Digitally Training Graph Viewers against Misleading Bar Charts

Claudia Ramly (cmatta2@wisc.edu)[2], Ayon Sen (asen6@wisc.edu)[1], Ved P. Kale (vpkale@wisc.edu)[1], Martina A. Rau (marau@wisc.edu)[2], Xiaojin Zhu (jerryzhu@cs.wisc.edu[1]

[1] Department of Computer Sciences, [2] Department of Educational Psychology
University of Wisconsin-Madison

Abstract

Bar charts are common visual tools used to convey statistical information. Even though bar charts are effective in making abstract concepts more accessible, poorly-designed bar charts - whether designed intentionally or unintentionally - can easily mislead the viewer. For example, a poorly-designed bar chart may only present part of the effective range on the vertical axis. This exaggerates the contrast among bars, leading an unsuspecting graph viewer to wrong conclusions. More broadly, misrepresentation in data visualization is becoming an increasing societal problem contributing to daily misinformation. This paper presents a computational and cognitive solution to this problem. Our idea is to train viewers by showing them a few dozen carefully designed bar charts that are misleading, together with guidance on why these bar charts are misleading. We then test whether the viewers identify similarly misleading bar charts in the future. Importantly, we use neural networks and cognitive models to optimize the training (i.e., the design of those few dozen bar charts). Our experiment shows that perceptual trainings can help viewers not be fooled by similar misleading graphs in the future.

Keywords: bar charts; misleading graphs; perceptual fluency; Neural networks; cognitive model.

Introduction

Graphs such as bar charts are commonly used to communicate statistical information in print, social media, and news media. While such graphs can be informative because they can make abstract concepts more accessible, they can also be misleading if they are poorly designed. For example, if a bar chart's y-axis starts at some value other than zero, it distorts the depicted information by exaggerating the difference between the categories shown by the individual bars (see Figure 1). Formally, we define a graph as misleading if the displayed size of an effect (e.g., the relative height of the two bars in the bar chart in Figure 1) does not correspond to the size of the true effect (Tufte, 2001). For example, in Figure 1 [see left bar chart], the orange bar is about 1.75 times higher than the white bar, whereas 29.4 percent is 42 times larger than 0.7 percent.

Since graph design software has become cost-effective and accessible to all (Tan & Benbasat, 1990), it is easier for anyone without experience to design graphs; and this can lead to intentionally or unintentionally misleading graphs (Cairo, 2019). For example, a review of graphs used in medical advertisements showed that about one third of the graphs presented information in a misleading way (Cooper, Schriger, Wallace, Mikulich, & Wilkes, 2003). Misleading graphs can have serious implications in healthcare, for example, where both patients and physicians already struggle to understand how graphs communicate health risks (Galesic & Garcia-Retamero, 2011). Even if viewers know that a graph is misleading, they still face challenges in extracting the correct answer, further compounding the problem (Harper, 2004). Hence, to prevent such graphical misinformation, viewers need to (1) notice that the graph is distorted, and (2) adjust their process of extracting information from the graph accordingly.

The goal of the present study is to investigate whether viewers can be trained to become perceptually fluent at extracting correct information from misleading graphs. To this end, we developed a perceptual training for misleading graphs. We compare different versions of this training with varying levels of feedback. We hypothesize that providing specific feedback on the correct answer improves viewers' benefit from the perceptual training. In the following section, we first review literature on perceptual fluency, as well as prior work on graph design and graph perception. Then, we discuss prior research on perceptual trainings that counteract perceptual biases — albeit not in the context of misleading graphs.

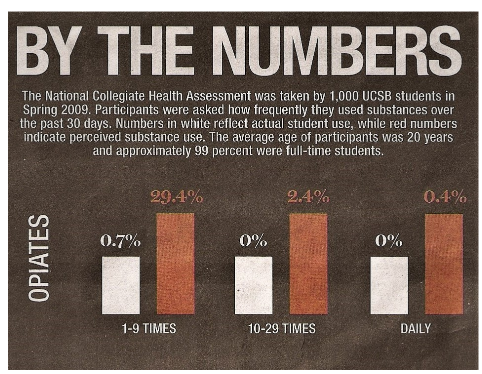


Figure 1: An example of misleading bar charts.

Related Work

Perceptual fluency is the ability to extract information quickly and effortlessly from visuals (Gibson, 2000). One benefit of perceptual fluency is that it frees cognitive resources so individuals can use these essential resources for higher-order thinking (Kellman & Massey, 2013). Perceptual fluency is acquired via nonverbal, inductive learning processes

(Goldstone & Barsalou, 1998). These processes are considered nonverbal because verbal reasoning does not aid but interferes with perceptual learning (Koedinger, Corbett, & Perfetti, 2012). Second, these processes are considered inductive because humans become more effective at recognizing visual patterns based on numerous experiences with visual stimuli (Gibson, 2000; Koedinger et al., 2012). The result is a highly efficient perceptual system allowing viewers to see meaningful chunks in a graph (as opposed to individual lines or areas) that are linked to conceptual information about graphs that can be retrieved from long-term memory (Richman, Gobet, Staszewski, & Simon, 1996).

In order to understand how viewers read graphs, we draw on the literature on graph design and graph perception. One foundational study by Cleveland (1985) explored visual features, such as color, length, position, and angle that affect graph decoding (i.e., how viewers read a graph). Cleveland (1985) identified a variety of "elementary graphical-perception tasks" (e.g., perceiving color, length, position, and angles) that viewers rely on to read a graph. In addition, prior research on graph perception shows that two types of processes are involved when viewers extract information from graphs: perceptual processing and spatial processing (Trickett & Trafton, 2006). Perceptual processing involves the ability to extract available information from a graph, such as the numbers on the x-axis in a bar chart. In other words, perceptual processing involves "reading the data" (Curcio, 1987). Spatial processing comes into play when the information is not directly in the graph and viewers must make inferences beyond what they see in the bar chart (Trickett & Trafton, 2006), such as comparing trends between bar charts over time. In other words, spatial processing involves "reading beyond the data" (Curcio, 1987).

In sum, these findings imply that graphs can become misleading if they incorrectly display a visual feature that viewers rely on to read a graph, such as position on a scale. In addition, since non-experts tend to rely more heavily on perceptual processing to extract information from graphs (Trickett & Trafton, 2006), they use whatever information is readily available in a graph. This can be problematic particularly when the available visual features are misleading.

Misleading graphs pose severe issues. A large body of research documents that poorly designed graphs can affect viewers' risk perception and consequently their decision making process (Ancker, Senathirajah, Kukafka, & Starren, 2006; Stone et al., 2003). Indeed, perceptual biases affect human decision making. A prominent example of perceptual biases relates to race perception where police officers perceive a black person as more threatening than a white person, resulting in disproportionate decisions to shoot black people [e.g., (James, Klinger, & Vila, 2014)]. Similarly, perceptual biases can affect perceptual experiences other than race, such as when reading graphs. For example, since viewers are worse at judging slopes compared to angles (Cleveland & McGill, 1985), they might experience more perceptual biases when answering questions about slopes in a graph.

Perceptual trainings offer a solution to such biases. For example, police officers undergo perceptual trainings that systematically expose them to racial variation in potential targets while ensuring that race is not indicative of whether the target is threatening or not. Several studies found that such trainings are effective in reducing perceptual biases among police officers [e.g., (James et al., 2014; Plant & Peruche, 2005)]. Perceptual trainings have also been used in the context of STEM learning to help individuals become efficient at extracting scientific information from visual representations (Kellman, Massey, & Son, 2010). The goal of such perceptual trainings is to allow individuals to quickly see meaningful information in visual representations they encounter in educational materials, thereby freeing cognitive resources to invest in further learning of STEM content knowledge.

Perceptual trainings are designed to engage individuals in nonverbal, inductive learning processes. To this end, these trainings expose individuals to numerous visual representations while asking them to perform quick judgment tasks, such as whether or not a given visual shows a certain chemical molecule. Throughout, they are encouraged to process the information visually without overthinking the answer, so as to encourage inductive processing. Further, these trainings provide specific feedback on individuals' performance on the perceptual task (Rau & Patel, 2018).

While perceptual trainings have proven effective in the context of police biases and STEM instruction, to our knowledge, they have never been used to train viewers to extract correct information from misleading graphs. One limitation of prior research is that it has focused more on optimizing graph design instead of training viewers to detect graphical misinformation when present and to adjust their information extraction processes accordingly. Further, there is no prior data on the extent to which specific feedback can help individuals identify misleading graphs. In sum, while perceptual trainings have proven an effective tool for perceptual learning, they have not yet been in the context of visual misinformation. The goal of the present paper is to address these limitations.

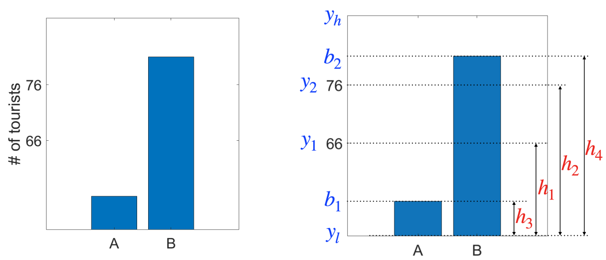
This experiment investigates the following research questions: (1) Are viewers susceptible to misleading bar charts? (2) Which type of feedback is most effective for a perceptual training? (3) At what point during the training are viewers able to identify misleading graphs?

Materials

Misleading Bar Charts

Our task asked participants to determine the ratio of two values in a misleading graph. We used bar charts as we expected all our participants to be familiar with such graphs in their daily lives.

One example of such a graph is given in Figure 2a: there are two bars showing the number of tourists on Island A ($b_1 = 56$) and Island B ($b_2 = 81$). We asked the participants to determine the ratio between the two values i.e., $y = b_2/b_1 \approx 1.4$. Note that the vertical axis intentionally does not show the entire range of values. This is misleading, as the ratio will be overestimated if a participant only pays attention to how the bars "look like" (i.e., bar pixel heights). In this example the pixel height ratio is $y^{\text{pixel}} = 5.2$, which is far from the true value ratio of $y = 1.4$.



(a) (b)

Figure 2: (a) Sample bar chart shown to human participants, with the cover question: "The graph shows the number of tourists who visit two islands on a Saturday. If Island A needs one ferry to transport all the tourists, how many ferries does Island B need? Quickly estimate the answer; you may enter decimal values." The number of tourists on Island A and Island B in this example are 56 and 81 respectively. (b) Annotated variables (not shown to participants).

Mathematically, each bar chart contains six values [1] of interest **(see Figure 2b):** y_l: the lower limit of the vertical axis; y_h: the upper limit of the vertical axis; y_1: the lower vertical tick value; y_2: the upper vertical tick value; b_1: value depicted by the first bar; b_2: value depicted by the second bar.

Visually, though, a participant would directly see only the values y_1, y_2, and perceive four pixel heights: h_1 : pixel height of the y_1 tick; h_2 : pixel height of the y_2 tick; h_3 : pixel height of bar 1; h_4 : pixel height of bar 2. There is indeed enough information to infer the correct y from visual perception:

$$y = \frac{(y_h - y_l)h_4 + y_l h_2 - y_h h_1}{(y_h - y_l)h_3 + y_l h_2 - y_h h_1}. \tag{1}$$

However, with poor y-axis labeling, a participant can be misled into estimating pixel height ratio y^{pixel} instead:

$$y^{\text{pixel}} = \frac{h_4}{h_3}. \tag{2}$$

We call a bar chart misleading if its y and y^{pixel} differ greatly.

In order to prepare a variety of bar charts with diverse misleading potentials, we used the following procedure. We only generated bar charts with two bars, both having positive values. Such a graph is designed by six values y_l, y_h, y_1, y_2, b_1 and b_2. We sampled these six values using the generative model given in Algorithm 1. We started by drawing the values of y_l and y_h uniformly between 10 and 100 (lines 1-3). Next we sampled b_1 and b_2 followed by y_1 and y_2. We ensured that $b_1, b_2 > y_l$ so that the neither bar has 0 pixel height in the generated image. Also we fixed $b_2 > b_1$, meaning the ratio between the two values has a fixed range: $1 < y < 10$. This ensures that the range is never too large or too small for the human participants to estimate. The function pixel$(\cdot)$ calculates the pixel height of the different elements in the graph. The checks in line 6 are performed to ensure that none of the numbers displayed in the plot overlap. We determined the pixel values in line 6 by generating several sample plots. The function generate_bar_graph$(\cdot)$ generates the bar chart using the values sampled. We used MATLAB for this purpose. We also generated multiple cover questions. An example of such a cover question is provided in the caption of Figure 2.

Algorithm 1 Generative Model for Bar Graphs

1: $y_l \leftarrow \mathrm{U}(\{10,\ldots,100\})$; $y_h \leftarrow \mathrm{U}(\{10,\ldots,100\})$
2: Redo step 1 **if** $|y_l - y_h| < 3$
3: $y_l, y_h = \mathrm{sort}(y_l, y_h)$
4: $b_1 \leftarrow \mathrm{U}(\{y_l+1,\ldots,\min(y_l+20, y_h)\})$; $b_2 \leftarrow \mathrm{U}(\{y_l+1,\ldots,y_h\})$
5: $b_1, b_2 = \mathrm{sort}(b_1, b_2)$
6: Redo step 4 if $b_1 = b_2$ **or** $\mathrm{pixel}(b_1, y_l, y_h) < 68$
7: $y_1 \leftarrow \mathrm{U}(\{y_l,\ldots,y_h\})$; $y_2 \leftarrow \mathrm{U}(\{y_l,\ldots,y_h\})$
8: $y_1, y_2 = \mathrm{sort}(y_1, y_2)$
9: Redo step 7 **if** $\mathrm{pixel}(y_2, y_l, y_h) - pixel(y_1, y_l, y_h) < 34$ **or** $\min(|\mathrm{pixel}(b_1,y_l,y_h)$ - $\mathrm{pixel}(y_1,y_l,y_h)|$, $|\mathrm{pixel}(b_2,y_l,y_h)$ - $\mathrm{pixel}(y_1,y_l,y_h)|$, $|\mathrm{pixel}(b_1,y_l,y_h)$ - $\mathrm{pixel}(y_2,y_l,y_h)|$, $|\mathrm{pixel}(b_2,y_l,y_h)$ - $\mathrm{pixel}(y_2,y_l,y_h)|) < 61$
10: **generate_bar_graph**$(y_l, y_h, y_1, y_2, b_1, b_2)$

Perceptual Training

Our perceptual training consisted of a sequence of up to 30 bar charts. For each bar chart, we required the participant to enter their estimate $\hat{y}$ of the true ratio y. Then, the training provided feedback on that chart. We compared three types of feedback:

1. Random-charts verbal-feedback. The 30 charts are randomly generated according to Algorithm 1. We displayed the same feedback for each bar chart by stating "Pay attention to where the vertical axis starts" after the participant entered their estimate $\hat{y}$. We did not show the correct y. An example is given in Figure 3(left).

2. Random-charts y-feedback. The 30 charts are randomly generated according to Algorithm 1. We provided the true ratio y and repeat the participant's estimate $\hat{y}$ as feedback. However, we did not explain which part of the graph the participant should pay attention to. An example is given in Figure 3(right).

3. Machine-teaching y-feedback. We generated a special sequence of 30 charts in an attempt to optimize learning of

[1] We distinguish *value*, which is the number that would have been shown on the y-axis, from *pixel height* which is the apparent height of a landmark.

a simple neural network model. The neural network is a computational approximation of how humans might learn to read the charts. The sequence is optimized using a machine teaching technique (Zhu, Singla, Zilles, & Rafferty, 2018). Details are presented in the Appendix. We provided the true ratio y and repeated the participant's estimate $\hat{y}$ as feedback.

Human Experiments

Participants

The human experiment was conducted using Amazon's Mechanical Turk (MTurk) (Buhrmester, Kwang, & Gosling, 2016). For this experiment, we recruited 79 master workers. Master is a certificate provided by MTurk to workers who have consistently performed well on previous tasks. Each participant was paid 5$ to go through 80 bar charts.

We randomly assigned each participant to one of four conditions:

- The Random-charts verbal-feedback condition
- The Random-charts y-feedback condition
- The Machine-teaching y-feedback condition
- The control condition. This condition consisted of 30 randomly-generated bar charts, but we did not provide any feedback. Thus the control condition can be thought of as simply consisting of 80 test questions. The goal of this condition was to check if participants can identify the misleading graphs on their own without the perceptual training.

Procedure

A participant experienced a total of 80 bar charts in sequence. The sequence was divided into three phases:

Phase 1: Pretest. In this phase, each participant was shown a series of 20 randomly generated bar charts like the one in Figure 2a. For the i-th bar chart ($i = 1 \ldots 20$) the participants had to enter their estimated bar ratio $\hat{y}_i$. There was no feedback and the task immediately moved on to the next bar chart.

Phase 2: Training. In this phase, each participant received 30 bar charts specific to the condition.

Phase 3: Posttest. In order to investigate whether and how soon they identify misleading bar charts, we interleaved Phase 2 and Phase 3. That is, each misleading bar chart was followed by a randomly-generated test bar chart. Bar charts $i = 21, 23, \ldots, 79$ were the charts where the participant entered estimate $\hat{y}_i$ then received feedback depending on the condition; bar charts $i = 22, 24, \ldots, 80$ were posttests where the participant entered $\hat{y}_i$ but received no feedback.

Results

Data cleaning Despite our MTurk task instructions, 7 participants misunderstood the task: instead of responding with an estimate of the b_2/b_1 ratio, they responded with the value of b_2. Such participants were easy to identify due to their large responses in the range of 40 to 100. We manually removed them leaving a total of 72 participants, which coincidentally left 18 participants per condition. For additional outlier removal, we pooled all 5760 remaining responses $\hat{y}$ together and calculated their mean and standard deviation. We then removed responses that were outside $2\times$ standard deviation around the mean. This removed 30 outlier responses out of 5760 responses.

People do misread bar charts We investigated the pretest phase to answer our first question: Are participants susceptible to misleading bar charts? In the pretest phase, the participants had not received different conditions yet. If the participants tend to estimate the bar ratio by the perceived bar heights instead of reading the y-axis labels carefully, then their response $\hat{y}$ should be closer to y^{pixel} than the true ratio y. We merged the pretest data from 72 participants across all 4 conditions. With outlier removal, this led to a total of $n = 1392$ pretest bar charts. The average absolute error between response and true ratio is $\frac{1}{n}\sum_{i=1}^{n} |\hat{y}_i - y_i| = 0.89$ (stderr=0.03), while that between response and bar pixel ratio is $\frac{1}{n}\sum_{i=1}^{n} |\hat{y}_i - y_i^{\text{pixel}}| = 0.43$ (stderr=0.02). A t-test revealed that the difference is significant at level $\alpha = 0.01$. Therefore, we conclude that, during the pretest phase, participants tend to be misled by the pixel height ratio in bar charts.

Table 1: Average absolute error $\frac{1}{n}\sum_i |\hat{y}_i - y_i|$ by condition

Condition	Pretest	1st-half Posttest	2nd-half Posttest
Random verbal feedback	0.89 $n = 348$	0.73 $n = 269$	0.71 $n = 270$
Random y feedback	0.99 $n = 359$	0.51 $n = 270$	0.42 $n = 269$
Machine teaching y feedback	0.91 $n = 356$	0.48 $n = 270$	0.39 $n = 269$
Placebo	0.81 $n = 356$	0.80 $n = 269$	0.79 $n = 270$

Perceptual training helps. We then addressed our second question regarding the effectiveness of perceptual training. First, we checked whether the training interventions led to learning gains. To this end, we conducted a repeated-measures ANOVA with a Greenhouse-Geisser correction. Results showed large significant gains from pretest to posttest (i.e., lower absolute errors), $F(1,68) = 41.824$, $p = .000$, $\eta_p^2 = .381$. Next, we tested the effects of each of the four conditions on learning gains. To this end, we conducted a one-way ANCOVA post-hoc Bonferroni. Results indicated that "Machine-teaching y-feedback" ($p = .005$) and "Random-charts y-feedback" and ($p = .004$) significantly improved participants ability to identify misleading charts compared to control condition. Their absolute error reduced by more than half. There were no significant results with the "Random-charts verbal-feedback" condition. This may be due to participants not receiving the true ratio y as feedback. Finally,

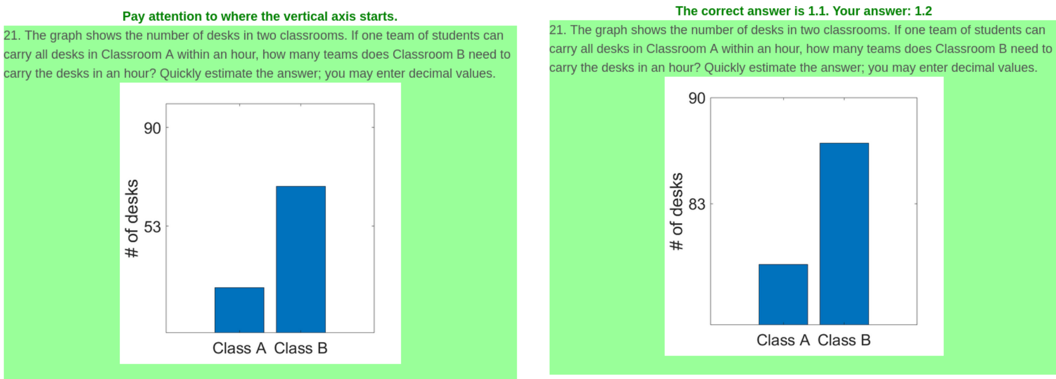


Figure 3: Example of (left) verbal feedback and (right) true ratio y feedback.

participants in the control condition did not improve on their own by being frequently exposed to the task. Recall that these participants did not receive any feedback. Therefore, we conclude that the perceptual training was effective for these types of misleading bar charts.

When are participants able to identify the misleading graphs. To investigate our third question regarding the speed in which this identification takes place, we split the posttest phase into two sections: a 1st-half of 15 charts and a 2nd-half of 15 charts. Recall that after the pretest phase, our 30 misleading charts interleaved with 30 posttest charts. We report the average absolute error $\frac{1}{n}\sum_i |\hat{y}_i - y_i|$ in Table 1. We conducted a repeated measures ANOVA to investigate for differences in absolute errors within the pretest, 1st-half of the posttest and the 2nd-half of the posttest. Results indicated a significant improvement in absolute errors between the pretest and the 1st-half of the posttest (p=.000) but no significant improvement between the 1st-half and the 2nd-half of the posttest (p=.548). We also saw evidence that the participants moved away from the wrong concept of pixel height ratio y^{pixel}: in the "Machine-teaching y-feedback" condition $\frac{1}{n}\sum_i |\hat{y}_i - y_i^{\text{pixel}}|$ *increased* over the pretest (0.36), 1st-half posttest (0.60), and 2nd-half posttest (1.00). Therefore, we conclude that participants were able to identify the misleading bar charts earlier (within the first 15 bar charts).

Discussion and Conclusion

The goal of this experiment was to investigate whether a perceptual training intervention could help viewers extract correct information from misleading graphs. The learning gain results between the pretest and the posttest indicated that both the"Machine-teaching y-feedback"and the "Random-charts y-feedback"versions of the training were effective. Viewers who received these types of feedback outperformed viewers who received either no feedback (control) or no specific feedback on the correct answer (verbal feedback). The superior version of the training provided feedback in the form of the true ratio y. Finally, results indicated that the perceptual training was effective earlier, as evident with the significant improvement of absolute errors between the pretest and the posttest items provided within the first half of the training.

These findings expand prior research on perceptual trainings in two ways. First, to our knowledge no prior research has investigated whether a perceptual training can help viewers become better at extracting information from misleading graphs. Our findings demonstrate that a perceptual training can improve the extraction of information from a graph even if the graph is misleading. Second, in line with prior research on perceptual trainings in STEM (Rau & Patel, 2018), we found that specific feedback in the form of the true ratio increased the effectiveness of the perceptual training. Thus, the design of perceptual training interventions for misleading graphs can benefit from specific feedback that directs participants towards the correct answer. Third, our results show that participants do not have to be exposed to many graphs to be able to extract correct information from misleading graphs.

Our results should be interpreted in light of the following limitations. First, we are using a relatively small sample (18 participants per condition). Second, the recruited MTurkers may have different motivations to complete the assigned tasks. Third, we used a simple form of bar graph that stays homogeneous throughout the training. Future research may focus on other commonly used graphs such as pie charts and line graphs. Despite these limitations, these results indicate that viewers are susceptible to being misled and that a perceptual intervention can help them quickly identify such graphs.

In spite of these limitations, our research shows that a perceptual training with specific feedback can be used to help viewers extract correct information from misleading bar graphs. This suggests that perceptual trainings have promise to help circumvent visual misinformation.

References

Ancker, J. S., Senathirajah, Y., Kukafka, R., & Starren, J. B. (2006). Design features of graphs in health risk communication: A systematic review. *Journal of the American Medical Informatics Association*, *13*(6), 608–618.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2016). Amazon's mechanical turk: A new source of inexpensive, yet high-quality data?

Cairo, A. (2019). *How Charts Lie* (1st Editio ed.). New York: W. W. Norton & Company.

Cleveland, W. S., & McGill, R. (1985). Graphical perception and graphical methods for analyzing scientific data. *Science*, *229*(4716), 828–833.

Cooper, R. J., Schriger, D. L., Wallace, R. C., Mikulich, V. J., & Wilkes, M. S. (2003). The quantity and quality of scientific graphs in pharmaceutical advertisements. *Journal of General Internal Medicine*, *18*(4), 294–297.

Curcio, F. R. (1987). Comprehension of Mathematical Relationships Expressed in Graphs. *Journal for Research in Mathematics Education*, *18*(5), 382–393.

Demuth, H. B., Beale, M. H., De Jess, O., & Hagan, M. T. (2014). *Neural network design*. Martin Hagan.

Galesic, M., & Garcia-Retamero, R. (2011). Graph literacy: A cross-cultural comparison. *Medical Decision Making*, *31*(3), 444–457.

Gibson, E. J. (2000). Perceptual learning in development: Some basic concepts. *Ecological Psychology*, *12*(4), 295–302.

Goldstone, R. L., & Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition*, *65*(2-3), 231–262.

Harper, S. (2004). Students' Interpretations of Misleading Graphs. *Mathematics Teaching in the Middle School*, *9*(6), 340–343.

James, L., Klinger, D., & Vila, B. (2014). Racial and ethnic bias in decisions to shoot seen through a stronger lens: Experimental results from high-fidelity laboratory simulations. *Journal of Experimental Criminology*, *10*(3), 323–340.

Kellman, P. J., & Massey, C. M. (2013). Perceptual Learning, Cognition, and Expertise. In *Psychology of learning and motivation - advances in research and theory* (Vol. 58, pp. 117–165).

Kellman, P. J., Massey, C. M., & Son, J. Y. (2010). Perceptual learning modules in mathematics: Enhancing students' pattern recognition, structure extraction, and fluency. *Topics in cognitive science*, *2*(2), 285–305.

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The Knowledge-Learning-Instruction Framework: Bridging the Science-Practice Chasm to Enhance Robust Student Learning.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105).

Plant, E. A., & Peruche, B. M. (2005). The consequences of race for police officers' responses to criminal suspects. *Psychological Science*, *16*(3), 180–183.

Rau, M., & Patel, P. (2018). A collaboration script for nonverbal communication enhances perceptual fluency with visual representations. In *Proceedings of international conference of the learning sciences, icls* (Vol. 1, pp. 272–279). International Society of the Learning Sciences (ISLS).

Richman, H. B., Gobet, F., Staszewski, J. J., & Simon, H. A. (1996). Perceptual and memory processes in the acquisition of expert performance: The epam model. *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games*, 167–187.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Stone, E. R., Sieck, W. R., Bull, B. E., Yates, J. F., Parks, S. C., & Rush, C. J. (2003). Foreground: Background salience: Explaining the effects of graphical displays on risk avoidance. *Organizational behavior and human decision processes*, *90*(1), 19–36.

Tan, J. K. H., & Benbasat, I. (1990). Processing of graphical information : A decomposition taxonomy to match data extraction tasks and graphical representations author. *Information Systems Research*, *1*(4), 416–439.

Trickett, S. B., & Trafton, J. G. (2006). Toward a comprehensive model of graph comprehension: Making the case for spatial cognition. In *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)* (Vol. 4045 LNAI, pp. 286–300).

Tufte, E. R. (2001). *The visual display of quantitative information* (Vol. 2). Graphics press Cheshire, CT.

Zhu, X., Singla, A., Zilles, S., & Rafferty, A. N. (2018, January). An Overview of Machine Teaching. *ArXiv e-prints*.

Appendix: Machine Teaching

We modeled the human learning using a feed-forward artificial neural network (ANN) (Demuth, Beale, De Jess, & Hagan, 2014). The network had two main components: a convolutional neural network (CNN) (Krizhevsky, Sutskever, & Hinton, 2012) to identify different components of the graph and an optical character recognition (OCR) component that read the two y-tick values. We use VGG (Simonyan & Zisserman, 2014) as our CNN. The weights of this CNN is held frozen as is standard practice. The goal of the CNN is to extract useful features from the graph images e.g., pixel height of the bars, pixel height of the vertical axis labels etc. The goal of the OCR is to extract the vertical tick values. Afterwards, we pass these information through further layers to extract the ratio y.

We now discuss how to construct a good training sequence of length 30 (i.e. the machine teaching doses) using the ANN as a surrogate participant. We first randomly generate a large pool of candidate training bar charts P, and another pool of test bar charts T. We train the ANN separately using each bar chart in P, then determine the ANN's performance on the test set T. This allows us to rank the bar charts in P based on their individual performance. Then we form the training sequence by taking the top 30 bar charts in that order. This is a greedy method; the training sequence so produced is likely suboptimal. Nonetheless, this method enjoys computational efficiency. Other, more computationally demanding methods can potentially produce better machine teaching doses, and are left for future research.