently and provides a more realistic estimation of the population mean by using repeated sampling from the collected data. We assume each subject's number of correct responses for each stimulus in each block to be binomially distributed. Next, 10,000 samples at random were drawn from these distributions. We carried out the analyses above on each of these estimates, resulting in a set of 10,000 estimates of accuracy difference. To handle multiple comparisons, we use Bonferroni corrected 99.9% confidence intervals for each distribution. The Confidence Intervals (CI) for each accuracy difference are shown by the error bars in Fig. 5(Left).

Contrary to our hypothesis, we found strong evidence that participants were more accurate with non-anthropomorphic glyphs as compared to anthropomorphic glyphs. The ***Pixel glyph visual encoding generated the most precise categorization performance and led to statistically significantly higher accuracy than the Face glyph***.

Task Load Survey Results: Our NASA-TLX based survey results are shown in Fig. 5(Right). We excluded responses to the questions of temporal effect and physical effort from analysis in our study due to their lack of relationship to our study's task. The Pixel and Petal glyph consistently recorded the lowest average value for the responses, indicating they were perceived to be the easiest to use to complete the task. To test for statistical significance, we used a non-parametric Kruskal Wallis test with a Bonferroni corrected alpha level of 0.01. The statistical tests yielded a non-significant result, i.e., $p > 0.01$ for all NASA-TLX questions.

6 DISCUSSION

In the present study we asked, "How should the data be visually represented to maximize categorization accuracy?". Our research explores a small but meaningful visual design space to understand the differences in categorization accuracy. Our results show that, contrary to our expectations based on published literature, we found that non-anthropomorphic glyph designs have higher categorization accuracy. To explain this trend, we conducted a posthoc strategy analysis to determine the categorization rule adopted by the participants in the categorization task as discussed in Sec. 2. For the posthoc analysis we use self reported strategy data (Sec. 4.2, Demographic Survey and Qualitative Feedback). Based on this strategy analysis, we speculate that participants used an optimal categorization strategy (Multi-Cue strategy) for the Pixel glyph and a sub-optimal strategy (Single-Cue) for the Face glyph. Our speculation is backed by comments like "'I thought that the heads belonged mainly to one family and same with the eyes" where participants explicitly point out that their strategy focused on looking for visually salient features such as eyes and heads in the glyphs. More specifically, we found that the "eye" feature is processed differently than the other elements of the Face glyph. Based on their anatomical significance, the eyes have a pop-out effect which aids the pre-attentive processing [31], i.e., eyes attract more attention than other features of the Face glyph. Consequently, eyes bias the categorization strategy of participants.

In the self-reported feedback, we found similar patterns exist for the head feature in the Human glyph. We predicted that anthropomorphic glyphs would aid memory and guide participants to higher categorization accuracy. Instead, we see the opposite effect: our anthropomorphic glyphs negatively affect performance and lead to biased information processing.

Based on the results of the posthoc strategy analysis, we propose two design recommendations for probabilistic categorization tasks:

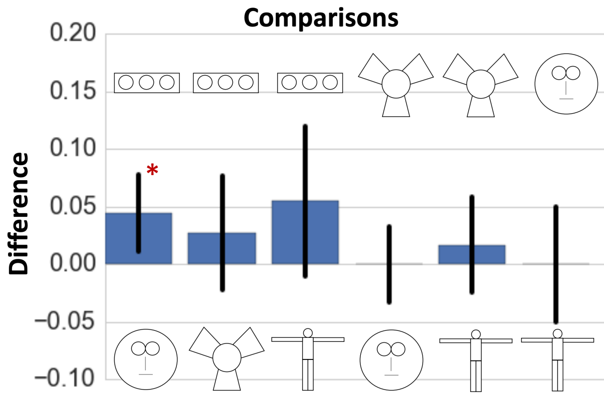
1. If the categorization task requires equal attention for all features, it is essential that glyph designers use an encoding in which all features are equally perceptually salient.

2. Alternatively, if the categorization task includes a subset of features that should be weighted more heavily in the observer's decision, such as in the doctor example indicating a symptom which is critical or deadly, then glyph designs with very salient or significant features (e.g., anthropomorphic designs) could be used to take advantage of human pre-attentive selection processes.

Our study takes an essential step in the direction of general results for glyph encoding performance with categorization tasks, including a comparison of abstract glyphs with anthropomorphic glyphs. Our study found that participants performed the categorization task significantly worse with the most anthropomorphic (Face) glyph. These results naturally lead to the question "Do other glyph encodings with anthropomorphic or natural features follow our results?" For example, if we use a dog-shaped glyph, will the dog head or face be processed differently than a human? What about a car-shaped glyph? These differences are hard to predict as there is significant variation in the participation strategy and may only be addressed with detailed future empirical studies. Meanwhile, our results can serve as a heuristic for glyphs where the design does not match the stimuli used in the experiment.

7 CONCLUSION

We found that the visual encoding of probabilistic categorization data as glyphs can affect human performance for completing categorization tasks. To find an effective glyph representation, we



Question	Pixel	Petal	Human	Face
How mentally demanding was the task? (1 = low, 7 = high)	4.09	4.08	4.54	4.26
How successful were you in accomplishing what you were asked to do? (1 = perfect, 7 = failure)	3.81	3.97	3.93	3.95
How hard did you have to work to accomplish your level of performance? (1 = low, 7 = high)	4.56	4.42	4.77	4.77
How insecure, discouraged, irritated, stressed, and annoyed were you? (1 = low, 7 = high)	3.36	3.19	3.42	3.72

Figure 5: Left: 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% CI is plotted, and asterisks (*) denote Bonferroni-corrected significance in accuracy of one stimulus over other. Right: Average subjective data responses to select NASA-TLX rated on a 7-point Likert scale.

evaluated four glyph designs ranging from abstract to anthropomorphic. Contrary to our hypothesis, our results show that human-like features negatively affect categorization accuracy. Through a posthoc 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 encoding in which all features are equally perceptually salient. In future work we hope to investigate if there are effects from other glyph designs, as well as conduct *in situ* evaluations with real decision making scenarios (e.g., medical diagnoses). We also hope the visualization community will further explore visualizations in support of probabilistic categorization learning and task completion in support of effective decision making.

ACKNOWLEDGMENTS

This work was supported in part by a Northeastern University Tier 1 Grant, and by the Khoury College of Computer Sciences.

REFERENCES

[1] Adam R Aron, Daphna Shohamy, Jill Clark, Catherine Myers, Mark A Gluck, and Russell A Poldrack. 2004. Human midbrain sensitivity to cognitive feedback and uncertainty during classification learning. *Journal of neurophysiology* 92, 2 (2004), 1144–52. https://doi.org/10.1152/jn.01209.2003

[2] F Gregory Ashby and Ralph E Gott. 1988. Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14, 1 (1988), 33–53. https://doi.org/10.1037//0278-7393.14.1.33

[3] Rita Borgo, Johannes Kehrer, David H. S. Chung, Eamonn Maguire, Robert S. Laramee, Helwig Hauser, Matthew Ward, and Min Chen. 2013. Glyph-based Visualization: Foundations, Design Guidelines, Techniques and Applications. In *Eurographics 2013 - State of the Art Reports*, M. Sbert and L. Szirmay-Kalos (Eds.). The Eurographics Association, -. https://doi.org/10.2312/conf/EG2013/stars/039-063

[4] Michelle A. Borkin, Zoya Bylinskii, Nam Wook Kim, Constance May Bainbridge, Chelsea S. Yeh, Daniel Borkin, Hanspeter Pfister, and Aude Oliva. 2016. Beyond Memorability: Visualization Recognition and Recall. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 519–528. https://doi.org/10.1109/TVCG.2015.2467732

[5] Michelle A. Borkin, Azalea A. Vo, Zoya Bylinskii, Phillip Isola, Shashank Sunkavalli, Aude Oliva, and Hanspeter Pfister. 2013. What Makes a Visualization Memorable? *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2306–2315. https://doi.org/10.1109/TVCG.2013.234

[6] Matthew Brehmer, Robert Kosara, and Carmen Hull. 2021. Generative Design Inspiration for Glyphs with Diatoms. arXiv:2107.09015 [cs.HC]

[7] Elizabeth J. Casey. 1986. Visual Display Representation of Multidimensional Systems: The Effect of Information Correlation and Display Integrality. *Proceedings of the Human Factors Society Annual Meeting* 30, 5 (1986), 430–434. https://doi.org/10.1177/154193128603000504 arXiv:https://doi.org/10.1177/154193128603000504

[8] Michael Chau. 2011. Visualizing Web Search Results Using Glyphs: Design and Evaluation of a Flower Metaphor. *ACM Trans. Manage. Inf. Syst.* 2, 1, Article 2 (March 2011), 27 pages. https://doi.org/10.1145/1929916.1929918

[9] Herman Chernoff. 1973. The Use of Faces to Represent Points in k-Dimensional Space Graphically. *J. Amer. Statist. Assoc.* 68, 342 (1973), 361–368. https://doi.org/10.1080/01621459.1973.10482434 arXiv:https://www.tandfonline.com/doi/pdf/10.1080/01621459.1973.10482434

[10] Bradley Efron and Robert J Tibshirani. 1994. *An introduction to the bootstrap.* CRC press, -.

[11] WK Estes. 1972. Elements and patterns in diagnostic discrimination learning. *Transactions of the New York Academy of Sciences* 34, 1 (January 1972), 84—95. https://doi.org/10.1111/j.2164-0947.1972.tb02662.x

[12] J. Fuchs, P. Isenberg, A. Bezerianos, F. Fischer, and E. Bertini. 2014. The Influence of Contour on Similarity Perception of Star Glyphs. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (Dec 2014), 2251–2260. https://doi.org/10.1109/TVCG.2014.2346426

[13] J. Fuchs, P. Isenberg, A. Bezerianos, and D. Keim. 2017. A Systematic Review of Experimental Studies on Data Glyphs. *IEEE Transactions on Visualization and Computer Graphics* 23, 7 (July 2017), 1863–1879. https://doi.org/10.1109/TVCG.2016.2549018

[14] Mark A Gluck and Gordon H Bower. 1988. From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General* 117, 3 (1988), 227–47. https://doi.org/10.1037/0096-3445.117.3.227

[15] Mark A Gluck, Daphna Shohamy, and Catherine Myers. 2002. How do people solve the "weather prediction" task?: Individual variability in strategies for probabilistic category learning. *Learning & Memory* 9, 6 (2002), 408–18. https://doi.org/10.1101/lm.45202

[16] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (2006), 904–908. https://doi.org/10.1177/154193120605000909 arXiv:https://doi.org/10.1177/154193120605000909

[17] Donald Homa, Bram Haver, and Terrence Schwartz. 1976. Perceptibility of schematic face stimuli: Evidence for a perceptual Gestalt. *Memory & Cognition* 4, 2 (1976), 176–85. https://doi.org/10.3758/BF03213162

[18] Phillip Isola, Devi Parikh, Antonio Torralba, and Aude Oliva. 2011. Understanding the Intrinsic Memorability of Images. In *Advances in Neural Information Processing Systems 24*, J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, and K. Q. Weinberger (Eds.). Curran Associates, Inc., -, 2429–2437. http://papers.nips.cc/paper/4451-understanding-the-intrinsic-memorability-of-images.pdf

[19] BT Kabulov. 1992. A method for constructing chernoff faces oriented toward interval estimates of the parameters. *Soviet journal of computer and systems sciences* 30, 3 (1992), 94–97.

[20] Dietrich Kammer, Mandy Keck, Thomas Gründer, Alexander Maasch, Thomas Thom, Martin Kleinsteuber, and Rainer Groh. 2020. Glyphboard: Visual Exploration of High-Dimensional Data Combining Glyphs with Dimensionality Reduction. *IEEE Transactions on Visualization and Computer Graphics* 26, 4 (2020), 1661–1671. https://doi.org/10.1109/TVCG.2020.2969060

[21] Robert S. Ledley and Lee B. Lusted. 1959. Reasoning Foundations of Medical Diagnosis. *Science* 130, 3366 (1959), 9–21. https://doi.org/10.1126/science.130.3366.9 arXiv:https://science.sciencemag.org/content/130/3366/9.full.pdf

[22] Michael D. Lee, Rachel E. Reilly, and Marcus E. Butavicius. 2003. An Empirical Evaluation of Chernoff Faces, Star Glyphs, and Spatial Visualizations for Binary Data. In *Proceedings of the Asia-Pacific Symposium on Information Visualisation - Volume 24* (Adelaide, Australia) *(APVis '03)*. Australian Computer Society, Inc., AUS, 1–10.

[23] Yi-Na Li, Dong-Jin Li, and Kang Zhang. 2015. Metaphoric Transfer Effect in Information Visualization Using Glyphs. In *Proceedings of the 8th International Symposium on Visual Information Communication and Interaction* (Tokyo, AA, Japan) *(VINCI '15)*. Association for Computing Machinery, New York, NY, USA, 121–130. https://doi.org/10.1145/2801040.2801062

[24] Donald MacGregor and Paul Slovic. 1986. Graphic Representation of Judgmental Information. *Hum.-Comput. Interact.* 2, 3 (Sept. 1986), 179–200. https://doi.org/10.1207/s15327051hci0203_1

[25] Douglas L Medin and Marguerite M Schaffer. 1978. Context theory of classification learning. *Psychological review* 85, 3 (1978), 207–238. https://doi.org/10.1037/0033-295X.85.3.207

[26] Juan E. Mezzich and David R.L. Worthington. 1978. A COMPARISON OF GRAPHICAL REPRESENTATIONS OF MULTIDIMENSIONAL PSYCHIATRIC DIAGNOSTIC DATA. In *Graphical Representation of Multivariate Data*, PETER C.C. WANG (Ed.). Academic Press, -, 123 – 141. https://doi.org/10.1016/B978-0-12-734750-9.50010-1

[27] Allan Paivio. 1990. *Mental representations: A dual coding approach*. Vol. 9. Oxford University Press, -.

[28] Michael I Posner and Steven W Keele. 1968. On the genesis of abstract ideas. *Journal of experimental psychology* 77, 3, Pt.1 (1968), 353–363. https://doi.org/10.1037/h0025953

[29] Timo Ropinski and Bernhard Preim. 2008. Taxonomy and Usage Guidelines for Glyph-based Medical Visualization. In *Simulation and Visualization 2008 (SimVis 2008), 18-29 Februar 2008, Magdeburg*, Helwig Hauser, Steffen Straßburger, and Holger Theisel (Eds.). SCS Publishing House e.V., -, 121–138. http://www.simvis.org/Tagung2008/sv-proceedings.html

[30] Roger N Shepard, Carl I Hovland, and Herbert M Jenkins. 1961. Learning and memorization of classifications. *Psychological monographs: General and applied* 75, 13 (1961), 1–42. https://doi.org/10.1037/h0093825

[31] Geen De Soete and Wilfried Do Corte. 1985. On the Perceptual Salience of Features of Chernoff Faces for Representing Multivariate Data. *Applied Psychological Measurement* 9, 3 (1985), 275–280. https://doi.org/10.1177/014662168500900305 arXiv:https://doi.org/10.1177/014662168500900305

[32] Moritz Stefnar and Dominik Baur. 2014. OECD Regional Well-Being. https://truth-and-beauty.net/projects/oecd-regional-wellbeing/

[33] Beichen Tian. 2016. Facing The Year 2016 Presidential Election. https://bit.ly/3q7ZYcE

[34] Leland Wilkinson. 1982. An Experimental Evaluation of Multivariate Graphical Point Representations. In *Proceedings of the 1982 Conference on Human Factors in Computing Systems* (Gaithersburg, Maryland, USA) *(CHI '82)*. Association for Computing Machinery, New York, NY, USA, 202–209. https://doi.org/10.1145/800049.801781