A Markov Model of Users' Interactive Behavior in Scatterplots

Emily Wall* Georgia Tech
Arup Arcalgud† Georgia Tech
Kuhu Gupta‡ Georgia Tech
Andrew Jo§ Georgia Tech

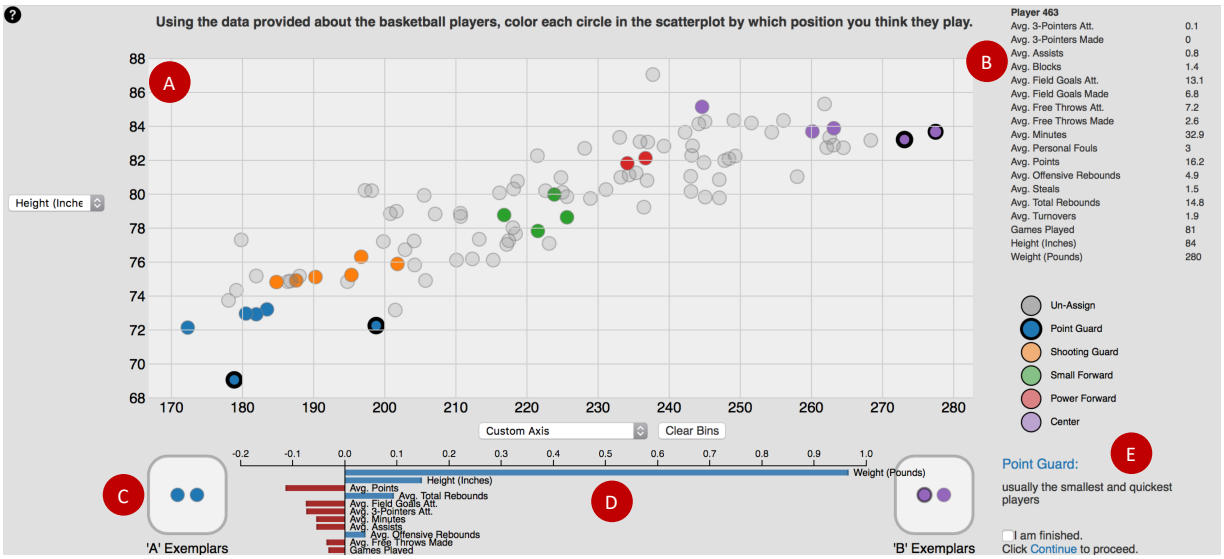


Figure 1: The interface, Interaxis [12], used in the experiment to explore and categorize a dataset of basketball players.

Abstract

Recently, Wall et al. proposed a set of computational metrics for quantifying cognitive bias based on user interaction sequences. The metrics rely on a Markov model to predict a user's *next* interaction based on the *current* interaction. The metrics characterize how a user's actual interactive behavior deviates from a theoretical baseline, where "unbiased behavior" was previously defined to be equal probabilities of all interactions. In this paper, we analyze the assumptions made of these metrics. We conduct a study in which participants, subject to anchoring bias, interact with a scatterplot to complete a categorization task. Our results indicate that, rather than equal probabilities of all interactions, unbiased behavior across both bias conditions can be better approximated by proximity of data points.

Index Terms: Human-centered computing—Human Computer Interaction (HCI); Human-centered computing—Visualization

1 Introduction

As data generally becomes larger and more ubiquitous, visualizations of data are increasingly used for analysis and decision making in domains such as health care, consumer products, government policy, and so on. However, it is unclear what the role of cognitive bias is as humans use these visualizations to analyze data. Recent work has begun to grapple with this question by verifying that users of visualizations are indeed susceptible to cognitive biases (e.g., the attraction effect [5], anchoring bias [2], selection bias [9], etc.), by introducing metrics to quantify those biases [8, 9, 23, 24], and by suggesting techniques to mitigate biased decision making with visualizations [4, 13].

Recently, Wall et al. introduced a set of six computational metrics for characterizing aspects of bias in a user's analytic process [23, 24]. These bias metrics are computed by comparing a user's interaction sequences to a baseline of unbiased behavior. In this paper, we analyze the assumptions made about how to model unbiased behavior in [23]. The baseline of unbiased behavior was theorized as a Markov model, where each combination of {data point, interaction type} constitutes a unique state. However, "unbiased behavior" was initially suggested to be represented as equal probabilities between all states in the Markov model. This assumes randomness in the user's interactive behavior, which we posit is an unreasonable assumption for most tasks and interfaces. Hence, we experimentally challenge the assumption of equal probabilities of interactions by exploring people's actual interaction sequences as they analyze data.

To do so, we conducted an experiment in which we induced anchoring bias (the tendency for people to rely too heavily on initial "anchoring" information) [7] before participants performed a categorization task with an interactive scatterplot [12] (Figure 1). From recorded interactions, we derive a Markov model representing users' observed interactive behavior across two bias conditions. Our analysis indicates that, rather than equal probabilities of all interactions, people's interactions can be better modeled roughly based on the proximity of data points. That is, people are more likely to interact with nearby data points than those that are far away.

2 Related Work

Cognitive bias is a term used to describe errors that result from the use of "rules of thumb" or heuristics in decision making [19]. The cognitive science community has identified hundreds of these heuristics [11]. Of particular relevance to the present work is **anchoring bias**, or the tendency to rely more heavily on information that is first presented [7, 10]. In one experiment to illustrate anchor-

*e-mail: emilywall@gatech.edu
†e-mail: arcalgud@gatech.edu
‡email:kuhu.gupta@gatech.edu
§e-mail: andrew.jo@gatech.edu

ing bias, participants were asked to judge whether the number of African countries in the United Nations was above or below a threshold X, then estimate the actual number [19]. The threshold X was determined randomly and thus should have no impact on people's decisions. However, researchers found that the initial anchoring number strongly impacted the estimates people made.

Bias impacts many aspects of the visual data analysis pipeline [18], including data sampling bias, algorithmic bias, and analysts' cognitive bias in decision making [25]. Herein, we focus specifically on cognitive bias in visual analytics. Recent work has focused on formalizing cognitive bias in the context of visual analytics [6, 21], demonstrating the existence of a bias during interactive data analysis [2, 5, 9, 22], or even proposing ways of mitigating biased decision making [4, 13]. Our work takes a step back from these formalizations to understand what commonalities exist in interactive user behavior under different bias conditions.

3 Wall et al.'s Bias Metrics

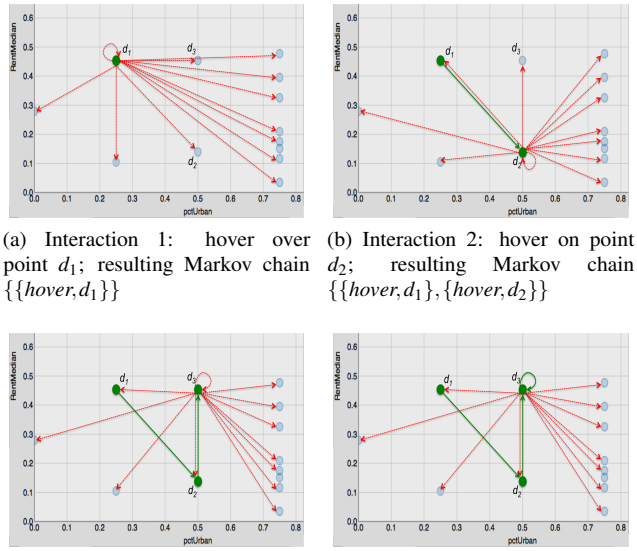
The bias metrics take real-time user interaction logs as input. Interactions, in this case, provide an externalized and approximate representation of a user's cognitive state. The detailed formulation of the metrics is outside the scope of the present work and can be seen in [23]. Instead, in this work we focus on defining a model of "unbiased" interactive behavior by characterizing commonalities of users in different bias conditions. This can ultimately improve the sensitivity of Wall et al.'s bias metrics [23, 24] to more precisely characterize the ways in which users are biased.

Defining a Markov Model. The bias metrics compare a user's *actual* interactive behavior to a *theoretical baseline of unbiased behavior* using a Markov model. A Markov model is comprised of a finite set of states and transitions that occur between the states. It can be conceptualized as a graph, where nodes are states and edges are transitions. A Markov model predicts the next state (the user's *next* interaction) based only on the current state (the user's *current* interaction). For each state (node) in the model, there are possible transitions to other states (outgoing edges), each with an associated probability of occurrence. The sum of probabilities from a given state to all other states must equal 1, representing all possible subsequent states. A Markov model was chosen to model interactive behavior due to its simplicity (predictions based on a single current interaction) and generalizability (a new state space can be defined according to the data and interactions relevant in a given task).

Example. To illustrate how a Markov model applies to interactive behavior in visualizations, consider a user interacting with a scatterplot (Figure 2). Each combination of data point (i.e., point on the scatterplot) and interaction type (e.g., click, hover, drag, etc.) comprises a state in the model. The green dots represent a data point that has been visited, and the red arrows represent all possible transitions from the current state. The green arrows represent the actual sequence of the user's interactions. First, the user hovers on data point d_1 (Figure 2a). Next, she hovers on data point d_2 (Figure 2b), then d_3 (Figure 2c), and lastly clicks on data point d_3 (Figure 2d). This sequence of interactions comprises a Markov chain, whose specific probability can be computed by comparing it to all possible sequences of interactions.

4 Experiment Methodology

We conducted a study to explore the assumptions of "unbiased" behavior in Wall et al.'s bias metrics. In the original formulation of the bias metrics [23] and subsequent experiment [24], users' interaction sequences were compared to unbiased behavior defined by *equal probabilities for all interactions*. However, we believe this assumption is likely ill-fit for most tasks and interfaces. In recent work, researchers constructed a theoretical Markov model based on size of data points (pixel area on the screen) as an approximation



(a) Interaction 1: hover over point d_1; resulting Markov chain $\{\{hover, d_1\}\}$

(b) Interaction 2: hover on point d_2; resulting Markov chain $\{\{hover, d_1\}, \{hover, d_2\}\}$

(c) Interaction 3: hover over point d_3; resulting Markov chain $\{\{hover, d_1\}, \{hover, d_2\}, \{hover, d_3\}\}$

(d) Interaction 4: click on point d_3; resulting Markov chain $\{\{hover, d_1\}, \{hover, d_2\}, \{hover, d_3\}, \{click, d_3\}\}$

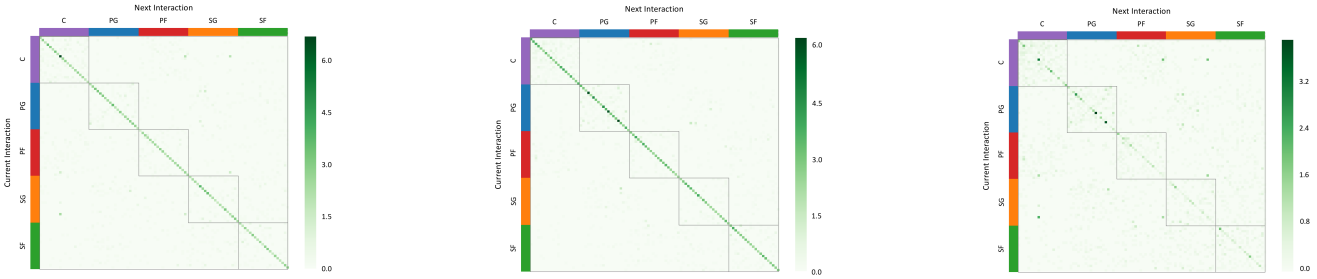
Figure 2: An illustration of a Markov chain produced by four interactions with a scatterplot. Figure reproduced from [23] with permission.

for probability of interaction [3]. We are motivated by such work to create a more precise model of unbiased user behavior based on experimental observations.

We hypothesize that *proximity* can be used to better model user behavior. That is, people will be more likely to interact with nearby data points than far away data points, by starting with what they know (the initial anchoring information) and expanding their analysis, analogous to local exploration of graphs [17]. To test this hypothesis, we replicated the experiment conducted by Wall et al. [24], described below, but refocused data analysis to examine probabilities of interaction sequences. Participants were randomly assigned to one of two task framing conditions, designed to *anchor* them on specific attributes of the dataset. They were tasked to utilize all of the data to categorize 100 anonymized basketball players by position (Center, Power Forward, Small Forward, Shooting Guard, or Point Guard) using InterAxis [12] (Figure 1). To our knowledge, there is no known way to explore truly "unbiased" or perfectly neutral user behavior. Users will be impacted by the framing of the task, prior biases and experiences, etc. Hence, we approximate unbiased behavior by examining the commonalities between two groups of participants who are biased in a controlled way.

4.1 InterAxis

Participants utilized a scatterplot-based visualization tool, InterAxis [12], the same version of the tool used in the experiment in [24]. In the dataset of basketball players, each player is represented in the scatterplot by a circle (Figure 1A), where details (statistics including Height, Weight, Rebounds, Free Throws, etc.) about a player can be seen on the right (Figure 1B) by hovering over a circle in the scatterplot. The axes of the scatterplot can be manipulated by selecting from a drop-down, or by dragging points into the bins on the left and right sides of the x-axis (Figure 1C). The system then computes a weighted combination of attributes representing the difference between the points in the bins. The weights can be further manipulated by dragging the bars beneath the x-axis (Figure 1D). Users can click one of the colored circles on the right (Figure 1E) to display a



(a) Role Condition (b) Size Condition (c) Size - Role

Figure 3: Aggregate probability transition matrices by condition. Rows (current interaction) and columns (next interaction) represent each of 100 basketball players, grouped by position. The highlighted squares along the diagonals indicate subsequent interactions with the same player position. Darker squares indicate higher probabilities.

description of that position. Subsequently clicking on a point in the scatterplot will color and categorize that player accordingly.

4.2 Analytic Task & Framing Conditions

As in the previous study by Wall et al. [24], we likewise focus on the task of data categorization. Participants were tasked to categorize 100 anonymized NBA basketball players[1], 20 players for each of the five positions: Center (C), Power Forward (PF), Small Forward (SF), Shooting Guard (SG), and Point Guard (PG). Participants were not shown the name or team of the players, but were given the following statistics: 3-Pointers Attempted, 3-Pointers Made, Assists, Blocks, Field Goals Attempted, Field Goals Made, Free Throws Attempted, Free Throws Made, Minutes, Personal Fouls, Points, Offensive Rebounds, Steals, Total Rebounds, Turnovers, Games Played, Height (Inches), and Weight (Pounds).

Participants were randomly assigned to one of two conditions. In each condition, we manipulated task framing [20] to impact users' analysis in a controlled way. The two sets of position descriptions in the task were designed to *anchor* participants on a specific set of attributes or statistics in the data (Figure 1E). Participants in the *Size* condition were shown descriptions of the five positions that used statistics about their physical size (i.e., Height and Weight), while participants in the *Role* condition were shown descriptions that used statistics associated with their typical role on the court. For full experimental details, including the specific language used in each framing condition, please refer to supplemental materials [2].

4.3 Participants

We recruited 13 participants to complete our study (7 in the Size condition). Eligible participants completed a screening questionnaire to demonstrate sufficient background knowledge about the domain (basketball) and visualization literacy (scatterplot interpretation) [1, 14]. There was no compensation to participants in the study.

4.4 Procedure

The procedure for this experiment followed the same as in [24], with differences detailed below. Participants provided informed consent and completed two questionnaires (demographic & interface usability). They were shown videos demonstrating how to use InterAxis. Different from the procedure in [24], participants in this study were given the opportunity to get accustomed to the interface for 5 minutes with a small dataset of 15 cars to be categorized by type (as either sedan, SUV, or sports car); they were also shown a refresher video on basketball positions. The main task took approximately 15-20 minutes, during which interactions in the interface were logged. Different from [24], we collected one additional piece of information in the interaction logs to aid our analysis: the locations of all data points at the time of each interaction. In total, the experiment took about 45 minutes.

[1] http://stats.nba.com/

[2] https://github.com/gtvalab/bias-markov

5 Data Analysis and Results

For simplicity in an initial model, we aggregated all interaction types (click, hover, drag) with a data point into a single Markov state. Next, we first filtered out some interactions. Hovers and drags less than 100ms were likely accidental interactions [16], while the user passed from one intentional point to the next; so we removed those interactions. Participants performed, on average, 1043 interactions ($SD = 390$) which filtered down to an average of 527 interactions ($SD = 148$). Participants had an average categorization accuracy of 54% ($SD = 12\%$). Two participants (P12 and P13) did not label all 100 players in the scatterplot. They categorized $\frac{89}{100}$ and $\frac{97}{100}$, respectively. Next we describe and visualize the probabilities resulting from our analysis. All results can be seen in greater detail in supplemental materials[2].

Comparing Conditions. Figure 3 shows aggregate matrices representing the probability of interacting with subsequent players in the scatterplot. Rows indicate the "current" interaction, and columns represent the "next" interaction. Hence, a cell is colored darker according to the probability of interacting first with the associated "current" player and then with the "next" player, where players in each matrix are ordered by their position. We see similar patterns across both conditions. Namely, there is a strong trend along the diagonal. That is, there is approximately a 50% chance that from a given state (player interaction), users next transition will remain in the same state (interact with the same player again), regardless of the condition (50.04% for role condition, 54.74% for size condition). The difference matrix between the two conditions is shown in Figure 3(C), revealing near-0 differences between most transition probabilities in the two conditions (98.5% of transition probabilities ¡ 0.1). Collectively these results suggest similar transition probabilities between states, regardless of condition.

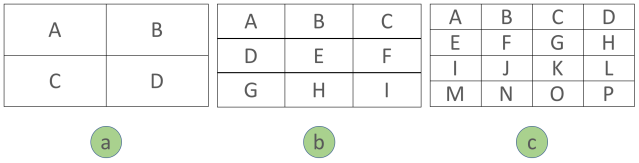
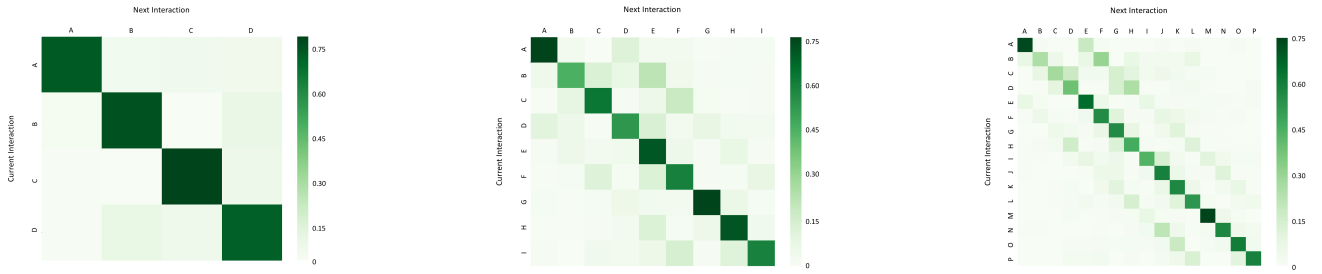


Figure 4: Interactions within the scatterplot were grouped into states in the Markov model by dividing the scatterplot into (A) a 2x2 grid, (B) a 3x3 grid, and (C) a 4x4 grid.



(a) 2x2 grid (4 states) (b) 3x3 grid (9 states) (c) 4x4 grid (16 states)

Figure 5: Aggregate probability transition matrices of all participants when Markov states are defined by grouping points in the scatterplot in a 2x2, 3x3, and 4x4 grid. Darker squares indicate higher probabilities.

Proximity Analysis. In this analysis, we wanted to approximate the probability of interacting with visually nearby data points. To do so, we defined new Markov states by dividing the scatterplot into equal size grids (Figure 4): 2x2 (4 states), 3x3 (9 states), and 4x4 (16 states), and assessed the probability of interacting with points within and between these fixed grid squares. We chose to use a fixed grid overlay for our analysis in order to examine proximity even when the position of individual points on the dynamic scatterplot may be changing. From the previous analysis, we know that multiple interactions with the same player are significantly more likely (e.g., hover on a player then click to label). Hence, in this analysis, we remove subsequent interactions with the same player to see if interactions with different basketball players tend to still follow trends of proximity. Furthermore, we observe no significant difference between conditions, so here we present results aggregated for all 13 participants. Figure 5 shows the results of this analysis. We observe the hypothesized pattern of proximity: users are more likely to interact with other data points within the same grid square (i.e., nearby data points) than data points in different grid squares (i.e., far away data points). This is evident by the stronger colors and hence higher probabilities along the diagonal. In $Markov_{2x2}$, we find that nearby interactions (diagonal probabilities in Figure 5) comprise, on average, 75.3% of subsequent interactions. Similarly, in $Markov_{3x3}$ and $Markov_{4x4}$, we find nearby interactions to comprise 64.36% and 54.29% of subsequent interactions, respectively. Apart from subsequent interactions within the same grid square (higher diagonal probabilities), we also observe a trend in $Markov_{3x3}$ and $Markov_{4x4}$ parallel to the diagonal, indicating that people often perform subsequent interactions with *adjacent* grid squares.

A New Baseline. Results of our experiment suggest that users are more likely to interact with nearby data points than far away data points when performing a categorization task with an interactive scatterplot. How do we now incorporate this information into a new probability matrix that represents a baseline of unbiased behavior?

We tend to favor simple models or modifications over more complex ones, with modest changes to the equal-probability baseline. Hence, we propose that in the context of our experimental task, a more accurate baseline of unbiased behavior could adjust from the equal-probability baseline by distributing interaction probabilities such that subsequent interactions with the *same* data point comprise roughly 50% of interactions from any given state. We could likewise account for proximity by grouping points in grid squares (as in Figures 4-5) and defining probabilities of subsequent interactions within each grid square (nearby interactions) as at least 50% of interactions from any given state, according to the grid size chosen.

6 Discussion and Future Work

Explaining Unbiased Interaction Sequences. This experiment provides a more accurate baseline of unbiased behavior in the context of our tool, dataset, and analytic task. However, we posit that these results may not be especially generalizable. Higher probability of interactions with a specific quadrant of the dataset could be explained by the structure of the task. For instance, because the player descriptions tended to point users to a specific part of the distribution (i.e., the tallest players, the players with the highest number of Assists, etc.), interactions with the high end of the axis likely all occurred within a given quadrant. With all else equal, a slightly different problem framing may likely have yielded a vastly different baseline model. Hence, it is important to account for the specific context of a problem when defining a baseline, including the tool, task framing, and so on. Our experiment provides a model by which more accurate baseline models can be derived through pilot studies for interfaces that may utilize the bias metrics [23, 24].

Other Notions of Proximity. In this work, we focused on understanding how proximity can be used to model users' interactive behavior. However, we only roughly estimated proximity by grouping interactions into Markov states based on a grid pattern. The purpose of this choice was the ease with which it could be computed using a Markov model. Future work could consider other notions of proximity (e.g., measure the precise pixel distance between points).

Future Models. While the current study focused on analyzing data from the perspective of proximity, there are many other variables that could impact user behavior. Future work could include an examination of how aspects of visual salience [15] impact interactive behavior (e.g., default size of data points in the scatterplot, variable encodings using hue or opacity, etc).

Overfitting. There are numerous ways to model unbiased behavior, as mentioned above. However, a common danger among them is to create models that are overfit to user data from a single experiment. Hence, we must exercise caution in how we define or alter models of unbiased behavior, keeping in mind that often the simplest approaches work best. The next step given the current work to improve the baseline is to implement and compare it against other potential baselines of unbiased behavior to see how well the resulting metrics are able to detect deviations in user behavior in real-time.

7 Conclusion

In this paper, we have described an experiment in which we assess the probability of users interacting with different sequences of data points in a categorization task with an interactive scatterplot. Our results indicate that, regardless of bias, users' interaction sequences tend to follow trends of proximity; that is, they are more likely to interact with nearby data points than far away data points. These results enable us to refine what unbiased interactive behavior looks like. This can ultimately be leveraged by bias metrics [23, 24] to more accurately detect when user behavior deviates from acceptable, unbiased behavior.

REFERENCES

[1] J. Boy, R. A. Rensink, E. Bertini, and J.-D. Fekete. A principled way of assessing visualization literacy. *IEEE transactions on visualization and computer graphics*, 20(12):1963–1972, 2014.

[2] I. Cho, R. Wesslen, A. Karduni, S. Santhanam, S. Shaikh, and W. Dou. The anchoring effect in decision-making with visual analytics. *IEEE Conference on Visual Analytics Science and Technology (VAST)*, 2017.

[3] J. A. Cottam and L. M. Blaha. Bias by default? a means for a priori interface measurement. *DECISIVe: Workshop on Dealing with Cognitive Biases in Visualizations*, 2017.

[4] E. Dimara, G. Bailly, A. Bezerianos, and S. Franconeri. Mitigating the attraction effect with visualizations. *IEEE transactions on visualization and computer graphics*, 25(1):850–860, 2019.

[5] E. Dimara, A. Bezerianos, and P. Dragicevic. The attraction effect in information visualization. *IEEE transactions on visualization and computer graphics*, 23(1):471–480, 2017.

[6] E. Dimara, S. Franconeri, C. Plaisant, A. Bezerianos, and P. Dragicevic. A task-based taxonomy of cognitive biases for information visualization. *IEEE transactions on visualization and computer graphics*, 2018.

[7] M. Englich and T. Mussweiler. Anchoring effect. *Cognitive Illusions: Intriguing Phenomena in Judgement, Thinking, and Memory*, p. 223, 2016.

[8] M. Feng, E. Peck, and L. Harrison. Patterns and pace: Quantifying diverse exploration behavior with visualizations on the web. *IEEE transactions on visualization and computer graphics*, 25(1):501–511, 2019.

[9] D. Gotz, S. Sun, and N. Cao. Adaptive contextualization: Combating bias during high-dimensional visualization and data selection. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*, pp. 85–95. ACM, 2016.

[10] K. E. Jacowitz and D. Kahneman. Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin*, 21(11):1161–1166, 1995.

[11] D. Kahneman. *Thinking, fast and slow*. Macmillan, 2011.

[12] H. Kim, J. Choo, H. Park, and A. Endert. InterAxis: Steering Scatterplot Axes via Observation-Level Interaction. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):131–140, 2016. doi: 10.1109/TVCG.2015.2467615

[13] P.-M. Law and R. C. Basole. Designing breadth-oriented data exploration for mitigating cognitive biases. In *Cognitive Biases in Visualizations*, pp. 149–159. Springer, 2018.

[14] P. Lee. Learning from Tay's introduction. `https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/`, 2016.

[15] L. E. Matzen, M. J. Haass, K. M. Divis, Z. Wang, and A. T. Wilson. Data visualization saliency model: A tool for evaluating abstract data visualizations. *IEEE transactions on visualization and computer graphics*, 24(1):563–573, 2018.

[16] A. Newell. *Unified theories of cognition*. Harvard University Press, 1994.

[17] R. Pienta, J. Abello, M. Kahng, and D. H. Chau. Scalable graph exploration and visualization: Sensemaking challenges and opportunities. In *2015 International Conference on Big Data and Smart Computing (BIGCOMP)*, pp. 271–278. IEEE, 2015.

[18] D. Sacha, H. Senaratne, B. C. Kwon, G. Ellis, and D. A. Keim. The role of uncertainty, awareness, and trust in visual analytics. *IEEE transactions on visualization and computer graphics*, 22(1):240–249, 2016.

[19] A. Tversky and D. Kahneman. Judgment under uncertainty: Heuristics and biases. *Science*, 185:1124–1131, 1974.

[20] A. Tversky and D. Kahneman. The framing of decisions and the psychology of choice. *Science*, 211:453–458, 1981.

[21] A. C. Valdez, M. Ziefle, and M. Sedlmair. A framework for studying biases in visualization research. In *DECISIVe 2017: Dealing with Cognitive Biases in Visualisations*, 2017.

[22] A. C. Valdez, M. Ziefle, and M. Sedlmair. Priming and anchoring effects in visualization. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):584–594, 2018.

[23] E. Wall, L. M. Blaha, L. Franklin, and A. Endert. Warning, bias may occur: A proposed approach to detecting cognitive bias in interactive visual analytics. *IEEE Conference on Visual Analytics Science and Technology (VAST)*, 2017.

[24] E. Wall, L. M. Blaha, C. Paul, and A. Endert. A formative study of interactive bias metrics in visual analytics using anchoring bias. *A Formative Study on Anchoring Bias in Visual Analytics Using Interactive Bias Metrics*, 2019.

[25] E. Wall, L. M. Blaha, C. L. Paul, K. Cook, and A. Endert. Four perspectives on human bias in visual analytics. In *Cognitive biases in visualizations*, pp. 29–42. Springer, 2018.