Eurographics Conference on Visualization (EuroVis) 2019
M. Gleicher, H. Leitte, and I. Viola
(Guest Editors)

Volume 38 (2019), Number 3

Investigating Effects of Visual Anchors on Decision-Making about Misinformation

R. Wesslen[1], S. Santhanam[1], A. Karduni[1], I. Cho[1], S. Shaikh[1], & W. Dou[1]

[1]University of North Carolina at Charlotte

Abstract

Cognitive biases are systematic errors in judgment due to an over-reliance on rule-of-thumb heuristics. Recent research suggests that cognitive biases, like numerical anchoring, transfers to visual analytics in the form of visual anchoring. However, it is unclear how visualization users can be visually anchored and how the anchors affect decision-making. To investigate, we performed a between-subjects laboratory experiment with 94 participants to analyze the effects of visual anchors and strategy cues using a visual analytics system. The decision-making task was to identify misinformation from Twitter news accounts. Participants were randomly assigned to conditions that modified the scenario video (visual anchor) and/or strategy cues provided. Our findings suggest that such interventions affect user activity, speed, confidence, and, under certain circumstances, accuracy. We discuss implications of our results on the forking paths problem and raise concerns on how visualization researchers train users to avoid unintentionally anchoring users and affecting the end result.

CCS Concepts

• ***Human-centered computing*** → *Empirical studies in visualization;*

1. Introduction

Visual Analytics (VA) combines statistical and machine learning techniques with interactive visualizations to facilitate high-level decision-making on large and complex data. An important attribute of an effective VA system is the support of *exploratory visual analysis* [Tuk77, Kei02]. Many VA systems designed for exploratory visual analysis often employ coordinated multiple views (CMV) to provide functionality including details-on-demand, linked navigation, and small multiples [Mun14]. These VA systems offer the user flexibility to use the VA system to solve problems through many possible strategy paths and "have a dialogue with the data" [Rob07]. However, user flexibility–like in CMV systems–can introduce trade-offs as well [LM10]. Zgraggen *et al.* [ZZZK18] find too much freedom in visualization systems can lead to spurious insights and high rates of false discoveries, also known as the multiple comparisons problem or the **forking paths problem** [PK18]. Pu and Kay [PK18] define the forking paths problem in visualizations as "unaddressed flexibility in data analysis that leads to unreliable conclusions." They argue cognitive biases may be one reason for users' susceptibility to the forking paths problem. In this paper, we consider the problem within the scope of one such cognitive bias, anchoring bias, and the possible effect pre-task training can have on the complex decision-making task of social media misinformation identification using a CMV VA system.

Cognitive biases are the result of the over-reliance of heuristics, or rules-of-thumb, for decision-making tasks to make decisions with relative speed [TK74]. An emerging topic within the VA community considers the role of cognitive biases in VA decision-making [WBFE17, VZS17, DFP*19]. Cognitive biases have been shown to affect decision-making processes in predictably faulty ways that can result in sub-optimal solutions when information is discounted, misinterpreted, or ignored [TK74]. One cognitive bias relevant to exploratory visual analysis with VA systems is **anchoring bias**. It refers to the human tendency to rely too heavily on one and most likely the first piece of information offered (the "anchor") when making decisions [Kah16]. Past studies from psychology and cognitive science have focused on numerical anchoring, in which an initial numerical value anchors judgment and the subsequent adjustment with updated information [Kah16, EG06, LGHG17]. Cho *et al.* [CWK*17] provided evidence anchoring transfers to VA; specifically **visual anchoring**, which is the over reliance on a single or subset of views during exploratory visual analysis.

To situate our experiment in real-world decision-making tasks with VA systems, we selected the application of misinformation identification. Recently, the topic of combating misinformation has received much attention in many fields including machine learning, psychology, journalism, and computational social science [LBB*18, PCR18, SSW*17]. While a variety of fully automated techniques have been developed, more direct interaction like laboratory experiments with users on misinformation decision-making is needed [VRA18].

Our work makes the following salient contributions:

1. We conducted an empirical study on the effects of visual anchoring in decision-making. Specifically, we investigated misinformation in social media in a between-subjects design laboratory experiment with 94 participants.
2. Introduction and formalization of strategy cues and visual anchors as treatments to intervene within the visualization training process.
3. Careful integration of strategy cues from psychology literature as hypotheses to test the interaction between visual anchoring and providing hypotheses in visual decision-making tasks.
4. Quantitative analysis on factors that affect anchoring bias in VA to measure visual anchors' impact on user decisions, confidence, time spent, and interactions.

Understanding the effect of cognitive biases like anchoring in visual analysis serve as an important first step to raising awareness and possibly mitigating cognitive biases with visual analysis. At the end of the paper, we connect findings from our experiment to practices of interacting with participants on a newly designed visual analytic systems. The findings of our experiments shed more light on how and when anchoring effects can occur in visual analytic systems and call for more careful consideration of training users or designing tutorials for a visual analytic system.

2. Background

In this section, we review past research on cognitive biases in visualizations. We also review literature that motivated our experiment design and research questions.

2.1. Anchoring & Cognitive Biases in VA

A cognitive bias is a systematic and involuntary cognitive deviation from reality [Poh04, DFP*19].[†] Introduced by Tversky and Kahneman [TK74], cognitive biases have since had a long history of investigation by various social scientists [BBM*15]. Ellis and Dix [ED15] provide an early case on exploring the role of cognitive biases in VA and, more recently, several papers have provided theoretical frameworks or taxonomies of cognitive biases in VA [WBP*17, WBFE17, VZS17, DFP*19]. Empirically, VA research on cognitive biases have developed studies on a variety of biases including attraction [DBD17], selection [GSC16], availability [DDB16], and confirmation bias [KWS*18, DJDO18]. In our study, we explore the phenomenon of anchoring [TK74], which is the tendency to focus too heavily on one piece of information when making decisions. Originally considered in the task of open-ended numerical decisions [TK74, EG06], anchoring has been studied in VA both in MTurk studies using scatterplots [VZS18] and more complex, lab-based experiments using a CMV system [CWK*17].

As one of the first studies on the effect of anchoring in VA, Cho *et al.* [CWK*17] employs an open-ended task of identifying protest-related events from social media data. They analyzed the impact of visual anchors on the reliance of views and analysis paths; however, users decisions were more affected by classical numerical anchors within the task. The impact of visual anchoring on user performance was not measured and limited information was captured at the time of each decision.

[†] Dimara *et al.* [DFP*19] provide a detailed discussion that "reality" stems from normative models of decision-making, which in itself leads possible controversies on assessing cognitive biases.

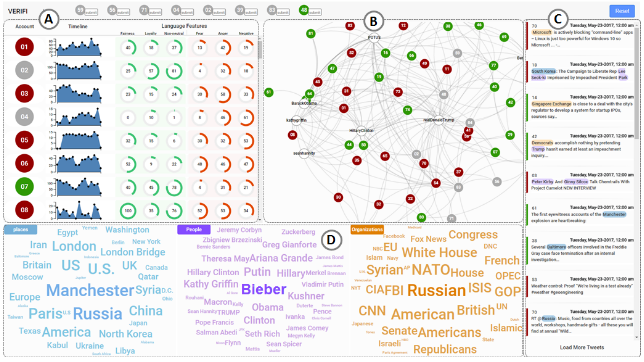

Figure 1: *Screenshot of Verifi. Verifi is comprised of four views: (A) Language Features View, (B) Social Network View, (C) Tweets Panel View, and (D) Entities View. Progress Bar and Form Submit buttons are at the top.*

2.2. Strategy Cues in Psychology Experiments

Our motivation for providing strategy cues in the experiment design is rooted in prior research from psychology [RWP15, AGP17, WSS17, SMD15, BBP17]. In their influential conceptual model for exploratory analysis, Pirolli and Card argued that identifying an exploration strategy or hypothesis is one technique that users of visual analytic systems can benefit from [PC05]. To illustrate, research has demonstrated that users preferred to devote attention to stimuli that matched a given hypothesis or template, even in the presence of alternate, more optimal strategies [RWP15]. Amer *et al.* [AGP17] designed experiments in which participants were given explicit and implicit spatio-temporal cues in a visual event coding task and found systematic effects of the explicit and implicit cues on users' attention within the visual analytic system and how these cues affected processing of information.

2.3. Possible Training Induced Biases

A recent survey of visualization evaluation practices from the Vis Community highlighted that many publications need to observe more evaluation reporting rigor by providing important methodological details [IIC*13]. In particular, there is a lack of consistent reporting on how the participants were trained (by experimenters, with or without a script, training videos, example strategies to complete the task, etc.). In our experiment, visual anchors are introduced during training the participants on how to use a visual analytic system to investigate misinformation. We will investigate the impact of the visual anchors on users' performance and behavior during analysis. In summary, how participants were trained may significantly impact their task completion, thus we argue for more consistent reporting of these details.

3. Experiment

In this section, we outline our experiment design including reviewing Verifi, the VA system used in the experiment, our research questions, variables, and hypotheses.

3.1. The Verifi System

For our study we use Verifi [KWS*18] (Figure 1), an interactive CMV system for identifying Twitter news accounts suspected of spreading misinformation. Verifi includes four views: Social Network, Language Features, Tweet Panel, and Entities. Each view provides users with different features in detecting misinformation [VSJH17]. The Social Network and the Language Features views are the two primary views; the Entity View and Tweet Panel are secondary views. Following Cho *et al.* [CWK*17], we selected Verifi to test visual anchoring as its an example of a complex CMV system. CMV systems inherently require users to make choices on which views to use and strategies to switch between views. Accordingly, visual anchoring may occur in such systems when a user is biased into over relying on one view and possibly leading to a sub-optimal decision.

The system includes two weeks of tweets from 82 Twitter news accounts. Each account name was converted to an integer code (1 to 82) and annotated as a misinformation account (red), real news outlet (green), or requiring user decision (grey). The annotations are based on independent, third-party sources.[‡] Each user's task is to make a decision on the veracity (real or suspected of spreading misinformation) for eight grey accounts within a one-hour session. Following [KWS*18], the eight accounts were qualitatively selected to provide a range of difficulty level as well as consistent and inconsistent information to challenge users in their decision-making processes. Table 1 provides the actual Twitter handles of the eight gray accounts along with a brief description.

3.2. Experiment Design

We highlight **two critical design decisions** in our experiment compared to Cho *et al.* [CWK*17]:

1. We collect direct feedback from users in a submission form (Figure 2) to capture input at the time of each decision (e.g., view importance, strategies, and open-ended comments). Unlike Cho *et al.* [CWK*17] who captured users' decision on paper and after the task, we designed the Form Submit view (Figure 2) to collect information regarding the factors that influenced each decision.
2. We provide **strategy cues**, or hypotheses, as a secondary condition in the form of written statements that reinforce functionality for each primary view in Verifi [KWS*18]. Strategy cues are initial hypotheses, provided on paper to users, of possible relationships between the data elements in a specific view. Strategy cues align to confirmatory data analysis as they provide a mechanism to control for possible hypotheses of the task and its functionality with an anchored view (e.g., real news accounts have lower anger, fear, or negativity). Strategy cues serve as interaction variables to the visual anchor as they may enhance the anchor's effect on a view if the user follows the view's strategy cues.

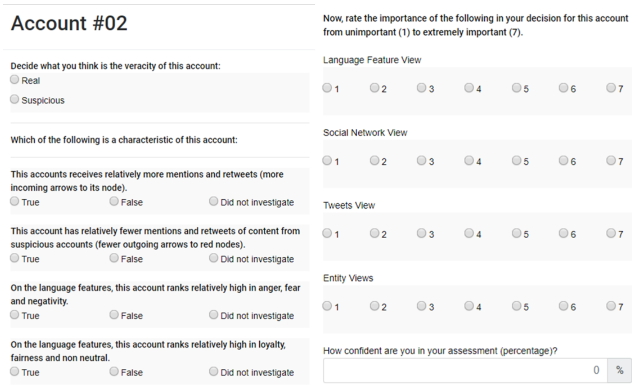

Figure 2: *Form Submit view of Verifi for Account #02 (@ABC). This pop-up provides an interface for the user decisions and feedback per account (e.g., strategy cues use, view importance, and open-ended comments (not shown).)*

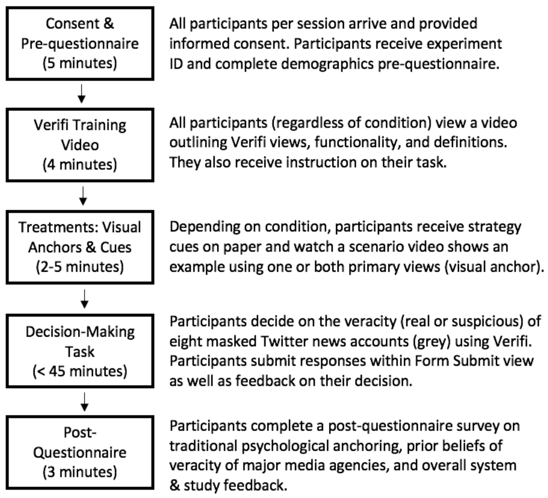

Figure 3: *The experiment flow for each participant session.*

To analyze the effects of visual anchors and strategy cues in decision-making, we conducted a between-subjects, repeated measures laboratory experiment. Figure 3 provides the experiment flow. Each user's task is to make a decision on the veracity (real or suspicious) of eight grey Twitter accounts (see Table 1). Users submit their decisions in the Form Submit view (Figure 2) along with their ratings of the visualization views and strategy cues. To control for learning effects, we randomized the order of the account icons in the Progress Bar per participant.

94 users participated in our study. The gender distribution was 68% male and 32% female. Users' ages were between 21 and 56

‡ Suspicious accounts are based on four websites as provided in [VSJH17]. 31 real news accounts are provided through the following links: https://tinyurl.com/yctvve9h and https://tinyurl.com/k3z9w2b

News Outlet	Description
@zerohedge	A financial blog with aggregated news and editorial opinions
@AddInfoOrg	An anti right-wing news blog and aggregator
@NatCounterPunch	An alternative media magazine and online news aggregator
@SGTreport	Anti corporate propaganda outlet with exclusive content and interviews
@ABC	A news division of a major broadcasting company
@nytimes	An American newspaper with worldwide influence and readership
@TIME	An American weekly news magazine
@Reuters	An international news agency

Table 1: *Eight Twitter news accounts for users' decisions (i.e., grey accounts in the interface). Accounts were anonymized in the study.*

(M = 28.7). A majority of users were pursuing a master's degree (n = 83), followed by undergraduate (n = 5), graduate certificate (n = 5), and Ph.D. (n = 1). Students were recruited through extra credit incentives offered in six courses: Visual Analytics (n = 40), Natural Language Processing (n = 25), Advanced Business Analytics (n = 14), Human Behavior Modeling (n = 6), Applied Machine Learning (n = 6), and Social Media Communications (n = 3).

Each participant session was capped at 45 minutes and averaged 27.1 minutes (SD = 7.524). Each session is identified through a participant ID and interactions (e.g., clicks, hovers, and scrolls) were saved to a MongoDB database. Computer specifications (browser, output/zoom) were controlled for to avoid them as confounding factors. Our study was approved under our institution's Institutional Review Board (IRB) policies (IRB #17-0251).

3.3. Research Questions

We investigate how users may be visually anchored on different *views* in a CMV system and how they might be anchored on specific interaction *strategies* based upon the training given to them. How does visual anchoring affect user performance, confidence and data coverage? Accordingly, our main research questions (RQs) are:

RQ1: What is the effect of visual anchors and strategy cues on participant performance (i.e., accuracy, speed, and confidence) and ratings (e.g., view importance and strategy usage)?

To analyze RQ1 from a participant-level, we use aggregated[§] non-parametric bootstrapped confidence intervals [WK16]. In our results, we focus on effects sizes rather than p-values [KNH16] and follow conventions provided by Dragicevic [Dra16]. Then we employ a hierarchical model to consider both participant and task-level effects on user accuracy and confidence. Following Kay *et al.*'s [KNH16] recommendation for Bayesian methods in HCI, we use Bayesian mixed-effects regressions with weakly-informed priors [FWM*18].

§ Given HCI's focus on people not tasks, Dragicevic [Dra16] advocates for calcuating confidence levels on a participant-level, not a task-level.

	Visual Anchor	Strategy Cues	Users	Decisions
Control	None	None	14	112
	None	1S, 2S, 1L, 2L	15	120
Balanced	SN -> LF	1S, 2S, 1L, 2L	17	134
	LF -> SN	1S, 2S, 1L, 2L	16	128
Partial	SN Only	1S, 2S	15	119
	LF Only	1L, 2L	17	135

Table 2: *Experiment treatments by condition groups.*

RQ2: Can users' analysis process (e.g., interaction logs) be linked to participant performance outcomes to infer user strategies?

For RQ2, we estimate condition effect sizes of user time spent per view and coverage metrics [WBFE17] using mean bootstrapped confidence intervals. To identify user behaviors with the coverage and time spent metrics, we used Ward's D2 Agglomerative Hierarchical Clustering [ML14] to cluster users and features using the R package `heatmaply` [GOSS17]. To determine the optimal number of clusters for the rows (features) and columns (users), we used the maximal average silhouette width method on the cophenetic distance of the dendrogram [Gal15]. The algorithm detected five clusters on the user-level, as identified by the five colors in the horizontal dendrogram. We then annotated the five clusters based on common attributes shared by users within a cluster.

3.4. Independent Variables (experimental conditions)

For our experiment design, we developed six treatments in three condition groups: Control, Balanced, and Partial (Table 2). The **Control** group did not receive any visual anchor (i.e., scenario video). The **Balanced** group received a visual anchor that reviewed a strategy using both primary views. The **Partial** group received a visual anchor that covered only one primary view but not the other.

In addition, the difference between each group condition was the *strategy cues* (or hypotheses) given to participants that reinforce each primary view. Each *strategy cue* is a hypothesis on how to identify real news accounts that aligns to one of the two primary views in the Verifi: Language Features view **(L)** and Social Network view **(S)**. The Language Features view presents predictive linguistic features for each account, such as fairness, loyalty, anger, and fear. The Social Network view provides retweet and mention relationships [KWS*18]. The cues are:

Cue 1L: "On the *language measures*, real news accounts tend to show a higher ranking in loyalty, fairness, and non-neutral."

Cue 2L: "On the *language measures*, real news accounts tend to show a lower ranking in anger, fear and negativity."

Cue 1S: "In the *social network graph*, real news accounts are less likely to mention and retweet content from suspicious accounts (fewer outgoing arrows to red nodes)."

Cue 2S: "In the *social network graph*, real news accounts tend to receive more mentions and retweets (more incoming arrows to their nodes)."

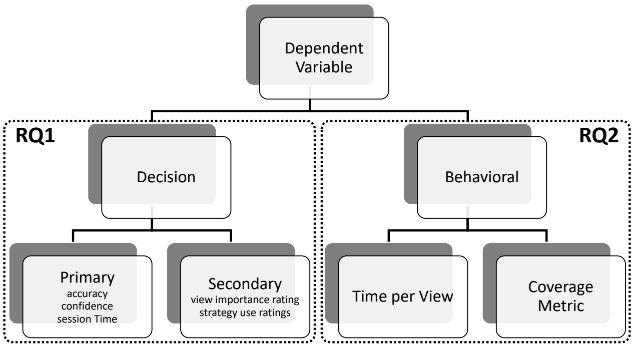

Figure 4: *Dependent variable groups in our experiment.*

3.5. Dependent Variables

We have two types of dependent variables: decision and behavioral metrics (see Figure 4). Decision metrics are provided by users and can be divided into two groups: primary and secondary outcomes.

Primary outcomes: We evaluated user performance based on three primary outcomes: (1) **accuracy** in correctly identifying misinformation accounts, (2) **confidence** of each misinformation decision as range from 100 (perfectly confident) to 1 (perfectly not confident), and (3) **session time** that is the time of the decision as minutes from the start of the experiment.

Secondary outcomes refer to the four **view importance ratings** and four **strategy use ratings** provided directly by each user at the time of each decision. The importance ratings use 1 (unimportant) to 7 (extremely important) Likert scale and the strategy cue ratings use a True, False, or Did Not Investigate value (see Figure 2).[¶]

In addition to decision metrics, we also consider participants' actions as dependent variables in two types of behavioral metrics.

Time spent metrics. We measured users' time spent per view through a mouse enter-exit log tracking. By using the enter-exit periods and allocating that to each view, we were able to measure participants time spent in the five views (two primary, two secondary, and Form Submit view).

Coverage metrics. Following Wall *et al.* [WBFE17], we created coverage metrics to measure participant use of key interface functionality. Specifically, we consider six primary actions: progress bar click, LF sort (combined for red/green features), and SN hovers (for grey, green, and red accounts).[‖]

3.6. Hypotheses

Based on the RQ's, we developed the following hypotheses:

H1: Balanced visual anchor users will have the highest accuracy as users will have more information on how to use both primary views. These users will use the primary views more than the secondary views as compared to the Control groups.

H2: Partially visual anchored users will have the worst accuracy as their anchors are one-sided. These users will disproportionately interact with the view associated with their anchor. By failing to consider the opposite view, their performance will diminish.

H3: When given scenario videos that include both primary views (i.e., Balanced group), order matters. The first view provided will have a larger effect than the second, leading to an increased use (time, coverage) of the first view introduced. To evaluate, we compare performance within the two Balanced conditions.

H4: Strategy cues will improve performance, confidence, and shorten session as more information is helpful. In this hypothesis, we'll compare treatments to the Control treatment with no cues.

4. Results

4.1. RQ1: Effects of Visual Anchoring and Strategy Cues on User Level

Contrary to **H1**, we do not find evidence that the Balanced visual anchored groups (70.2%-71.88%) have the highest accuracy. In fact, we find that the Control groups (i.e., no visual anchor/scenario video) performed just as well in terms of accuracy. Figure 5 provides the means and bootstrapped confidence intervals for the primary outcomes relative to the experiments conditions. Alternatively, we find some evidence in support of **H2** as the Partial-Social Network treatment had a lower accuracy (M = 61.3%) than the other groups, but not outside of 95% bootstrapped CIs. We do find evidence that the visual anchors provide a positive effect on user confidence relative to the Control conditions.

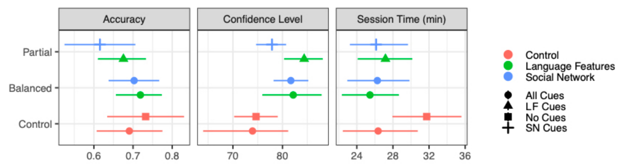

Figure 5: *Primary outcomes means and bootstrapped 95% confidence intervals on a user-level (n = 94).*

For **H3**, we find little difference in primary outcomes between the two Balanced groups, indicating that order doesn't appear to affect final decision outcomes. For **H4**, we find no evidence that the strategy cues provided an advantage in accuracy, in fact the opposite as the Control/no cues condition has the highest average accuracy. We find the cue groups do tend to have higher accuracy, but their effects may interact with the visual anchors as we find little difference between the Control groups. Last, it does seem that cues may shorten the session as the Control/no cues group, the only without any cues, had the highest average session time (M = 31.7), well above all other groups (ranging from 25.4 to 27.2). We'll explore this more in **RQ2** results when we decompose the session time by views.

[¶] Consistent with [KWS*18], we recoded strategy cue ratings to ensure whether the cues were consistent or not, depending on whether the account was Real or Misinformation. In this way, the cues can be interpreted as 1 = cue used consistently, -1 = cue used inconsistently, 0 = cue not investigated.

[‖] We removed hovers less than one second after a previous action to remove unintentional actions.

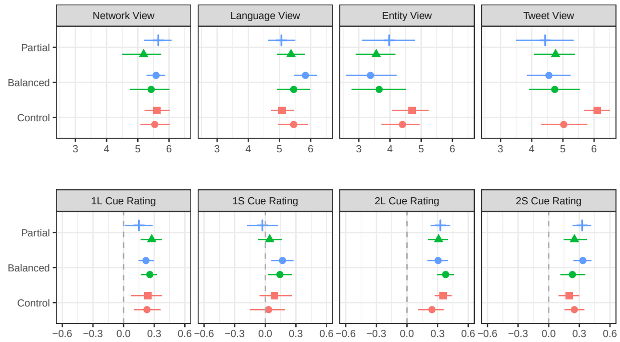

Figure 6: *Secondary outcomes means and bootstrapped 95% confidence intervals on a user-level (n = 94). The figure uses the same color and shape encodings as Figure 5.*

As for the secondary outcomes, we find evidence for **H1** that visual anchors seem to diminish users' value of the secondary views. For example, Balanced (and Partial) anchored users tend to rate both the Entities and Tweet view less than Control groups (Figure 6 top row). In fact, we find the Control/no cues condition valued the Tweet view the highest (M = 6.1), suggesting that without any anchors or cues, users valued the qualitative secondary view the most (i.e. reading individual tweets).

Last, we find little variance across cue ratings by the six conditions (Figure 6 bottom row). Most average ratings range from 0.2 - 0.3, indicating a slightly above average (0) use of the cue in their decision. The one exception is **1S**, in which all but the Balanced groups average rating was within 0 for its confidence interval.

4.2. RQ1: Effects of Visual Anchoring and Strategy Cues on User & Task Level

One weakness of the user-level analysis is that it ignores the task-level. In Figure 7, we find that user accuracy varied drastically by each account (task). For example, nearly all participants correctly predicted @nytimes while most users incorrectly predicted @TIME, especially those receiving visual anchors. To consider both the user- and task-level, we use mixed-effects regressions for both accuracy and confidence.[**]

We use a Bayesian generalized linear mixed-effects regression for each of the two outcome values using the R packages `brms` [B*17] and `tidybayes` [Kay18]. Our fixed effects are each *treatment* (Table 2), *time of decision* (in minutes), and their interactions.[††] For the random effects, we use *account* (Table 1) and *participant*. We use account as a random effect given the variability in difficulty from the qualitative ground truth [KWS*18].[‡‡]

** We did not investigate total session time due to the problem of allocating time to each actions for each decision. Therefore, we only investigate accuracy and confidence as dependent variables in regression.

†† We do not report the fixed effects of time of decision as we did not have a prior hypothesis to evaluate. However, these effects can be observed 04-regressions.Rmd | .html in the supplemental materials.

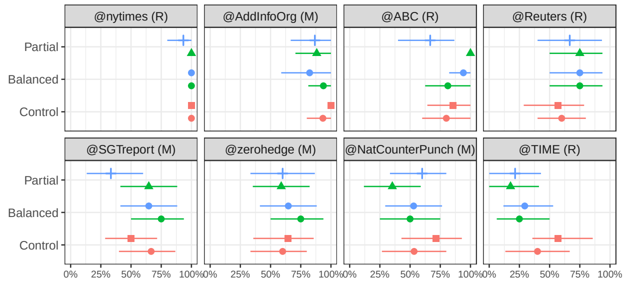

Figure 7: *Accuracy by Twitter account and bootstrapped 95% confidence intervals on decision-level (n = 748). (R) indicates a "real" news account and (M) indicates a "misinformation" account. The figure uses the same color and shape encodings as Figure 5.*

For each regression we use a slight variant depending on the outcome variable format. For accuracy, a binary 1 (correct) or 0 (incorrect) variable, we use a logistic mixed-effects regression. Alternatively, confidence is a continuous variable between 0 (no confidence) to 1 (perfect confidence) and, hence, we use a linear mixed-effects model.

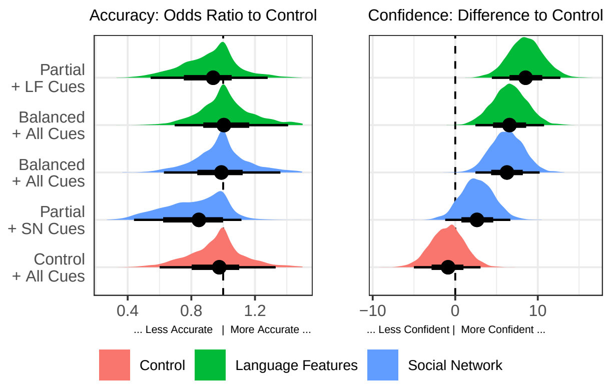

Figure 8: *Posterior distributions of differences in means of user accuracy and confidence level. For both plots, the conditions are relative to the Control (no cues) treatment. CIs of differences are at 95% and 66%.*

From Figure 8, we find that the treatments had a strong effect on user confidence but a smaller effect on accuracy. For instance, the Partial (LF cues) and the Balanced (SN) had effects larger than the 66% CI compared to the reference level, Control (no cues) treatment. This provides evidence that the visual anchors tend to produce higher user confidence levels. However, for accuracy, we find

‡‡ We only included participant as a random effects for confidence, not accuracy, following a significant effect via ANOVA testing with frequentist mixed-effects modeling. See 04-regressions.Rmd | .html in the supplemental materials.

small effects of the treatments as nearly all odds ratio CIs are within 1 (i.e., as likely as the reference level). The one exception is the Partial (SN Cues) treatment in which its 95% CI is nearly out of 1.

4.3. RQ2: Time Spent & Coverage Metrics

To evaluate the behavioral effect of visual anchors, we explore effect sizes using bootstrapped confidence intervals to identify differences in participants' time spent and coverage metrics, Figure 9 and 10). To consider **H1** and **H2**, we compare the visual anchored groups, Balanced and Partial, to the Control groups. First, we find that the visual anchored groups tended to spend more time on the Language View than the Control groups; however, time on the Network View was mixed as the Control Groups spent around 8-9 minutes on average, nearly the same as the anchored groups. This provides some evidence for our hypotheses, but only for Language View anchoring. The one anchored group that spent little time in the Language View was the Partial-Social Network treatment, where users averaged only 4 minutes (M = 4.01 minutes) as compared 6 to 7.5 minutes for the other anchored groups. This makes sense given these users' anchors only included the Social Network cues and videos, not the Language treatments.

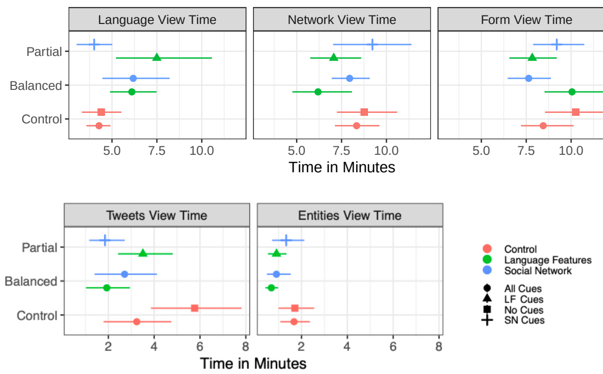

Figure 9: *Time spent per view means and bootstrapped 95% confidence intervals on user-level (n = 94).*

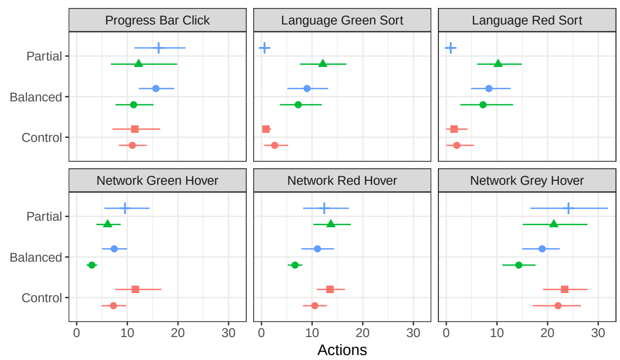

Figure 10: *Coverage metrics means and bootstrapped 95% confidence intervals on user-level (n = 94). The figure uses the same shape and color encodings as Figure 9.*

Considering users' coverage metrics, we consistently find that visual anchored groups (except the Partial-Social Network treatment) had many more Language interactions (Green and Red sort) than the Control groups. However, Social Network interactions (i.e., hovers) are similar between the visual anchored groups and the controls. Both of these points lead to partial evidence for **H1** and **H2**. That is, we find that users can be visually anchored to the Language View, but not the Social Network View.

4.4. RQ2: Clustering Users based on Interactions

We find users' actions can provide indications of different interaction behaviors (Figure 11). For example, consider the 'Slow and Steady' cluster. In Figure 11, these users are mostly yellow, indicating a high rank across all metrics. These users were very active, exploring the entire interface's functionality for an extended period of time. On the other hand, the 'Fast and Quick' group is mostly dark blue as they ranked low in coverage metrics and time spent. The bottom two rows of the dendrogram provide the treatment conditions for each user. Comparing these rows to the clusters, we find some evidence for **H1** and **H2**. Take 'Anchored to Social Network' group as an example. Only one user who was treated with a LF visual anchor (dark blue) is within this cluster. As we would expect, many are SN groups (light red) that received the SN visual anchors. However, what is peculiar is the number of Control users (dark blue), particularly those without any strategy cues. Perhaps one interpretation is that these users are naturally drawn to the social network view more than other views.

Descriptive statistics can also provide more context on each cluster. We find that the 'Slow and Steady' cluster users averaged much longer session times (M = 36.0 minutes). These users tended to have longer initial exploration periods, as they averaged nearly 10 minutes before their first decision submission. As context, other users typically made their first decision between 3 and 7 minutes. We also find that these users actively used the Progress Bar (M = 21.5 times), indicating a more organized strategy and using both primary views frequently. Interestingly, this cluster has, on average, the highest accuracy of 82.8%. Alternatively, we identified two clusters as users who focus more on either the SN (#1) or LF (#2). For example, cluster #1 spent 2.3x more time on the Social Network view than the Language Features view while the opposite holds for cluster #2.

Cluster validation: To validate the clusters, we compared them to post-questionnaire and decision data that was not included in the clustering process. For instance, we find that the clusters provide a range of different ratings for the language features and social network functionality in the post-questionnaire. Users in the 'Anchored to Social Network' (#1), 'Highly Confident' (#3), and 'Fast and Quick' (#4) generally preferred the social network over the language features. However, the 'Anchored to Language Features' cluster (#5) was the only cluster to prefer, on average, the LF over SN. Alternatively, we find distinct differences in user motivation, interest, and challenge between clusters like 'Slow and Steady' (#2) and 'Fast and Quick' (#4). The 'Slow and Steady' cluster tended to be the most motivated, interested, and challenged out of all of the clusters. This makes sense given their longer session times and heavy usage. On the other hand, the 'Fast and Quick' cluster was

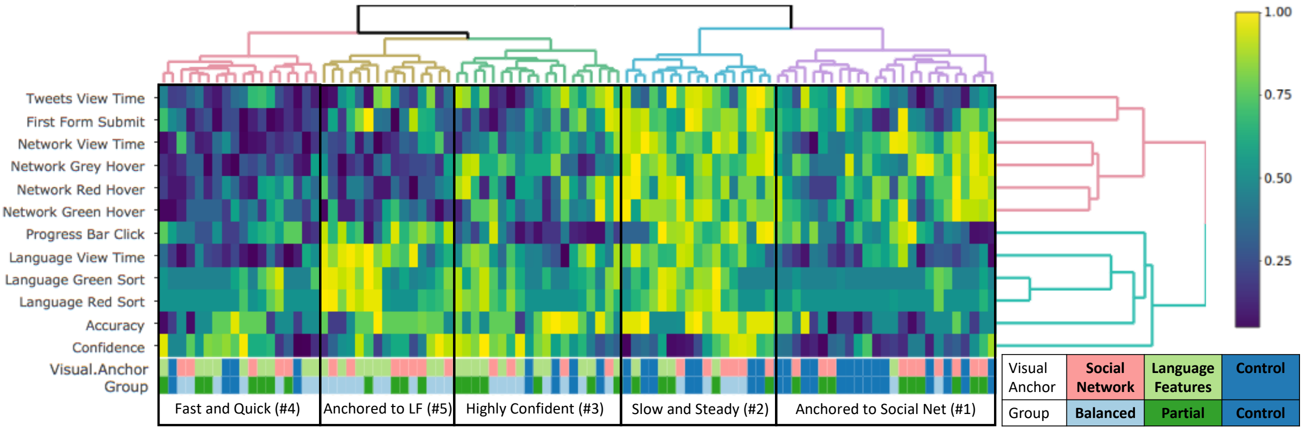

Figure 11: *Heatmap clustering of interaction logs (Ward.D2 [ML14]) by columns (users) and rows (metrics). Each column is normalized for its percentile ranks. Users with a high feature rank are yellow while users with a low rank usage are dark blue. The bottom two rows indicate user's group and anchor condition. Both metrics were not used in clustering and provided for comparison.*

the least motivated and interested. Perhaps lack of interest led to shorter session times and may factor in their lower accuracy.

Last, we explored the user-level interaction logs through scatter plots to validate our clusters. Figure 12 provides a scatter plots of fifteen user sessions. In each plot, a dot represents an action for each of the six views across session time (x-axis) and view (y-axis), with slight y-axis jittering to avoid overlapping actions. Each column includes three user sessions per cluster and chart row order represents, in descending order, highly accurate to inaccurate users.[§§]

We were able to identify general patterns and outliers from these plots. For example, the left-most column provides three users who are clustered to the 'Anchored to Social Network' group. These users tend to have many more actions in the Social Network view as compared to the Language Features, Tweet Panel, or Entities view. They seldomly use the Progress Bar (e.g., S104 and C1 use it somewhat while S108 never used the Progress Bar). Alternatively, we find examples in the 'Slow and Steady' group to have much longer user sessions, lasting well over thirty minutes (some even near forty minutes or more). These users tend to use a combination of all views like the Language Features, Social Network, and even the Tweet Panel views. Alternatively, we were able to identify outlier behaviors, like L103, who almost exclusively used the Language Features view. Even more interesting, the user waited until the end of the session to make all decisions.

Post-Questionnaire Feedback. We also evaluated open-ended feedback from users to assess user strategies. For instance, some participants identified a lack of trust in the language features because of a lack of clarity of their composition: "I did not like making a decision based on you saying whether the language measures were good or bad, I wanted to understand the language measures better." Others commented on the need for additional interface features, like a help menu, to aid in this intensive cognitive process: "it would be beneficial to have a 'help' section ON the platform to look at when needing the reminder of things the video mentioned." Some users commented on the usability of views in general, like the Entities and Tweet View. For example, one user commented "I didn't really understand the need of entities to determine fake articles." While another user admitted that "I did not use the tweets or entity features of the interface." Both comments explain users' limited use of that view but was expected given the limited training to functionality for these views.

5. Discussion and Limitations

In this section, we discuss implications of our findings on VA evaluation practices as well as consider limitations of our study along with avenues of future work.

5.1. Implications for VA Evaluation Practices

Our findings are informative for guidance on training and tutorial during visualization evaluation with human subjects. Our findings show that visual anchors and strategy cues can significantly impact users' confidence and time spent investigating in each view when performing tasks. Anchoring to a subset of views may lead to the over-reliance on (often incomplete) information presented in those views, thus preventing users from getting a comprehensive picture.

Such anchoring effects could occur due to how participants are trained to use the visual interface before carry out the tasks. First, providing a general training video is a good idea, however, careful considerations are needed when devising a script or training video. The experimenter may want to make sure that all important features/views get equal coverage in the script and video.

Since our experiments show that visual anchors can indeed impact multiple performance metrics (confidence, accuracy, time to decision), we would like to raise awareness of participants possibly being unintentionally anchored and suggest careful consideration on how to train users to use a visual interface.

[§§] See 03-logs.html in the supplemental materials for all 94 users' plots.

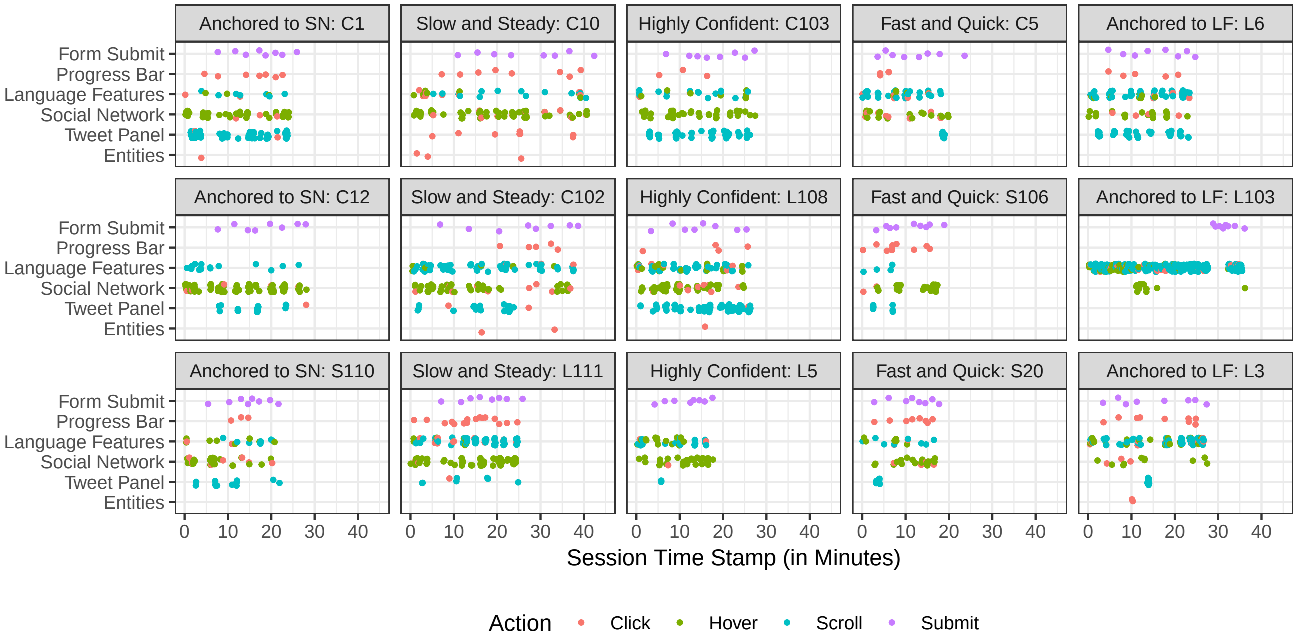

Figure 12: *Experiment interaction logs of Verifi. Each plot is a user's interaction log. Each dot is a user action: click (red), hover (green), scroll (blue), and submit (purple). The x-axis is the time of the action. The y-axis is the respective view associated with that action. The order corresponds to critical functionality (e.g., Form Submit) to primary view (e.g., Language Features vs. Social Network) to secondary views (e.g., Tweet Panel or Entities). Chart columns indicate user-level strategies based on user-level dendrogram clustering. Chart row order represents, in descending order, highly accurate users (7+ out of 8, top row), average users (5-6 out of 8, middle row), and inaccurate users (4 or less of 8, bottom row).*

5.2. Limitations and Future Work

While we attempted to avoid negative impacts to validity, there are several limitations to generalizing our results.

First, there are limits to studying users' behavior through interactions. A different approach to tracking visual anchoring could be through eye tracking to detect users' attention directly rather than through interactions. However eye tracking too presents challenges of its accuracy (especially for a multi-view interface). Future work in visual anchors should consider additional ways to analyze and measure user interactions.

Second, given the complex nature of the interface, we recruited highly trained students in computer science and data science who had some experience in visual analytics, machine learning, or social media communications. Consequently, our results may not generalize for a broader population (i.e., no experience in visual analytics). Future work could develop simpler interfaces that could be more appropriate for testing within broader participants pools like crowdsourcing (e.g., MTurk). Related, the choice of accounts can affect the difficulty of the decision-making task. If we selected different Twitter accounts, we may find our treatments have a different effects for our decision-making task.

Third, our study did not consider manipulating the interface design. While the training process differed between groups, all users received the same interface. As argued by Pu and Kay [PK18], design may have a significant effect on the forking paths problem as well. A future study could provide control interface layouts to identify the marginal value of each view in the decision-making process (e.g., testing whether the strategy cues with only the Tweet view – which mimics everyday social media usage – can measure a baseline accuracy). With such a baseline, a more precise estimate of the effect of the visualizations can be inferred. Another possible design enhancement could include adding uncertainty, like [FWM*18], encoding to the visualization (e.g., confidence intervals of each account based on past users' accuracy).

Last, cognitive science has developed Bayesian, rational computational models for understanding cognitive biases like numerical anchoring [LGHG17]. Such theoretical models–like Wu *et al.* [WXCW17]–can provide testable hypotheses that may aid future studies of cognitive biases in visual analytics. Two promising avenues to facilitate such cognitive modeling is through the incorporating prior knowledge [KRH17,KRH18] and the addition of incentives and decision-theory within visualization tasks [FWM*18].

6. Conclusion

In this paper, we presented an experiment on the role of visual anchoring in misinformation decision-making in a CMV VA system. We find that providing visual anchors and strategy cues can greatly affect users' confidence but have mixed results on users' speed and decision accuracy. Visual anchors can also play a role in secondary outcomes like users' view importance ratings and use of provided strategy cues. Last, exploration of user interaction logs can provide hints to users' strategies and the effects such treatments can have for certain users. While we find that some users are susceptible to such anchoring, others can ignore such treatments–perhaps due to uncertainty or a lack of trust–leading user attributes like motivation or interest can explain more of the users' knowledge seeking behaviors.

References

[AGP17] AMER T., GOZLI D. G., PRATT J.: Biasing spatial attention with semantic information: an event coding approach. *Psychological research* (2017), 1–19. 2

[B*17] BÜRKNER P.-C., ET AL.: brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software 80*, 1 (2017), 1–28. 6

[BBM*15] BLACKWELL S. E., BROWNING M., MATHEWS A., PICTET A., WELCH J., DAVIES J., WATSON P., GEDDES J. R., HOLMES E. A.: Positive imagery-based cognitive bias modification as a web-based treatment tool for depressed adults: a randomized controlled trial. *Clinical Psychological Science 3*, 1 (2015), 91–111. 2

[BBP17] BONARETTI D., BARTOSIAK M. Ł., PICCOLI G.: Cognitive anchoring of color cues on online review ratings. 2

[CWK*17] CHO I., WESSLEN R., KARDUNI A., SANTHANAM S., SHAIKH S., DOU W.: The anchoring effect in decision-making with visual analytics. In *IEEE Conference on Visual Analytics Science and Technology (VAST)* (2017). 1, 2, 3

[DBD17] DIMARA E., BEZERIANOS A., DRAGICEVIC P.: The attraction effect in information visualization. *IEEE transactions on visualization and computer graphics 23*, 1 (2017). 2

[DDB16] DIMARA E., DRAGICEVIC P., BEZERIANOS A.: Accounting for availability biases in information visualization. *arXiv preprint arXiv:1610.02857* (2016). 2

[DFP*19] DIMARA E., FRANCONERI S., PLAISANT C., BEZERIANOS A., DRAGICEVIC P.: A task-based taxonomy of cognitive biases for information visualization. *IEEE transactions on visualization and computer graphics* (2019). 1, 2

[DJDO18] DELPISH R., JIANG S., DAVIS L., ODUBELA K.: A visual analytics approach to combat confirmation bias for a local food bank. In *International Conference on Applied Human Factors and Ergonomics* (2018), Springer, pp. 13–23. 2

[Dra16] DRAGICEVIC P.: Fair statistical communication in hci. In *Modern Statistical Methods for HCI*. Springer, 2016, pp. 291–330. 4

[ED15] ELLIS G., DIX A.: Decision making under uncertainty in visualisation? In *IEEE Conference on Visual Analytics Science and Technology (VAST)* (2015). 2

[EG06] EPLEY N., GILOVICH T.: The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological science 17*, 4 (2006), 311–318. 1, 2

[FWM*18] FERNANDES M., WALLS L., MUNSON S., HULLMAN J., KAY M.: Uncertainty displays using quantile dotplots or cdfs improve transit decision-making. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (2018), ACM, p. 144. 4, 9

[Gal15] GALILI T.: dendextend: an r package for visualizing, adjusting, and comparing trees of hierarchical clustering. *Bioinformatics* (2015). doi:10.1093/bioinformatics/btv428. 4

[GOSS17] GALILI T., O'CALLAGHAN A., SIDI J., SIEVERT C.: heatmaply: an r package for creating interactive cluster heatmaps for online publishing. *Bioinformatics* (2017). doi:10.1093/bioinformatics/btx657. 4

[GSC16] GOTZ D., SUN S., CAO N.: Adaptive contextualization: Combating bias during high-dimensional visualization and data selection. In *Proceedings of the 21st International Conference on Intelligent User Interfaces* (2016), ACM, pp. 85–95. 2

[IIC*13] ISENBERG T., ISENBERG P., CHEN J., SEDLMAIR M., MÖLLER T.: A systematic review on the practice of evaluating visualization. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (Dec 2013), 2818–2827. doi:10.1109/TVCG.2013.126. 2

[Kah16] KAHNEMAN D.: 36 heuristics and biases. *Scientists Making a Difference: One Hundred Eminent Behavioral and Brain Scientists Talk about Their Most Important Contributions* (2016), 171. 1

[Kay18] KAY M.: *tidybayes: Tidy Data and Geoms for Bayesian Models*, 2018. R package version 1.0.0. URL: https://mjskay.github.io/tidybayes/. 6

[Kei02] KEIM D. A.: Information visualization and visual data mining. *IEEE transactions on Visualization and Computer Graphics 8*, 1 (2002), 1–8. 1

[KNH16] KAY M., NELSON G. L., HEKLER E. B.: Researcher-centered design of statistics: Why bayesian statistics better fit the culture and incentives of hci. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (2016), ACM, pp. 4521–4532. 4

[KRH17] KIM Y.-S., REINECKE K., HULLMAN J.: Explaining the gap: Visualizing one's predictions improves recall and comprehension of data. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (2017), ACM, pp. 1375–1386. 9

[KRH18] KIM Y.-S., REINECKE K., HULLMAN J.: Data through others' eyes: The impact of visualizing others' expectations on visualization interpretation. *IEEE Transactions on Visualization & Computer Graphics*, 1 (2018), 1–1. 9

[KWS*18] KARDUNI A., WESSLEN R., SANTHANAM S., CHO I., VOLKOVA S., ARENDT D., SHAIKH S., DOU W.: Can you verifi this? studying uncertainty and decision-making about misinformation in visual analytics. *The 12th International AAAI Conference on Web and Social Media (ICWSM)* (2018). 2, 3, 4, 5, 6

[LBB*18] LAZER D. M., BAUM M. A., BENKLER Y., BERINSKY A. J., GREENHILL K. M., MENCZER F., METZGER M. J., NYHAN B., PENNYCOOK G., ROTHSCHILD D., ET AL.: The science of fake news. *Science 359*, 6380 (2018), 1094–1096. 1

[LGHG17] LIEDER F., GRIFFITHS T. L., HUYS Q. J., GOODMAN N. D.: The anchoring bias reflects rational use of cognitive resources. *Psychonomic bulletin & review* (2017), 1–28. 1, 9

[LM10] LAM H., MUNZNER T.: A guide to visual multi-level interface design from synthesis of empirical study evidence. *Synthesis Lectures on Visualization 1*, 1 (2010), 1–117. 1

[ML14] MURTAGH F., LEGENDRE P.: Ward's hierarchical agglomerative clustering method: which algorithms implement ward's criterion? *Journal of classification 31*, 3 (2014), 274–295. 4, 8

[Mun14] MUNZNER T.: *Visualization analysis and design*. CRC press, 2014. 1

[PC05] PIROLLI P., CARD S.: The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis* (2005), vol. 5, McLean, VA, USA, pp. 2–4. 2

[PCR18] PENNYCOOK G., CANNON T., RAND D. G.: Prior exposure increases perceived accuracy of fake news. 1

[PK18] PU X., KAY M.: The garden of forking paths in visualization: A design space for reliable exploratory visual analytics. 1, 9

[Poh04] POHL R.: *Cognitive illusions: A handbook on fallacies and biases in thinking, judgement and memory*. Psychology Press, 2004. 2

[Rob07] ROBERTS J. C.: State of the art: Coordinated & multiple views in exploratory visualization. In *Coordinated and Multiple Views in Exploratory Visualization, 2007. CMV'07. Fifth International Conference on* (2007), IEEE, pp. 61–71. 1

[RWP15] RAJSIC J., WILSON D. E., PRATT J.: Confirmation bias in visual search. *Journal of experimental psychology: human perception and performance 41*, 5 (2015), 1353. 2

[SMD15] SHAFFER D. M., MCMANAMA E., DURGIN F. H.: Manual anchoring biases in slant estimation affect matches even for near surfaces. *Psychonomic bulletin & review 22*, 6 (2015), 1665–1670. 2

[SSW*17] SHU K., SLIVA A., WANG S., TANG J., LIU H.: Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter 19*, 1 (2017), 22–36. 1

[TK74] TVERSKY A., KAHNEMAN D.: Judgment under uncertainty: Heuristics and biases. *science 185*, 4157 (1974), 1124–1131. 1, 2

[Tuk77] TUKEY J. W.: *Exploratory data analysis*. Addison-Wesley Pub. Co., Reading, Mass., 1977. 1

[VRA18] VOSOUGHI S., ROY D., ARAL S.: The spread of true and false news online. *Science 359*, 6380 (2018), 1146–1151. 1

[VSJH17] VOLKOVA S., SHAFFER K., JANG J. Y., HODAS N.: Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on twitter. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (2017), vol. 2, pp. 647–653. 3

[VZS17] VALDEZ A. C., ZIEFLE M., SEDLMAIR M.: A framework for studying biases in visualization research. In *DECISIVe: Workshop on Dealing with Cognitive Biases in Visualizations* (2017). 1, 2

[VZS18] VALDEZ A. C., ZIEFLE M., SEDLMAIR M.: Priming and anchoring effects in visualization. *IEEE transactions on visualization and computer graphics 24*, 1 (2018), 584–594. 2

[WBFE17] WALL E., BLAHA L. M., FRANKLIN L., ENDERT A.: Warning, bias may occur: A proposed approach to detecting cognitive bias in interactive visual analytics. In *IEEE Conference on Visual Analytics Science and Technology (VAST)* (2017). 1, 2, 4, 5

[WBP*17] WALL E., BLAHA L., PAUL C. L., COOK K., ENDERT A.: Four perspectives on human bias in visual analytics. In *DECISIVe: Workshop on Dealing with Cognitive Biases in Visualizations* (2017). 2

[WK16] WOBBROCK J. O., KAY M.: Nonparametric statistics in human–computer interaction. In *Modern statistical methods for HCI*. Springer, 2016, pp. 135–170. 4

[WSS17] WRIGHT W., SHEFFIELD D., SANTOSA S.: Argument mapper: Countering cognitive biases in analysis with critical (visual) thinking. In *Information Visualisation (IV), 2017 21st International Conference* (2017), IEEE, pp. 250–255. 2

[WXCW17] WU Y., XU L., CHANG R., WU E.: Towards a bayesian model of data visualization cognition, 2017. 9

[ZZZK18] ZGRAGGEN E., ZHAO Z., ZELEZNIK R., KRASKA T.: Investigating the effect of the multiple comparisons problem in visual analysis. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (2018), ACM, p. 479. 1