THE ROLE OF EMOTION IN VISUALIZATION

by

Lane Harrison

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Computing and Information Systems

Charlotte

2013

Approved by:

____

Dr. Aidong Lu

____

Dr. Jing Yang

____

Dr. Zachary Wartell

____

Dr. William Ribarsky

____

Mr. Eric Sauda

ABSTRACT

LANE HARRISON. The role of emotion in visualization. (Under the direction of DR. AIDONG LU)

The popular notion that emotion and reason are incompatible is no longer defensible. Recent research in psychology and cognitive science has established emotion as a key element in numerous aspects of perception and cognition, including attention, memory, decision-making, risk perception, and creativity. This dissertation centers around the observation that *emotion influences many aspects of perception and cognition that are crucial for effective visualization.*

First, I demonstrate that emotion influences accuracy in fundamental visualization tasks by combining a classic graphical perception experiment (from Cleveland and McGill) with emotion induction procedures from psychology (chapter 3). Next, I expand on the experiments in the first chapter to explore additional techniques for studying emotion and visualization, resulting in an experiment that shows that performance differences between primed individuals persist even as task difficulty increases (chapter 4). In a separate experiment, I show how certain emotional states (i.e. frustration and engagement) can be inferred from visualization interaction logs using machine learning (chapter 5). I then discuss a model for individual cognitive differences in visualization, which situates emotion into existing individual differences research in visualization (chapter 6). Finally, I propose an preliminary model for emotion in visualization (chapter 7).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION: EMOTION AND VISUALIZATION

Visual reasoning involves many cognitive processes including memory, attention, learning, creativity, decision-making, and problem-solving [21, 27, 32, 43]. Yet research in psychology has found that short-term changes in emotion influence many of these same cognitive processes [38, 54, 39]. While researchers have begun to identify key factors that influence cognition (e.g. emotion and personality traits), the impact that such factors have on a user's ability to perceive and reason with information remains largely unexplored.

Emotion, typically referred to as *affect* in psychology, is often defined by values along two dimensions: *valence* (positive or negative feelings) and *arousal* (the intensity of those feelings). To study the impact of affect on various cognitive tasks, many studies in psychology involve manipulating valence and/or arousal via emotional stimuli. This technique of inducing emotion in human participants is referred to as *affective priming*.

Affect has been shown to influence both low-level cognitive abilities (attention, memory) [4, 113] as well as high-level abilities (creativity, decision-making, problem-solving) [7, 12, 66]. Research in visualization has identified a close relationship between a user's cognitive abilities with their ability to perceive and understand information in visualizations [21, 27], which suggests that there might a relationship between affect and visual judgment.

To determine whether emotion impacts basic visual judgment (e.g. comparing bars in a bar chart), I designed the experiment presented in chapter 3. This experiment replicates a classic visualization study and adds affective priming techniques (using stories) to induce either positive or negative emotion in the participants during the visualization tasks. The results of this experiment indicate that emotion can significantly impact low-level, fundamental visualization tasks.

The experiments in chapter 3 also raised new questions, such as whether alternate priming techniques could be used, if the chart tasks themselves induced an emotional change in participants, and how resilient the prime was to task difficulty. I conducted additional experiments to address these questions, described in chapter 4. These experiments show that a) images are also suitable primes for visualization tasks, b) that chart tasks themselves can cause a change in emotional response, and c) that increasing task difficulty leads to a large performance difference between primed groups.

Another line of experiments explores whether emotion is reflected in visualization interaction patterns (chapter 5). Participants first completed several tasks using a visualization tool. Next, they specified at what points during the tasks they became noticeably frustrated by viewing a recording of their face and screen during the tasks. Afterwards, we used the annotated data along with machine learning techniques to build a model for frustration. Moments of frustration or engagement were detected with approximately 70% accuracy.

Next, to provide necessary context for how emotion fits into the existing individual differences research in visualization, I discuss a model that describes individual cognitive differences in visualization in three dimensions: cognitive traits, cognitive states, and experience/bias (chapter 6). Specifically, emotion is a cognitive state, which is different than cognitive traits (e.g. personality traits) and any experience or biases a user may have. Yet all of these aspects can impact how a user performs with any given visualization.

To conclude, I leverage the results of the previous chapters to propose a preliminary model for emotion in visualization. I also discuss how emotion may relate to existing perceptual and cognitive models in visualization. This model lays the groundwork for future experiments in order to better establish the role of emotion in visualization.

CHAPTER 2: RELATED WORK

2.1 Affect and Cognition

There is a large body of research on affect and its influence on various cognitive processes. These processes include attention, memory, creativity, problem-solving, and decision-making [7, 38, 54, 113, 114]. Early literature in this area focused on behavioral studies, but has recently been supplemented by studies which have identified the neurological factors related to emotion and cognition [114]. Here, we discuss research as it relates to the cognitive processes used in visualization.

Several studies have explored the relationship between emotion and attention [82, 114], and specifically how attention relates to visual processes [113]. These studies show that when affective priming precedes cognitive tasks, the effect of the prime is carried over and influences attention regulation [82, 90]. Positive priming is usually linked to better performance [96], but there is also evidence that negative priming can produce desirable behavior [39] such as increased caution in decisions involving risk. For a discussion on possible neurological factors and other attention regulation processes, see Vuilleumier and Huang [114].

In addition to impacting attention, emotion has been shown to play a role in memory [56]. Studies show that emotion can have a significant impact on long-term recall [7], as well as influence working memory [4, 59]. Recent research also shows that negative affect can support memory of details, while positive affect increases working memory performance and supports memory of the “big picture” of situations [4, 59]. Another experiment found that anxiety (which falls under negative affect) can disrupt visuospatial working memory, leaving verbal working memory intact [101]. The experiment described in chapter 3 is partially based on these prior findings, and applies the same text-based affective priming techniques as presented by Goeritz and Verheyen [41, 111].

Judgment (or decision-making) is defined as the process by which one assesses future outcomes and examines various alternatives, typically through cost-benefit analysis [8]. Emotion is a significant component in decision-making, as each possible outcome is associated with emotional consequences [8, 69]. In fact, emotion is so strongly linked to decision-making that persons who are perceptually and cognitively normal, except for an inability to experience emotions, are incapable of making even basic decisions [30]. Other studies have shown that affect also plays a role in planning and risk perception [12, 69]. The observation that emotion impacts higher level cognitive process motivates the experiment described in chapter 5: If emotion influences how we plan and make decisions, are we able to infer emotional state from visualization interaction logs, which reflect decision making and planning?

Positive emotions have been shown to support creativity and open-ended reasoning [38]. A recent crowdsourced experiment from the HCI community showed that positive priming (using images) increased the quality of ideas generated on a range of tasks requiring creativity [66]. These results were used to make design suggestions for creativity-support software. Positive emotion is also linked to better performance in creative and general problem-solving [38, 54, 55].

In general, positive emotions lead to better cognitive performance, and negative emotions (with some exceptions) lead to decreased performance. Yet we currently do not know to what extent affect can impact visualization use and performance. Hence the motivation for the experiment in chapter 3: to prime participants with positive or negative affect and to measure their performance on visual judgment tasks.

As indicated earlier, emotion is typically defined along two axes: valence and arousal. These emotions are mapped to varying levels of arousal and valence in [97], and the corresponding two-dimensional diagram is shown in figure 1.

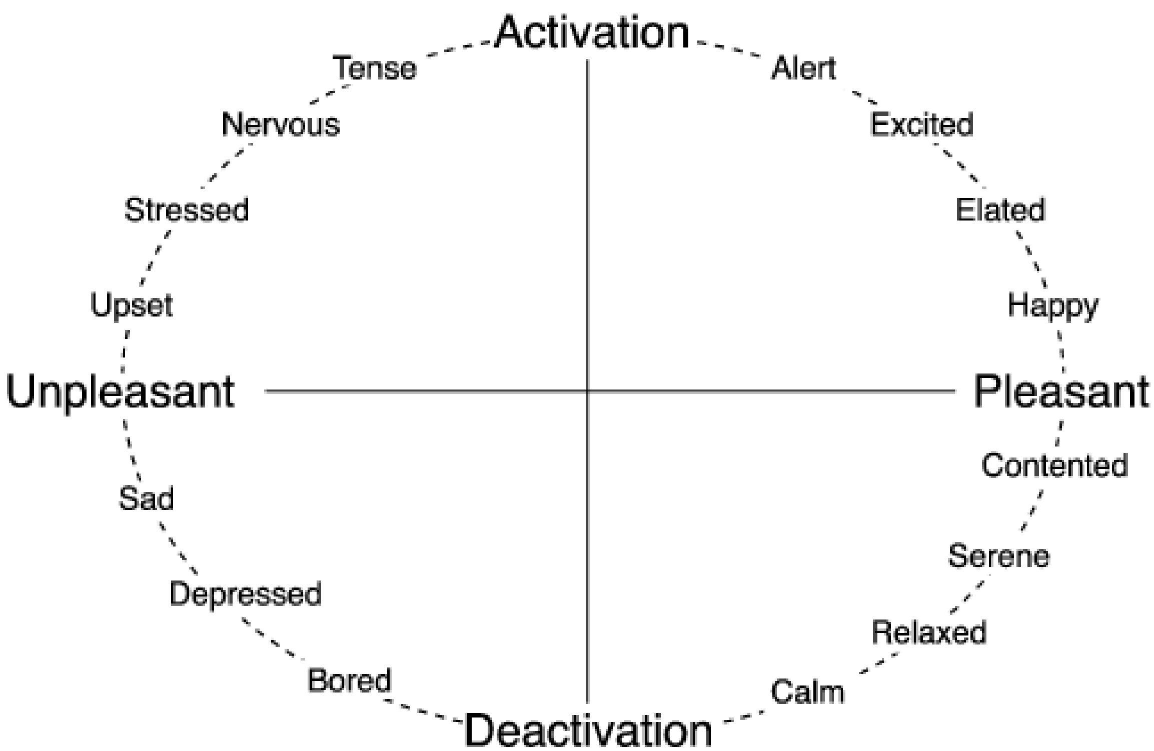

Figure 1: Relative positions of common emotions on valence and arousal scales [97].

In psychology research, several methods have been developed for measuring valence (positive/negative feelings) and arousal (the strength of the feelings). Some of these include physiological sensors, face-video processing, and neurological sensors [65, 84]. Physiological sensors excel at detecting arousal, but less so at valence [83]. Given the current limitations of physiological sensors, the most widely used and validated method for reporting affect is the survey/self-report, especially for assessing immediate emotion (as opposed to, for instance, emotion experienced a week prior) [83]. The Self-Assessment Manikin (SAM) scale, shown in Figure 2, has been successfully used in crowdsourced experiments on affective priming and creativity in HCI by Lewis et al. [66]. Based on these results, we also use SAM in our experiments.

2.2 Affective Computing

The area of affective computing seeks to equip computer systems with means to recognize and respond to human emotion. Affective computing techniques are currently used in virtual humans and intelligent tutoring systems [13]. The implications of emotions when interacting with such systems, however, are different than the implications of emotions when an analyst interacts with a visual analytics system.

Much work has been done in the affective computing community regarding how best to extract and track a user's affective state when interacting with a computer. Many of these techniques require expensive physiological sensors, such as a modified mouse and keyboards that measure pressure, or computationally expensive facial image processing techniques [83]. This kind of equipment may not always be available in visual analytics contexts, but interaction logs can be easily implemented, collected, and analyzed. Therefore, it is important to explore how much of the analysts' reasoning and mental processes can be captured by visualization interaction logs (which

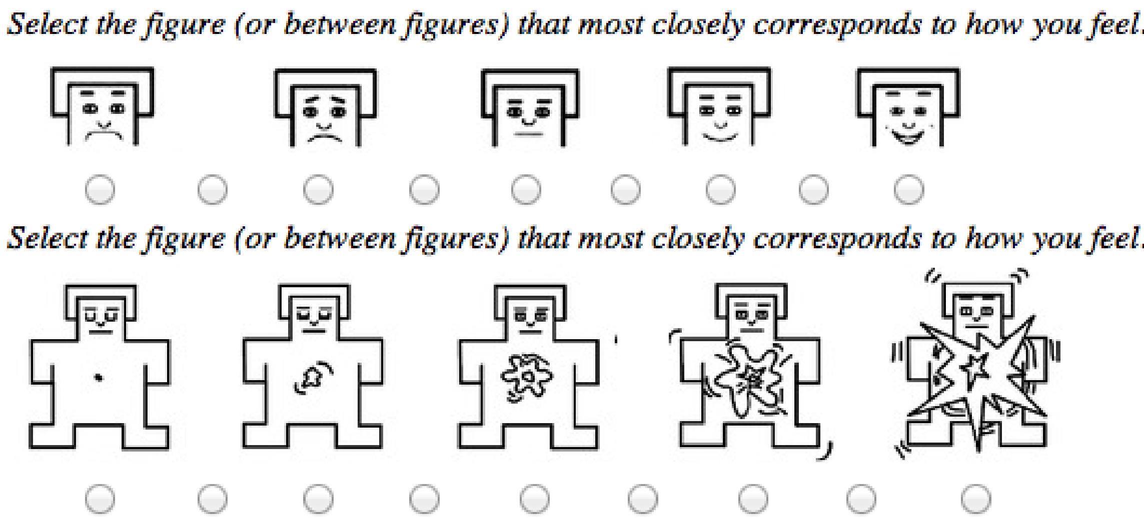

Figure 2: The 9-point Self-Assessment Manikin (SAM) scale for measuring valence (top) and arousal (bottom) [63].

is the focus of the experiment in chapter 5).

2.3 Graphical Perception

Graphical perception is a fundamental component of visualization, and is the subject of several recent experiments [47, 53]. The results from these studies have been used extensively in both the development of automatic presentation systems [71] and as a basis for visualization design principles [10].

Recently, Heer and Bostock replicated and extended a portion of Cleveland and McGill's experiments [47] on various chart types using AMT (see Figure 3). The rankings obtained in their crowdsourced studies were similar to Cleveland and McGill's rankings. These results validated AMT as a viable, scalable platform for visualization research.

Hullman et al. also leverage AMT and revise Heer and Bostock's graphical perception experiment [53] to explore the impact of social information on visual judgment.

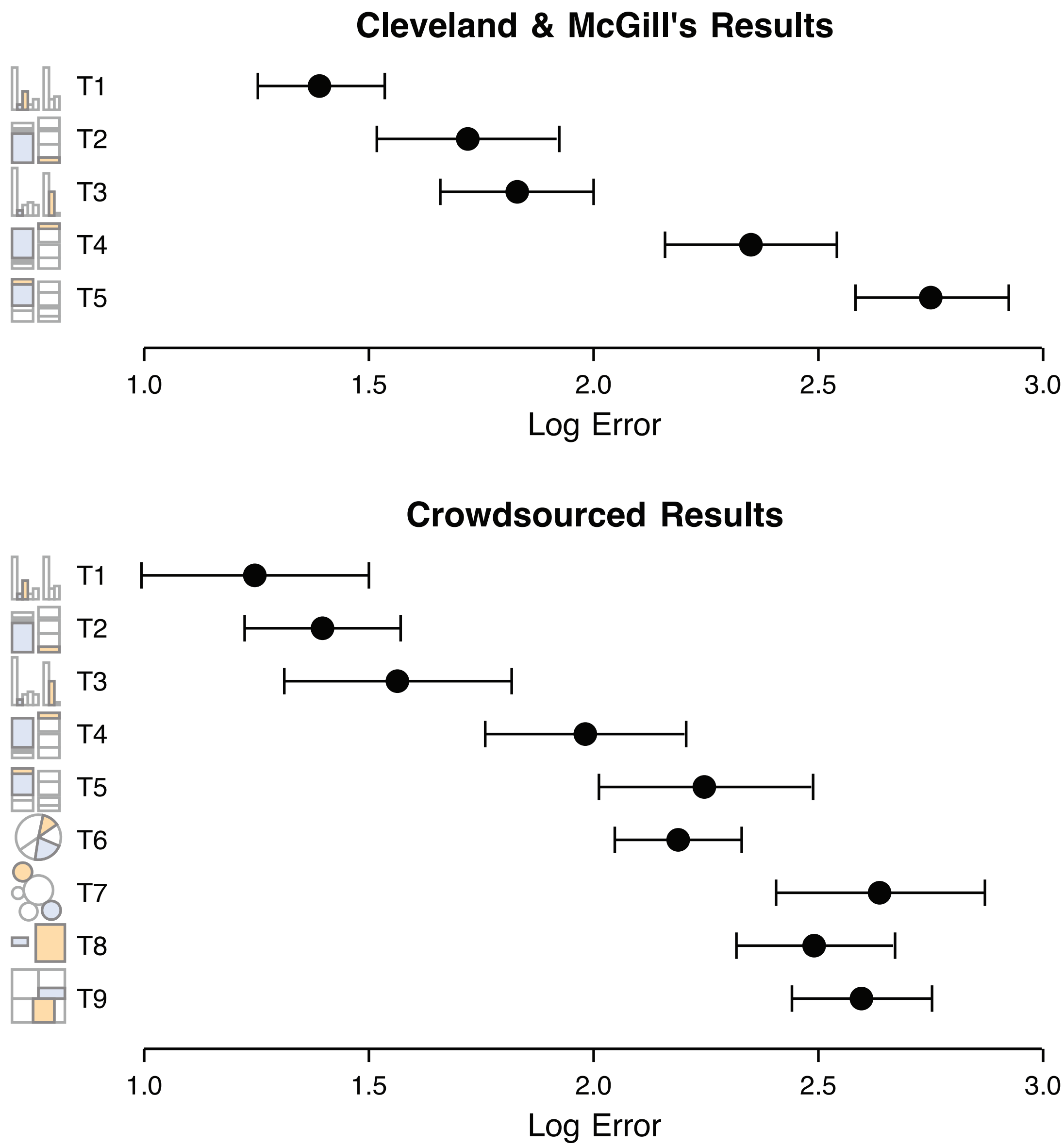

Figure 3: Cleveland and McGill's results and Heer and Bostock crowdsourced replication (from [47]).

Their results suggest that visual judgment can be influenced by adding social information, which has implications in collaborative and social visualization environments. Similarly, in the following chapter, we show that some of the most fundamental graphical perception tasks can be influenced by emotion.

CHAPTER 3: EMOTION INFLUENCES VISUAL JUDGMENT

Cleveland and McGill's seminal results [25] serve as a basis for this graphical perception experiment, as they have been replicated and extended by Heer and Bostock [47] as well as adapted to study the effect of social information on visual judgments by Hullman et al. [53]. Specifically, we follow Cleveland and McGill's experiments which compare user accuracy with various position-length chart types (see Figure 1). We also extend their work by building on both recent crowdsourced graphical perception studies [47, 53] and affective priming studies [66, 41] to examine how emotion influences visual judgment [45].

This experiment combines affective priming techniques with visual judgment tasks using Amazon's Mechanical Turk (*AMT hereafter*). We initially performed two pilot studies to verify our study design, emotional stimuli, and emotion-measurement metrics. For our main experiment, we recruited 963 participants to be either negatively or positively primed and to complete several visual judgment tasks similar to those found in Cleveland and McGill [25], Heer and Bostock [47], and Hullman et al. [53]. Our results indicate that affective priming influenced the visual judgments of participants, and that positive priming improved accuracy.

This experiment makes a number of contributions to our understanding of how emotion, a fundamental component of human cognition, relates to graphical perception, a fundamental component of visualization:

- First, we present experimental evidence demonstrating that affective priming can significantly influence visual judgment accuracy.
- Second, we corroborate these findings with related research in cognitive psychology and discuss how changes in affect arise in visualization contexts.
- Finally, we describe several cases in which these results can be used to help choose effective visualizations for different audiences and contexts.

To avoid confusion, we note a caveat about the use of the word *perception* in graphical perception studies. In these studies, participants must not only identify (perceive) the graphical elements (bars, pie slices), but also make *judgments* on the proportion of one element to another. Since these judgments involve attention, working memory, and mental translation/manipulation, they are cognitive in nature. One recent study uses the term "visual judgment", which we adopt here [53]. Throughout this section, we refer to *graphical perception studies* which involve *visual judgments*.

3.1 Experiment Design

There are several key differences between existing techniques for research in affective priming and graphical perception that render their integration non-trivial. For example, affective priming studies are often between-subjects, since both the exact duration of a prime and the interactions between subsequent negative and positive primes is uncertain [48]. In contrast, graphical perception studies are typically within-subjects and require participants to interact with multiple charts, making participation time longer [47]. Given the intrinsic differences between these two types of studies, we discuss how we altered previous graphical perception experiment design

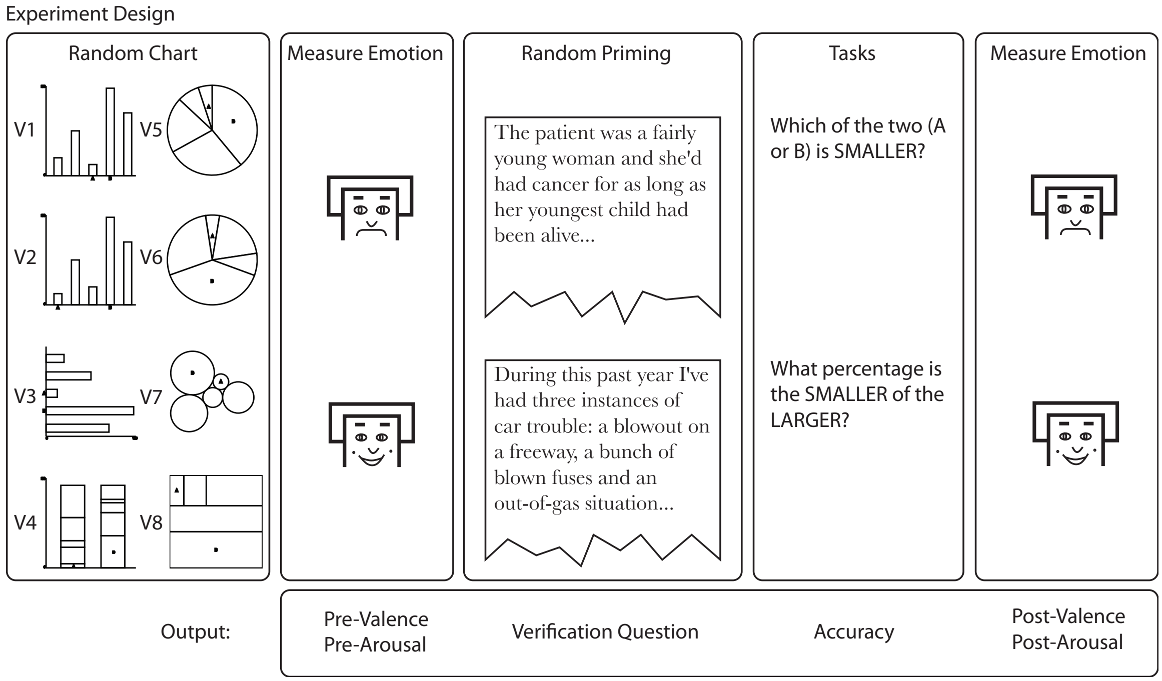

Figure 4: In our main experiment, participants were given either negative or positive text-based emotional-stimuli and completed visual judgment tasks with one of the eight chart types.

to be more suitable for affective priming. Figure 4 shows an overview of the resulting study design and procedure.

3.1.1 Charts

Several previous studies on graphical perception included 10 datasets and allowed participants to work with all datasets and chart types. As our first task was to adapt a graphical perception study for affective priming, we found the length of previous studies to be a concern. The longer the study, the more risk of losing a successful prime [48].

To address this, we first reduced the number of randomly-generated test datasets from 10 to 5. We ensured that the proportions of the marked elements that participants were to judge were similar to Cleveland and McGill's study, with ratios ranging from 13% to 85%. Furthermore, we also restrict our participants to working with one chart type and priming condition, making our study a between-subjects design. Similar to traditional priming experiments, a between-subjects design is necessary since it ensures that no participants see the priming stimulus more than once and become accustomed to it.

In Heer and Bostock's study, some charts have 5 graphical elements[1], such as pie charts, while some include 10, such as bar charts, stacked bar charts, and treemaps. We sought to have even amounts of data elements in each chart. Therefore, all of our charts consisted of 5 graphical elements, with the exception of the stacked-bar chart. The stacked bar chart contains two columns, so including only 5 elements would create an uneven number of elements in each column (e.g. 3 in one bar and 2 in the other), possibly leading to confusion. To create an even chart, we increased the number of elements in the stacked-bar to 10.

There are more length-based and rectangular charts in Heer and Bostock's study (shown in Figure 3). Also, the pie charts used in both Cleveland and McGill's as well as Heer and Bostock's study are unordered and do not start at 0 degrees, which may affect accuracy. To create a more even representation of length, angle, and area judgments, we reduce the number of stacked bar cases from three to one, and include an additional pie chart that is both ordered and starts at 0 degrees. We also include a bar chart with bars running horizontally instead of vertically, to test for a possible connection between chart orientation and embodied cognition based on work

[1]A graphical element is a component of a visualization. For example, a slice in a pie chart or a bar in a bar chart.

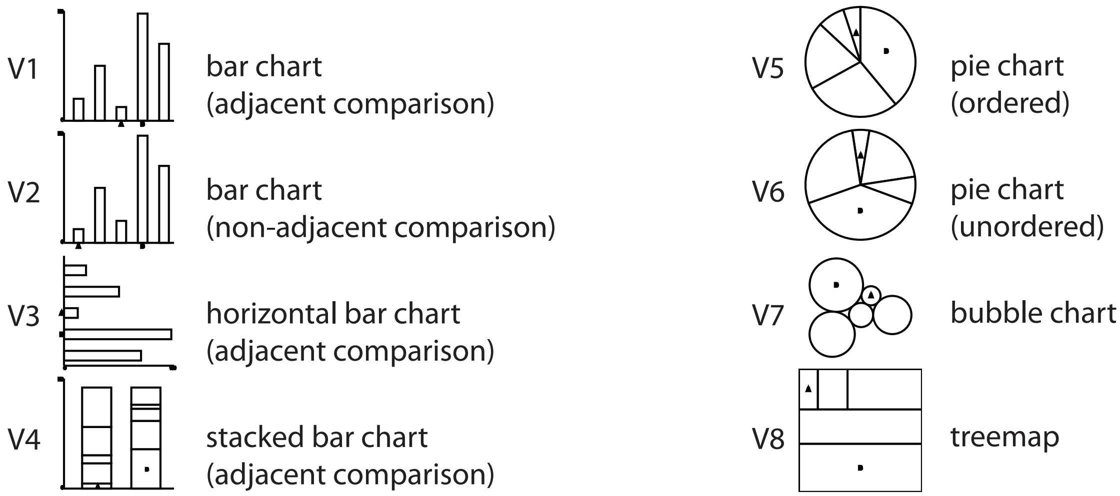

Figure 5: The charts used in our crowdsourced experiment.

by Ziemkiewicz and Kosara [122]. An example of each the 8 charts we use is shown in Figure 5.

3.1.2 Affect Measurement

We chose the Self-Assessment Manikin (SAM) scale for testing the valence and arousal of the participant. The SAM scale is widely used in psychology, in part because graphical depictions of emotion are cross-cultural [63, 66]. Unlike other surveys such as the Positive and Negative Affect Schedule (PANAS), SAM uses graphical depictions of emotions, which are both cross-cultural and take less time to complete [105]. This makes SAM a widely used tool in psychology and HCI research, especially for studies that use the AMT platform, which often involves international participants [66, 72].

Figure 2 shows the images used in SAM. Selections can be made in between the figures, for a total of 9 possible choices. Our work is primarily concerned with the effect of valence. However, we also collect arousal scores to ensure that only valence is affected by the stimuli.

3.1.3 Affect Stimuli

Several methods have been introduced for priming participants with positive or negative affect. Common methods include images, text (stories), videos, sounds, word-association tasks, and combinations thereof. Such methods are commonly referred to as Mood-Induction Procedures (MIPs). We chose text-based stories as our MIPs for several reasons.

First, verbal memory is separate from visual memory [101]. This separation is important for visual judgment tasks because using text-based stimuli helps ensure that there is no interaction or interference between the priming section of the study and the visual judgment part. Another reason is that using stories makes it easier to hide the true nature of a study. Whereas graphic images are often an obvious cue of trying to elicit an emotional response, stories can be more subtle[2]. Finally, we believe that emotional stimuli similar to our text-based MIPs are more likely to be encountered by users than the more graphic images and videos used in other studies [41].

For our main experiment, we evaluated several short stories found in the New York Times website for their emotional content. The stories were carefully selected to be relatively short (around 800 words) to ensure that participants could complete the study in a reasonable amount of time. Additionally, stories had to contain content that would be considered negative or positive by participants with divergent

[2] In chapter 4, we devised an image tagging task, which is a common Turker task, to test image priming.

cultural/ethnic backgrounds. In the end, we chose two stories from the New York Times, one negative story describing hospice care, and one positive story about the kindness of strangers.

To validate that these stories elicit the desired affective response from the participants, we ran a pilot study on AMT with $n = 40$ participants. To help ensure that the reported valence and arousal was not an effect of satisficing (i.e. participants trying to submit the answers they think the experimenters want), participants were only given the SAM as a post test in this pilot study. Satisficing was less of a concern in the full study because the stories were already determined (via the pilot) to induce the desired primes.

After accepting the task, participants read the story, with instructions to pay attention to the overall story more than the details. After the reading was complete (we made no time restrictions), they were asked a verification question (multiple choice) about the overall content of the story. The story was hidden during the verification questions. Next, they were given the SAM as a post-test. Finally, they were invited to share any additional comments about the story.

The results of the story validation showed a significant difference between the positive and negative groups for valence for both stories ($p < .05$). We also checked for significant differences in arousal, since having one story that produced low arousal and one that produced high arousal could be a confounding factor in the experiment [4]. However, no significant difference was found between stories for arousal ($p > .05$).

3.2 Pilot

In another pilot study, we tested text-based affective priming with Cleveland and McGill's position-angle experiment, which compared performance between bar charts and pie charts. The goal of this experiment was to determine whether to use raw valence scores or change in reported valence. Change in reported valence in the intended positive or negative direction is the same measure used in other affective priming studies [41, 111]. However, other studies have grouped participants by raw valence scores [66]. For instance, on our 9-point SAM scale, anyone below a 4 would be considered negatively primed, and anyone above a 7 would be considered positively primed. Alternately, if a participant reported a positive change in emotion, they would be in the positive group (and vice-versa).

After running the pilot and analyzing participants' performance, we found that performance on both charts was affected when the reported pre- and post-valence scores were changed in the intended direction ($post > pre$ for positive, and $post < pre$ for negative).

In this work, we call such a change as a "successful" prime. Based on this pilot, we formed the hypothesis that *successful affective priming will significantly influence error in visual judgment across all chart types.*

Below we will describe the experiment used to test this hypothesis. In general, we incorporate text-based affective priming techniques and follow Cleveland and McGill's study on graphical perception and Heer and Bostock's recreation of the study on AMT. In the following section, we discuss the changes we made to these studies to better suit them for affective priming.

3.3 Experiment

To validate our hypothesis that affect influences accuracy in graphical perception, we conducted a large-scale experiment using AMT. We chose to follow Heer and Bostock's crowdsourced replication of Cleveland and McGill's seminal study on graphical perception [25, 47], making necessary changes to effectively incorporate affective priming, as described in the study design. We verified the effectiveness of our emotional stimuli and determined that change in valence should be examined through two pilot studies.

Amazon's Mechanical Turk (AMT) is a service that allows workers to complete various web-based tasks in exchange for micro payments. It has been used for several graphical perception studies [47, 53], as well as studies exploring other cognitive aspects of visualization [121]. Another recent work used AMT as a platform for exploring affective priming and creativity [66].

3.3.1 Participants

We recruited 963 participants through AMT. It took a approximately five days to gather all responses.

Our study used a total of eight charts with the two priming groups (positive/negative). This resulted in a total of 16 groups (V1-Positive, V1-Negative, V2-Positive, etc.). It is well-documented that positive priming is more difficult to obtain than negative [41]. Therefore, we assigned more participants to the positive groups, to increase the number of successfully positively primed participants. Participants were randomly assigned to a priming group (with a 1/3 chance of being in the negative group, and 2/3 chance of being in the positive group) and randomly assigned to a chart group (charts shown in Figure 4).

The task was advertised on AMT as involving reading a story and answering several short questions afterwards. The task description did not reveal that we would be measuring mood or working with charts. Although we used stories found in newspapers, participants were instructed that they could quit the task at any time if they found the material offensive. After completing the study, participants were initially paid $0.20, which was later increased to $0.35 to speed completion time.

3.3.2 Procedure

After a participant has been assigned to a chart type and priming group, the participant was first given the SAM scale as a pre-test to record their valence and arousal. Next, participants were given the positive or negative story to read with the instructions that "the overall story is more important than the details". After reading was complete, participants were asked a simple verification question regarding the content of the story (the story was hidden at this point). Next, participants began the visual judgment tasks.

Each of the five tasks (all being the same chart type: bar, pie, or one of the others in Figure 5) were presented in random order. Recall that the datasets for each chart were generated in the same manner as Cleveland and McGill's and other graphical perception studies. Participants were instructed to make quick but accurate visual judgments, and to avoid using any outside tools. Similar to Heer and Bostock's study, participants were asked the following two questions about the charts: First, "Which of the two (A or B) is SMALLER?" followed by "What percentage is the SMALLER of the LARGER?".

After the visual judgment tasks were completed, participants were given the SAM again as a post-test. The SAM was given after the judgment tasks to ensure that the prime lasted through the duration of the visual judgement tasks. Finally, they were invited to submit any additional comments about the task.

3.4 Results

Our study adhered to a between-subjects design, since participants were given a single chart and priming combination. We excluded participants who incorrectly answered the story verification question, as well as those who either failed to answer or put the same answer for multiple questions (e.g. putting 50% for each judgment). In total, 299 of the 963 participants were removed, which is consistent with prior findings of error rates in AMT studies (see Bernstein et al. [9]).

Following previous graphical perception studies of Cleveland and McGill, Heer and Bostock, and Hullman et al., we use the midmeans of log absolute errors (MLAEs) using the following formula:

$$log_2(|judged\_percent - true\_percent| + .125) \tag{1}$$

Similarly, we also follow these studies in using bootstrapping to calculate the mean confidence intervals for each chart and priming combination. The resulting confidence intervals are shown in Figures 6 and 8.

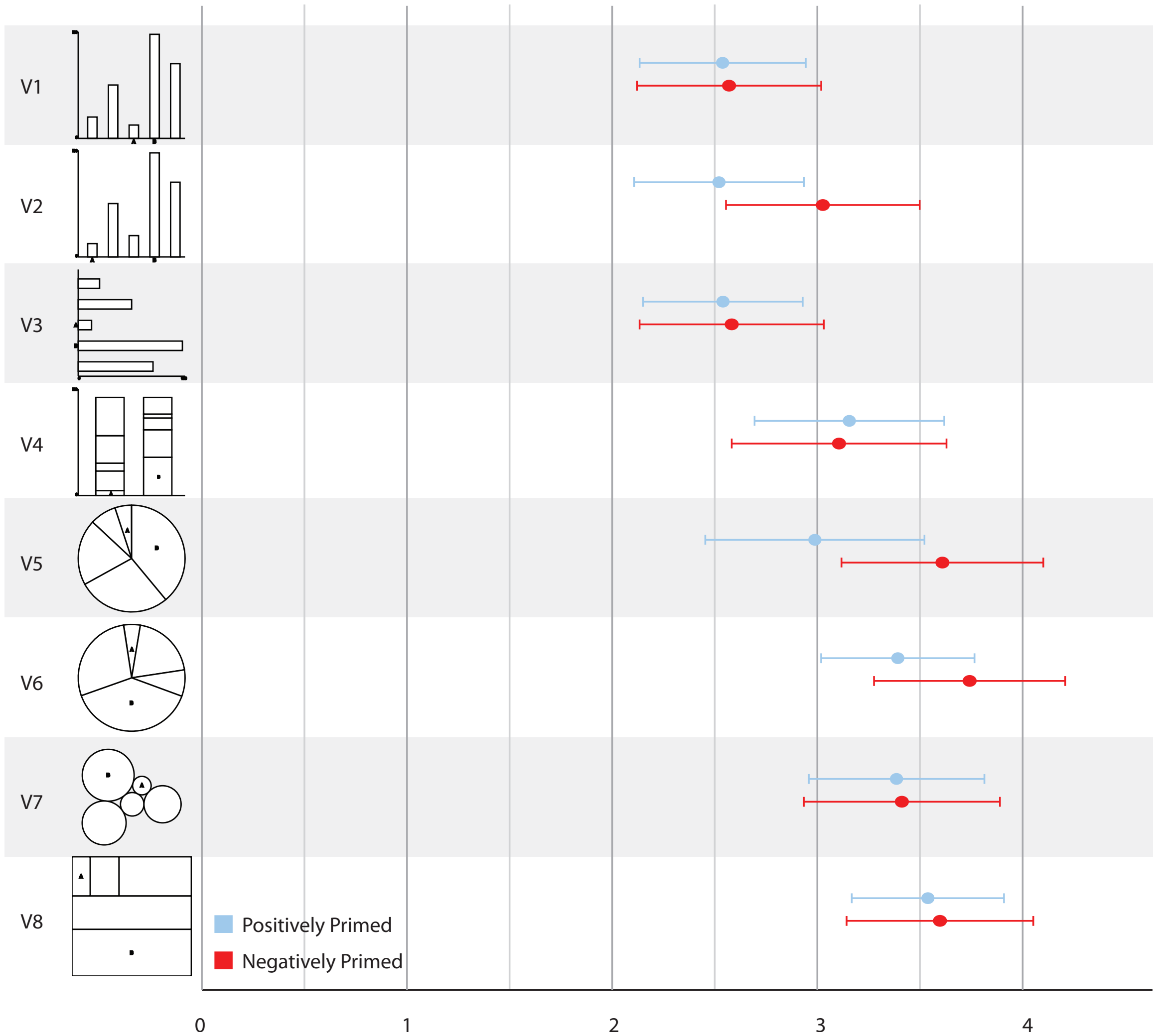

Figure 6: Results for all participants.

We divide our analysis into two cases, the first including all participants in the negative and positive groups, and the other including only participants who were successfully primed in the intended direction. The implications of each of these cases will be explored further in the discussion.

3.4.0.1 Case 1: All participants

This case includes every participant: 359 in the positive group and 305 in the negative group. Each chart and priming group has an average of 41 participants.

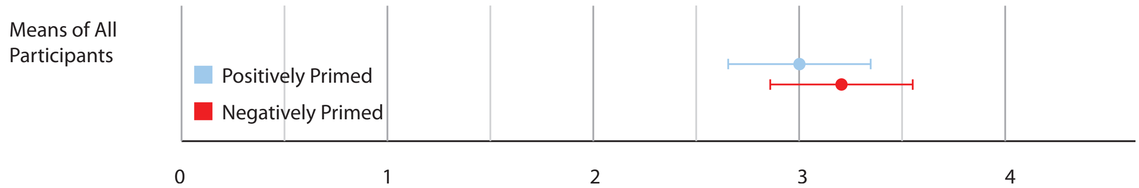

Figure 7: Total means for all participants.

This analysis consists of 3320 individual judgments.

For each chart in the graphical perception section, two data items were collected: which chart was perceived as smaller, and a number representing the percentage the smaller was of the larger (between 0 and 100). For each of these judgments, we compute the MLAEs using the formula mentioned previously. We then compute the mean error and 95% confidence intervals for each chart and priming type.

The results of these calculations are shown in Figure 6. Following Hullman et al. [53], we also take the results for all charts in each priming group together, including all chart error scores for each participant in the negative- and positive- primed groups. A t-test on these groups[3] was not significant $t(662) = 1.8318, p = .067$, This is not surprising because the difference between positive (M = 3.00, SD = 1.39) and negative (M = 3.20, SD = 1.475) is only 0.2, which is a small difference in error.

Recall that in our pilot study, we explored two common metrics for defining negative and positive groups. One approach divides groups based on raw valence scores, and the other on whether the change in pre- to post- valence scores occur in the intended direction [41, 111]. The results of the pilot study found a possible influence for change in valence on visual judgment. Based on these results, we include the change in valence metric in our second case.

[3]An unpaired t-test, corrected for multiple groups and two comparisons (requiring $p < 0.025$).

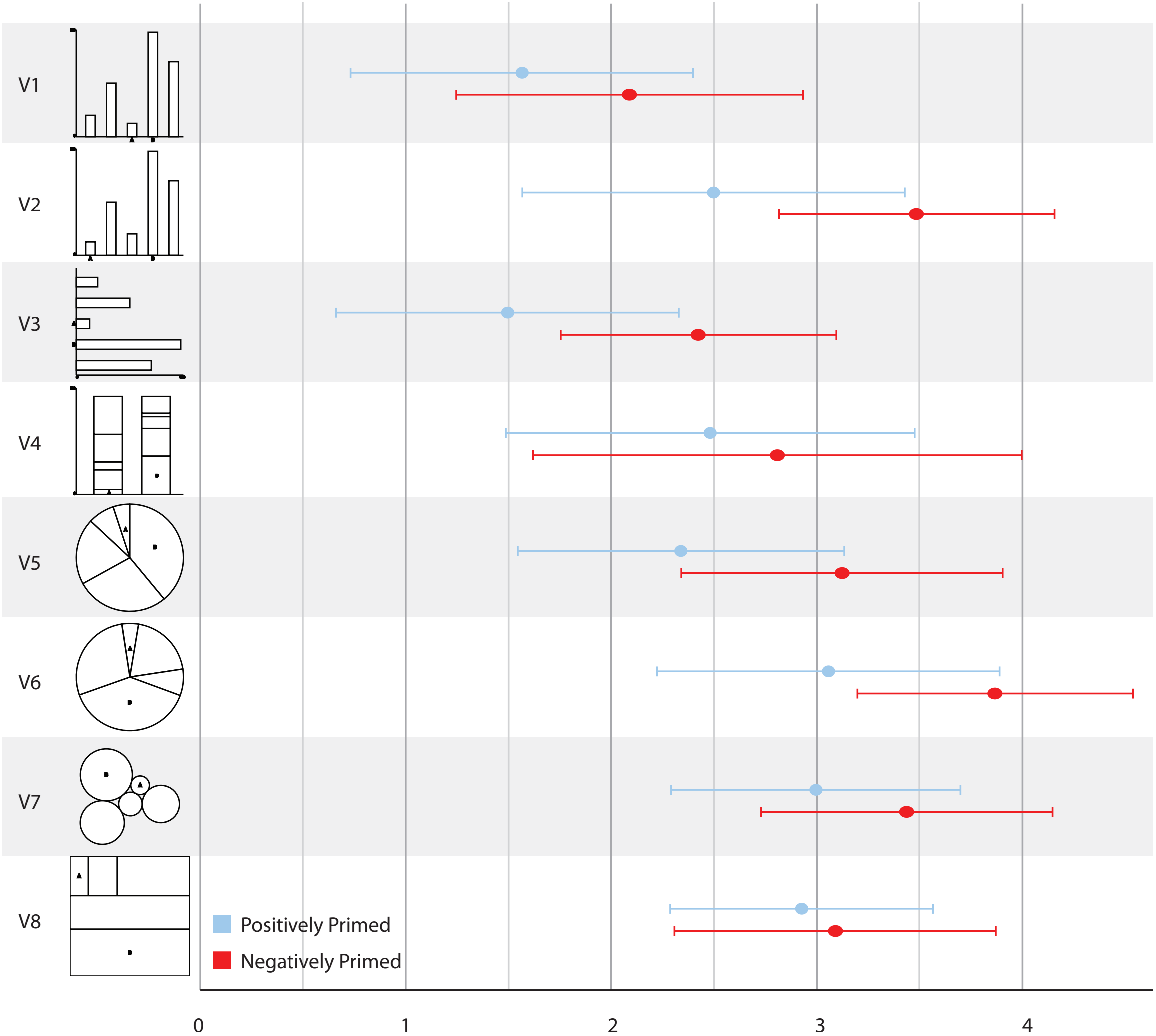

Figure 8: Results for successfully primed participants.

3.4.0.2 Case 2: Successfully primed participants

In this case we include those who were succesfully primed in the intended negative or positive direction. Specifically, we include those whose post-valence is higher than pre-valence for the positive group, and vice-versa for the negative group. This resulted in 87 participants in the positive condition and 120 in the negative condition for a total of $n = 197$. There was an average of 13 participants for a single chart and priming condition (out of 16 total possible combinations). Each participant

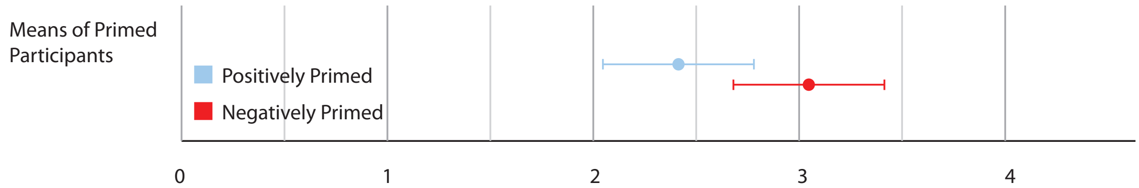

Figure 9: Total means for successfully primed participants.

made judgments on 5 charts. Therefore, this analysis consists of approximately 985 individual judgments.

As before, we compute the MLAEs for each judgment and combine these to produce a mean errors and confidence intervals for each chart and priming condition. The results of these calculations are shown in Figure 8.

We then compare positive error with negative error using a t-test. This yields a significant effect for error $t(205) = 3.1560, p = .0018$, with error in the negative group being higher than that of the positive group. The lowest error appeared in the positive group (M = 2.42, SD = 1.47), meaning participants in the positive group performed better on visual judgment tasks than those in the negative group (M = 3.05, SD = 1.49). These results are consistent with our hypothesis that affective priming can impact visual judgment performance in participants who report a change in valence scores.

An interesting finding is that it appears that positive priming tends to improve performance, rather than negative priming decreasing performance. We see this by comparing the data in Figures 7 and 9. It appears that, on average across all charts, negative priming stays about the same. To test this, we compare the successfully primed positive group (M = 2.42, SD = 1.47) against all positively primed participants (M = 3.00, SD = 1.39). Doing so, we see a significant effect for error $t(444) =$

$3.5634, p < .001$, with the successfully-primed participants having significantly lower error than the total positive group. This supports the notion that positive priming can improve visual judgment accuracy.

Other interesting findings can be seen by looking at the individual difference in charts in the Figure 6. Following Cleveland and McGill as well as Heer and Bostock, rather than test each chart pair for significant differences, we will simply discuss the relative rankings between charts and the effects of affective priming in the following section.

3.5 Discussion

Consistent with past research on the impact of affect and low level cognitive processes [82, 96], we found that affective priming also influences performance on visual judgment tasks involving several chart types commonly used in visualization [25, 47, 53]. Furthermore, as much research on affect has focused on the relationship between positive affect and increased cognitive performance [54, 55, 96], we also found that successful positive priming yielded significantly better visual judgment performance.

There are many ways in which the affect might impact performance. Given the low-level nature of the visual judgment tasks described in our experiment (the perception of larger/smaller elements and the estimation of the difference in their sizes) we find it useful to focus on low-level cognitive processes such as attention and working memory, which can be enhanced by priming positive affect [113, 4, 59].

While positive affect might improve these processes partly through higher overall engagement and alertness [96], it is also likely that different moods place observers in states that are better or worse for given tasks. For instance, positive moods can expand the scope of the conceptual 'spotlight', in creativity tasks where observers must generate new uses for tools [66], identify words that form compounds with an arbitrary set of other words [96], or solutions to anagrams [103].

But positive moods can also expand the scope of the perceptual spotlight of attention [34]. They encourage an observer to process a larger spatial area of the world in a single glance [96, 40], relative to negative or anxious moods that constrict this spatial area [35]. These effects are often demonstrated by asking observers to make a decision about a small spatial area (e.g. discriminate a letter at the center of a screen), and showing that incompatible information within the broader spatial area (e.g., another letter that tempts the 'wrong' answer) interferes more after positive mood priming than after negative or anxious mood priming.

Work from the perceptual psychology literature also suggests that some visual operations are more efficient under a broad spatial scope. These include determining the distance between two objects [16], deciding whether two objects are visually similar [37], or extracting statistical summary information (about e.g., size or orientation) from a collection of objects [23]. In contrast, a more narrow spatial scope may be necessary for processing spatial relations between objects [16, 37] or accessing fine details [82].

These differences in the perceptual scope of attention might serve as a mediating factor that causes mood to impact perceptual performance. A broad scope of attention might be beneficial for tasks where relevant information is more spatially distant, such as a bar chart with a non-adjacent comparison (v2). Unordered pie charts (v5, v6), which are positively impacted even in the experiment case with all participants (Figure 5), may similarly benefit from larger perceptual scope.

Mood may also impact perceptual performance via altering the performance of visual working memory. Visual working memory capacity may be highly influenced by the scope of information that an observer attempts to encode at once. Attempting to encode a large number of objects at once may lead to lossy encoding of each item, in contrast to attempts to encode a smaller number of objects at once with more precise encodings [112]. Changing mood might change perceptual scope, which could in turn alter the balance between the capacity and level of detail of visual memories for previously seen information. Depending on the visual task, information about more objects (but lossier) or information about fewer objects (but more detailed) might be more beneficial to observer performance.

Finally, in examining Figure 7, we can consider the trend where some charts appear to be less impacted by the affective primes - specifically treemaps (v8), bubble charts (v7), and stacked bar charts (v4). One possible reason for this stability effect is that the difficulty of these charts increases cognitive effort, which Hullman et al. suggests may lead participants to think about their decisions more carefully [52]. Specifically, both the treemaps and the bubble charts do not have axes, which Heer and Bostock discuss as a possible difficulty for users [47]. Also, the stacked bar charts used more graphical elements than the other charts, which Hollands and Spence found to increase cognitive effort [50].

3.5.1 Comparison to Previous Graphical Perception Studies

Our results preserve the general ranking from Heer and Bostock's study. However, equivalent charts in Heer and Bostock's study have lower error, despite the fact that we mimic their chart types and data-generation methods closely. We have identified possible reasons for this global increase in error, which can inform future work involving priming and graphical perception.

First, our tasks were longer than previous studies. In Heer and Bostock study [47] as well as Hullman et al. [53], each chart was given as an individual task. This means users could complete each tasks at will, rather than as part of one task. In contrast, due to our use of affective priming, our stories and charts had to be given together in one task. Additionally, Heer and Bostock as well as Hullman et al. make use of qualification tasks to verify and train participants for the chart-perception tasks. Qualification tasks serve two main functions: to ensure participants accurately follow instructions and to briefly train participants on the tasks. However, we omit a qualification since it would reveal the nature of the experiment and extend the study length. Future experiments could explore how to effectively incorporate additional quality assurance methods into crowdsourced priming experiments, perhaps based on those discussed by Mason and Suri [72].

Learning effects could also account for better performance in Cleveland and McGill as well as Heer and Bostock. In Cleveland and McGill's experiment, each participant was given 50 tasks (10 charts and 5 datasets) as opposed to our 5, so some amount of learning is possible. Heer and Bostock also used a training task, the purpose of which is to help participants learn how to work with each chart type. This may have impacted the participant's abilities in performing visual judgments.

3.5.2 Comments on Affective Priming Studies

It is worth noting that the stories we used as emotional primes came from everyday newspapers. We followed this restriction in order to create a prime that is more common in everyday life. Similarly, we did not pursue images as a priming technique, since images that tend to produce strong changes in valence are both less common and tend to cause a significant change in arousal, which can be a confounding factor in affective priming studies [114]. Based on the results of this study, which has shown that common affective primes can significantly impact visual judgment, we believe that alternate priming techniques could be evaluated in future work as a means to further explore the relationship between emotion, cognition, and visualization.

In our pilot study, we found that although it is easier to prime people with negative affect, more benefits are gained from positive priming. However, because the average pre-valence across both groups was between 6 and 7, it was more difficult to obtain a positive delta in valence. This finding is corroborated by previous studies which have noted that most people begin in a slightly positive state, and that more even results can be obtained by assigning more participants to the positive priming group [111].

3.6 Implications

We have found that affect influences visual judgment, and that positive priming increases visual judgment performance. Based on these results, we discuss some possible implications regarding visualization design and evaluation.

One specific design implication comes from our finding that non-adjacent comparisons appear to be both less accurate and more influenced by affective priming in both the all-participants and successfully-primed participants case. Therefore, when designing interactive visualizations where emotion can be a factor, designers should ensure the user should be able to interactively change a non-adjacent comparison situation to an adjacent comparison situation or to use other interactions or views to better support non-adjacent comparisons.

Other implications are more broad. For instance, as visualizations become more common in everyday life, care must be taken to ensure that information is communicated accurately for a variety of audiences and contexts. In general, we note that designers have no control over the possible priming-effects users experience before encountering a visualization. It is possible that a user could be primed positively or negatively through a variety of means, such as stories read in a newspaper or interactions with coworkers. Since the results of our study suggest that positive priming improves visual judgment accuracy, future research could explore how to integrate positive priming into design.

On the other hand, visualization tool design often includes knowledge about the *environments* in which users interact with the system. For instance, a user in a disaster-response setting may be more subject to negative priming, whereas a user in gaming may be more subject to positive priming. Such information can be used to assess the probability of negative or positive emotions and their subsequent impact on cognitive processes. One approach designers could follow is that of Lewis et al. [66], who suggest that embedded images and other interface design elements could help manage affect in software systems.

Another possible situation is where strongly-emotional content is unavoidable, such as visualizations dealing with data that has strong potential for unintentional affective priming. For example, a study by Elting et al. [33] compared visualization formats for communicating clinical trial risk. In clinical trials, a significant difference in error can often mean a number of lives lost (or saved). Other common high-stress areas where visualization has been used include criminal investigation, finance (high risk), and disease control. Given the complex nature of the decisions made in these situations, we believe that it is necessary to investigate the relationship between visualization, emotion, higher-level cognitive processes.

3.7 Limitations and Experiment Summary

While positive priming is generally associated with better cognitive performance, negative priming can also impact higher-level cognitive processes. An example a recent study by Livingston et al. which found that negative content had more impact than positive content on a user's subsequent review of a game [68]. While positive priming can be used to positively influence perceptual accuracy, it is possible that the residual effects of positive emotions may negatively influence a higher-level task.

We have described a crowdsourced experiment in which affective priming is used to influence low-level visual judgment performance. Our results suggest that affective priming significantly influences visual judgments, and that positive priming increases accuracy.

CHAPTER 4: EMOTION INFLUENCES VISUAL JUDGMENT (II)

The experiments in chapter 3 showed that inducing emotion can impact performance on Cleveland and McGill's classic graphical perception task [25], but these experiments are just a first step, and they raised several new questions:

- Do the chart tasks themselves induce an emotional change in participants?
- Could alternate priming techniques be used and do they produce the same effect as stories?
- Do the performance differences from priming hold as task difficulty increases?

In this chapter I describe additional experiments that address these questions. The first experiment tests whether chart tasks induce an emotional change by having participants report their emotional state and perform chart tasks, but *without* any priming task in between. The second experiment evaluates image-based priming on a chart task. This chart task is similar to those in chapter 3, but with a low and high difficulty condition.

In short, these experiments show that a) chart tasks themselves can cause a change in emotional response, b) images are also suitable primes for visualization tasks, and c) increasing task difficulty leads to a large performance difference between primed groups.

While these experiments may seem as somewhat incremental improvements over those in chapter 3, they serve an important purpose by replicating the main result of chapter 3 (that positively primed participants perform better), by exploring alternate methods of studying emotion and visualization, and by evaluating more difficult tasks, which helps us better understand how emotion might play a role in everyday visualization use.

4.1 Experiment Design

The general design for the two experiments in this chapter is the same, except that the first experiment includes no priming task, and the second experiment includes image priming. They share the same chart types, the same number of chart tasks, and the same emotion-measurement techniques. This section describes these components of the two experiments together, before discussing the details of each experiment.

4.1.1 Charts

One of more interesting results of the experiments in chapter 3 is that adjacent bars (v1) were less affected than non-adjacent bars (v2) (see Figure 11). In the discussion, we identified emotion's known impact on attentional scope as a possible explanation for this effect. Namely, negative emotion has been shown to narrow attentional scope, while positive emotion broadens it. Since the experiment in chapter 3 did not focus on testing attentional scope from the start, it was necessary to design an experiment that tests adjacent and non-adjacent bars specifically.

In the previous experiment, the adjacent bar condition kept the bars the participants had to judge (target bars) next to each other, and the non-adjacent bar

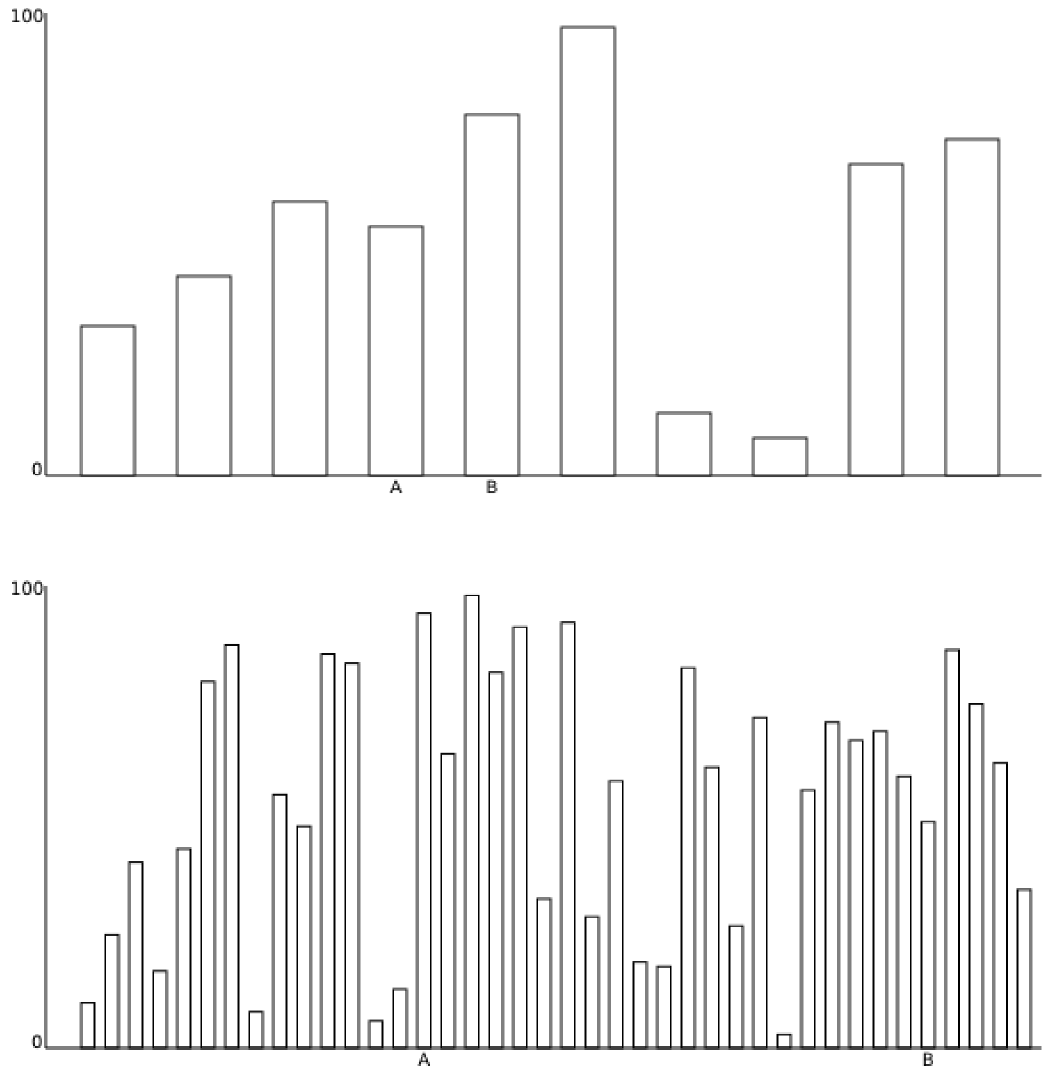

Figure 10: Barchart styles used in these experiments. Top: adjacent condition; Bottom 20-40 non-adjacent condition.

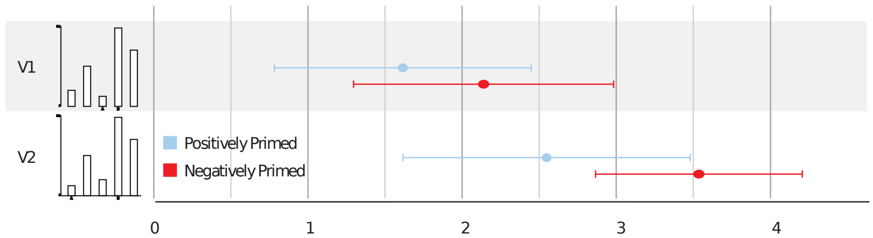

Figure 11: Adjacent bars were seemingly less affected than non-adjacent bars in the previous experiment (chapter 3).

condition put at least one bar in between the target bars (as Cleveland and McGill did in their original experiment). Additionally, all bar conditions in previous experiments have had exactly five bars. Therefore, in order to test conditions with more than three separating bars, the present experiment required an increase in the total number of bars.

To explore the effect of various non-adjacent conditions, we conducted a pilot study. Several non-adjacent conditions were created, which we refer to via two numbers (e.g. 5-10). The first number is the number of bars that separate the target bars, and the second number is the total number of bars in the chart. For example, referring to a "5-10 non-adjacent" condition means that there were 10 total bars in the chart, and that 5 bars separated the target bars (for examples, see Figure 10). In this pilot study, we compared adjacent (0-10) to 1-10, 3-10, 5-10, 10-20, and 20-40 non-adjacent conditions, for a total of 6 conditions. To accommodate a variety of screen sizes, we chose 600 pixels as the maximum chart width.

As expected, we found the largest error difference between the adjacent (0-10) and 20-40 non-adjacent condition. We therefore chose to compare these two in the larger experiments (no-priming and image-priming). Because the 20-40 task is noticeably more difficult than the 0-10 task, we refer to them as having high and low difficulty, respectively.

Increasing task difficulty is a necessary step that will help us better determine how emotion plays a role in everyday visual analysis. Additionally, since more difficult tasks demand more attention, increasing task difficulty also allows us to explore whether the differences we see in negative and positively primed individuals is due to emotion influencing their attentional-scope.

Because the experiments in this chapter require more statistical rigor, we also increase the number of total judgments a participant made during the experiment. In the experiment described in chapter 3, participants made only five visual judgments after the priming task. Five was chosen as a lower limit to help ensure that a) a variety of graphical element proportions were seen and b) successful emotional primes would persist through the visual judgment tasks. Additional pilots with the story priming, however, suggested that we can increase the number of visual judgment tasks to twenty without seeing a significant decrease in prime effectiveness.

4.1.2 Affect Stimuli and Measurement

In the experiments described in chapter 3, stories from the New York Times were used as priming material, in part because they allowed us to hide the nature of the study better than more explicit emotion-inducing forms, such as images or audio. Images have been used as priming material in other HCI studies [66, 77], and so we found them to be a good fit for the present experiment.

Ten images (five negative, five positive) were selected from the International Affective Picture System (IAPS) database [64]. The IAPS provides users with hundreds of images that have been rated by a diverse set of people and verified as inducing the desired emotional state. Because IAPS use requires that their pictures do not appear in publications or be posted online for download, we do not display them here. Using the IAPS, we selected five images that were often rated negatively (e.g. disaster scenes), and five often rated positively (e.g. smiling babies). During the experiment, Turkers were asked to write a description of each image, to ensure that they paid attention to its content. Image tagging tasks such as the ones used in our experiment are very common on Mechanical Turk. This made it possible to avoid fully disclosing the intention of the experiment until the debriefing.

For emotion measurement, we used the SAM scale [14], which was used in chapter 3, as well as other HCI studies [66].

4.2 Experiment: No Priming

To test whether the performing the chart tasks alone (with no priming task) can induce an emotional response, we conducted an experiment using Amazon's Mechanical Turk (AMT) service.

4.2.1 Participants

We recruited 200 participants through AMT; it took approximately one day to gather all responses. The task was advertised as "Answering questions about charts", and did not reveal that we would ask about participants' emotional state. A complete debriefing followed the task. Each participant was paid $0.50 for completing the study, which was calculated by measuring the average time it took participants to complete a pilot and adjusting to meet the US minimum wage. Since we had two chart conditions (adjacent and 20-40 non-adjacent), participants were randomly assigned to a condition when the experiment began.

4.2.2 Procedure

After a participant was assigned to a chart type (adjacent or 20-40 non-adjacent), they were first given the SAM scale as a pre-test to record their valence and arousal. Next, participants were given detailed instructions for the chart task and asked to complete a practice chart task. Participants were instructed to make quick but accurate visual judgments, and to avoid using any outside tools. Upon successful completion of the practice task, participants began the set of 20 chart questions. As in previous studies, the datasets were generated in the same manner as Cleveland and McGill's experiment.

After the visual judgment tasks were completed, participants were given the SAM again as a post-test. Finally, they were invited to submit any additional comments about the task.

4.3 Results

Our study adhered to a between-subjects design, since participants were given a single chart type. As before, we calculate judgment error using the midmeans of log absolute errors (MLAEs). To control for outliers, participants were excluded if their error or average task completion time exceeded three median-absolute deviations from the median[4].

Because we are concerned with whether the chart tasks produce an emotional

[4] This is similar to the three standard-deviations from the mean rule, but more robust to large initial outliers.

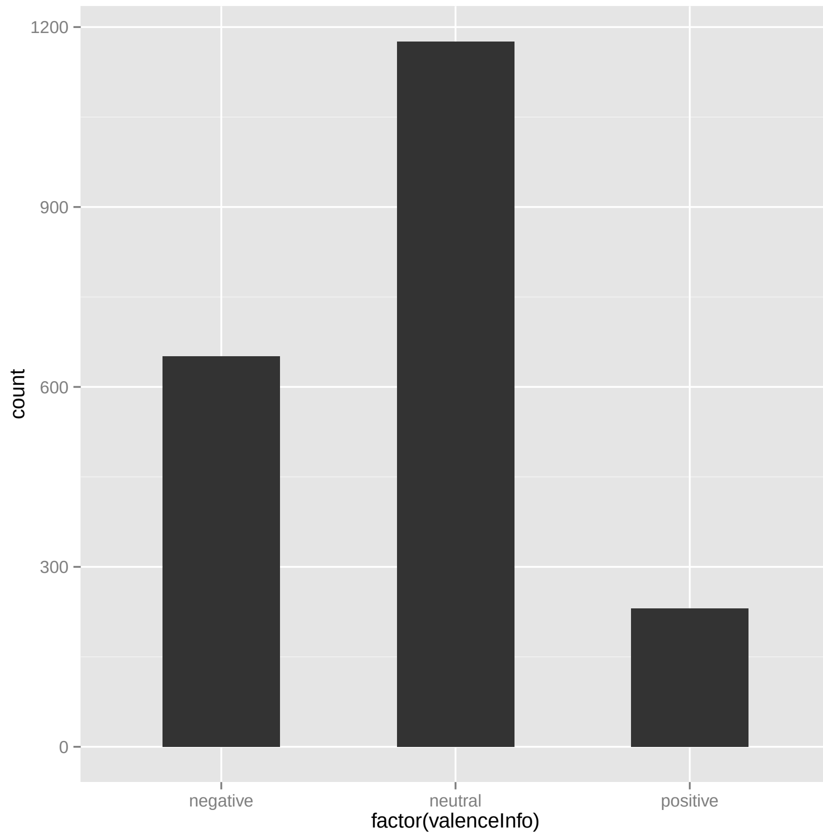

Figure 12: Distribution of emotional change in the control study. No priming was used in this study. Participants indicated their emotional state on a scale, did 20 chart tasks, then indicated their emotional state again. These changes were found to be an indicator of how accurate the participant was during the chart tasks.

change, we define three possible outcomes for the pre-test and post-test for valence. Namely, we define a decrease in valence scores as negative, an increase in valence scores as positive, and equal valence stores as neutral. The resulting distribution of valence groups from this experiment is shown in Figure 12. This distribution suggests that emotion can be induced from completing chart tasks. To investigate this effect further, we turn our attention to the relationship between resulting valence groups and error.

To analyze valence groups and error, we first separate participants by chart. This

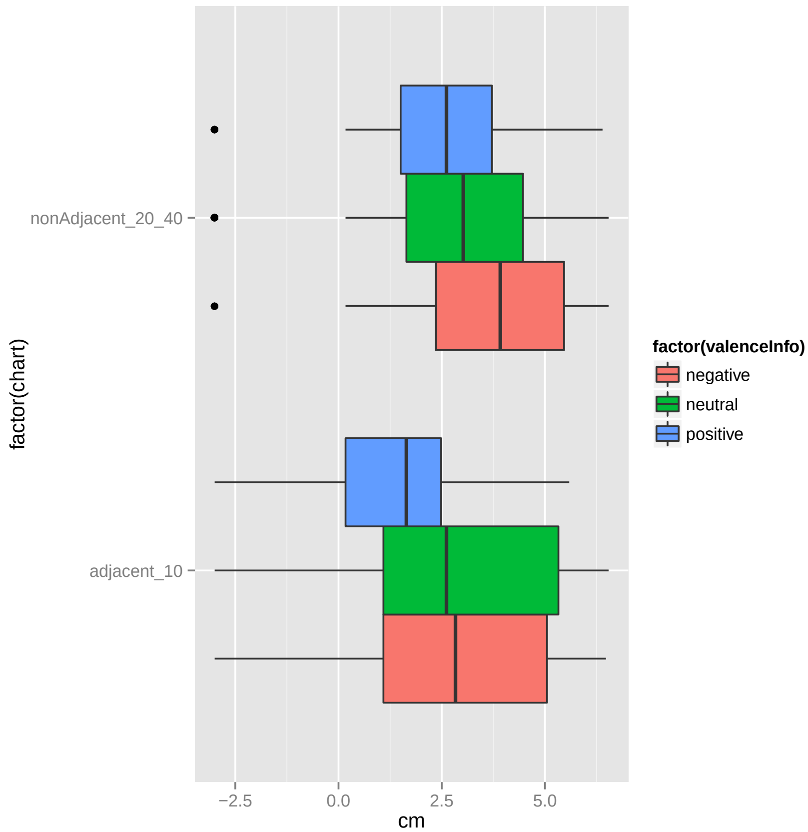

Figure 13: Error for different emotional-change groups in the control study. Participants who reported a positive change in emotion performed better than those who reported a neutral or negative change.

separation is necessary since we know that average error scores between charts is different. Boxplots of errors for the two chart types and valence groups is shown in Figure 13. Interestingly, the positive group performed best, followed by the neutral group and the negative group. An ANOVA yields significant results for both adjacent ($p = 5.87e - 06$) and 20-40 non-adjacent ($p = 1.69e - 12$). Following the ANOVA, we perform a Tukey's HSD for pairwise comparisons[5]. All comparisons between error and valence group were statistically significant, except for neutral and negative in the adjacent group.

[5]Tukey's HSD is conservative with respect to uneven groups.

4.4 Experiment: Image Priming

The purpose of this experiment was to test whether image priming produces the same effect as story priming and whether this effect holds when task difficulty is increased.

4.4.1 Participants

We recruited 450 participants through AMT; it took approximately three days to gather all responses. The task was advertised as "Describing images and answering visual questions", and did not reveal that we would ask about participants' emotional state. A complete debriefing followed the task. Each participant was paid $0.70 for completing the study, which was calculated by measuring the average time it took participants to complete a pilot and adjusting to meet the US minimum wage. Since we had two chart conditions (adjacent and 20-40 non-adjacent), participants were randomly assigned to a condition when the experiment began. Additionally, participants were randomly assigned to a priming group (positive or negative).

4.4.2 Procedure

After a participant was assigned to a chart type (adjacent or 20-40 non-adjacent) and priming group (positive or negative), they were first given the SAM scale as a pre-test to record their valence and arousal.

Next, participants were asked to write a short description of five images. All five images were selected to produce either a strong positive or negative emotional response. After describing the images, participants immediately began the chart task, in order to better preserve any successful prime.

As in the control study, participants then were given detailed instructions for the chart task and asked to complete a practice chart task. After the visual judgment tasks were completed, participants were given the SAM again as a post-test, and invited to submit any additional comments about the task.

4.5 Results

This experiment also adhered to a between-subjects design, since participants were given a single chart type and priming group. As before, to control for outliers, participants were excluded if their error or average task completion time exceeded three median-absolute deviations from the median.

As in the no-priming study, we define three possible outcomes for the pre-test and post-test for valence: positive, negative, and neutral. The resulting distribution of valence groups from this experiment is shown in Figure 14. This distribution suggests that the negative priming is easier to achieve than positive priming, similar to the effect we saw in chapter 3 and in previous psychology research [111].

To analyze priming groups, valence groups, and error, we first separate participants by chart, as we did in the previous experiment. Boxplots of errors for the two chart types, priming groups, and valence groups is shown in Figure 15. As we saw in chapter 3, the positive group tended to perform best, even accounting for the valence group effect we saw in the no-priming experiment. An ANOVA yields significant overall effect for error in both adjacent ($p = 0.003301$) and 20-40 non-adjacent ($p = 4.54e-09$) groups. However, the Tukey's HSD shows that some of the comparisons in

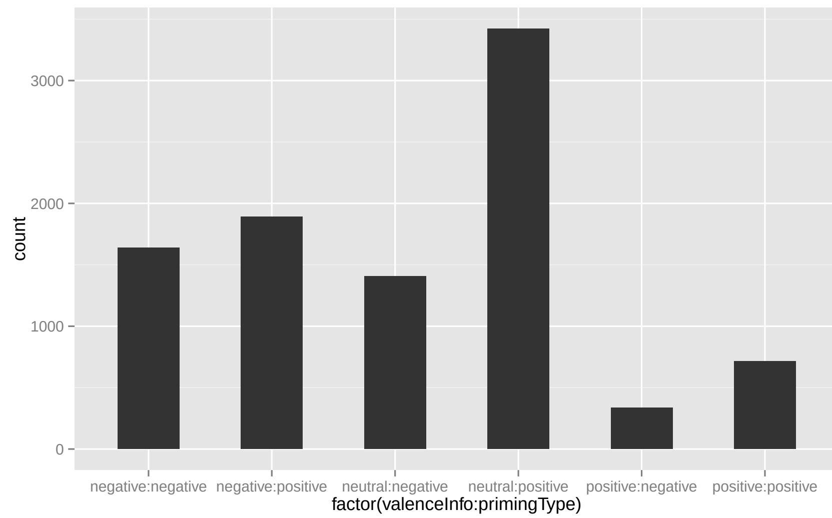

Figure 14: Distribution of emotional change and priming groups in the image-priming study. As expected, there are many more negative primes in the negative group. However, there are very few positive primes in the positive group, and numerous neutral. This indicates a limitation of either the SAM scale or the effectiveness of the priming material.

the adjacent group are not significant, particularly between the positive and negative priming/valence groups ($p = 0.9918565$), yet the differences between these in the 20-40 non-adjacent group is strongly significant ($p = 0.0000001$).

4.6 Discussion and Conclusion

The results of these experiments help us address each of the questions posed at the beginning of this chapter, so we discuss each in turn:

Do the chart tasks themselves induce an emotional change in participants? Yes. The results of the no-priming experiment indicate that participants who perform graphical perception tasks sometimes experience a change in emotional state. Further analysis showed that the direction of the emotional change also relates to performance, with the positive change group performing better than the neutral or negative groups. This implies that future experiments in emotion and visualization should take the

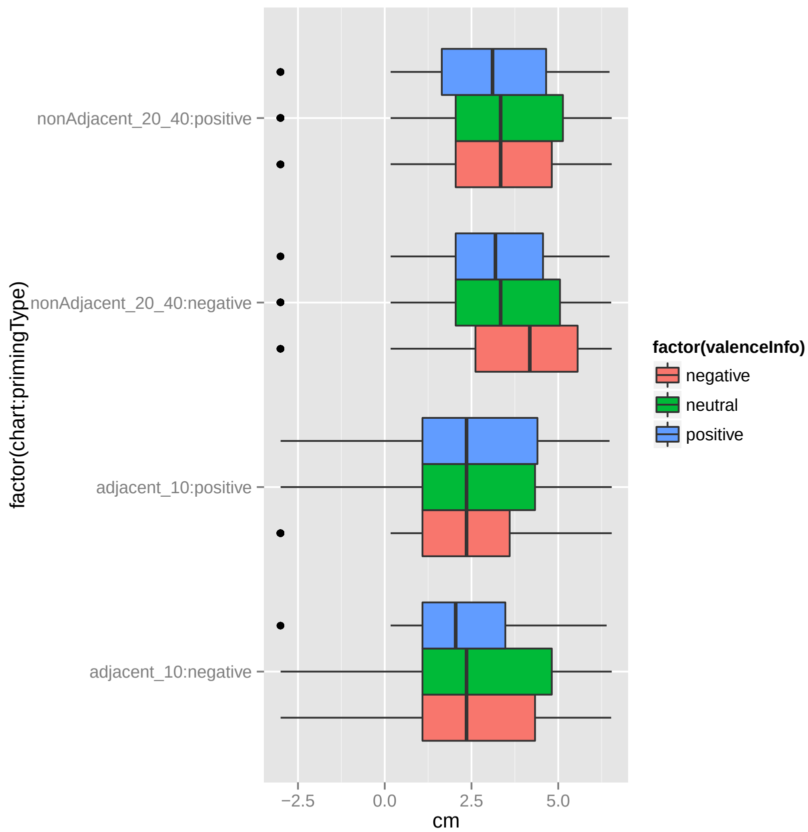

Figure 15: Error for different priming, emotional-change, and chart groups in the image-priming study. The effect of the no-priming study holds, but there is an additional change between the positive and negative groups for the more difficult (twenty separating bars) task.

emotional impact of the visualization task into account, which we do in the image-priming experiment in this chapter.

Could alternate priming techniques be used and do they produce the same effect as stories? Yes. We found that images from the IAPS database [64] were suitable primes for visualization tasks. Not only did they produce a similar performance impact as stories, but they also :w shortened the amount of time it took participants to complete the online experiment.

Do the performance differences from priming hold as task difficulty increases? Yes. The results of the image-priming experiment mirror those of the story-priming experiment of chapter 3 in that the adjacent bar condition was less affected by priming than the non-adjacent bar condition. Increasing task difficulty is an necessary step towards moving from controlled experiment style visualizations (like those of Cleveland and McGill), to visualizations that better reflect real-world visualization use.

The experiments of this chapter answer several questions raised by chapter 3. Charts tasks are shown to impact the reported emotional state of participants, which reflects how accurate they were on the task. Images are found to be as effective as stories for priming in visualization studies, which will allow for more cost-effective studies in the future. Finally, experiment results show that more difficult chart tasks are more influenced by priming, lending support to the hypothesis that positive emotion broadens attentional scope. These lay the groundwork for future experiments that explore the perceptual and cognitive processes involved in low-level visualization.

CHAPTER 5: EMOTION INFLUENCES INTERACTION

While human emotion has been shown to be an important component in decision making [93], there has been little discussion regarding the dynamics of human emotion in visual analytics. Recent work in visual analytics has explored the extent to which information regarding analyst action and reasoning can be inferred from interaction. However, these methods typically rely on humans instead of automatic extraction techniques. We demonstrate that action-level visualization interactions serve as a suitable predictor of user frustration and propose an automatic prediction method that performs similarly across different users and datasets.

To address this, we conduct an experiment to collect data that accurately reflects user emotion transitions and corresponding interaction sequences. This data is then used to build Hidden Markov Models whose capabilities in predicting user frustration from observed interaction sequences are tested accordingly. A discussion on the implications of the model results and additional findings regarding human emotion and frustration follow the experiment results.

5.1 Motivation and Contributions

The key to successful visual analytics lies within the analyst. However, analysts are human, and humans are influenced by their emotions. Thus far, there has been little discussion in the visual analytics community about the role of human emotion in decision making when using visual analytics systems. However, research in psychology indicates that emotions serve as a strong influence, especially in critical analytical situations [93]. Additionally, Pike et al. identified capturing user intent as a core challenge for visual analytics systems [86]. Yi et al. define intent as "what a user wants to achieve" [119]. However, an exploration sequence in which an analyst is frustrated may reflect a different high-level reasoning process than one in which they are engaged or interested. Coincidentally, human intent has been shown to be ambiguous and difficult to discern in the absence of emotional intelligence [85]. Therefore, the ability to detect emotions such as frustration could prove useful to visual analytics systems.

A goal of this experiment is to relate findings from psychology and affective computing regarding emotions and decision-making to visual analytics. Previous work in visual analytics has explored the extent to which high-level actions can be inferred from analyst interaction [93, 17]. Other work investigated the abilities of novice users to infer the reasoning strategies of expert analysts [31], and to assist expert analysts in recalling their own reasoning processes [67]. Similarly, our work seeks to explore the extent to which interaction logs can be used to infer the emotional state of analysts. In contrast to previous work, we propose a method of automatically predicting engagement and frustration without the aid of human coders.

We hypothesize that human-to-visualization interaction patterns can be used to infer and predict user frustration. For example, in human-to-human discourse, people are often able to detect frustration in others by observing the way the others interact with them. Similarly, in human-to-visualization discourse, we wish to see if a system can make similar inferences. Complicating the problem is the fact that the system can only detect a limited number of the user's actions, mainly through the keyboard, mouse, and loggable actions in the interface. While other input devices can be used, it is not practical to assume that such devices are available. Our work demonstrates that by tracking basic visualization and interface interactions, a statistical model can be built that effectively infers when a user is frustrated.

To accomplish this, we collect ground-truth data that contains visualization interaction logs and emotional state transition timelines. Then, we train Hidden Markov Models using this data and evaluate their predictive capabilities using evaluation methods from the machine learning community. We argue that predictive analytics applied to user interaction may serve as a first step towards visual analytics systems that can react intelligently to analyst's actions.

The contributions of this experiment to the visual analytics community can be summarized as follows:

- Demonstrating that interaction can be used to predict frustration in visualization-specific interaction patterns.
- An automatic method for inferring and predicting frustration using visualization interaction logs.
- Discussing the implications of emotion (and its impact on cognition) for visual analytics.

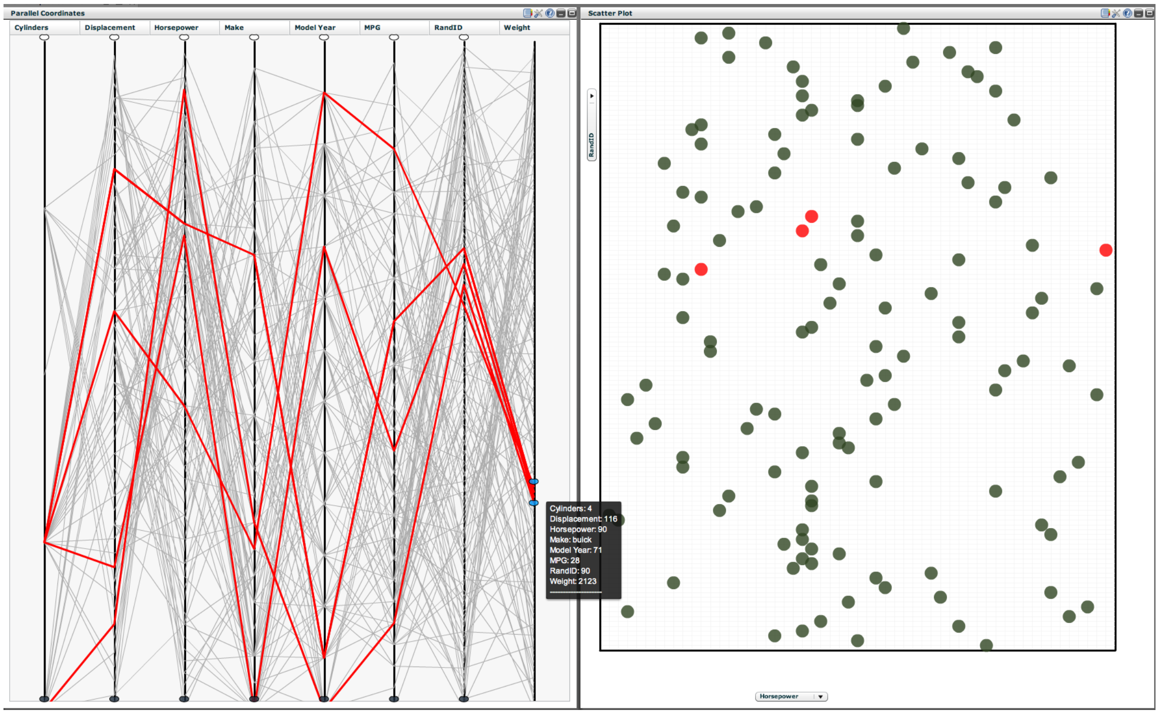

Figure 16: An overview of the visualization system participants used

5.2 Experiment

To build a predictive model of frustration, we first gathered ground-truth data that accurately reflects the emotional state transitions and interaction sequences of humans using a visualization system. Participants were asked to complete three to four tasks using either the cereal or cars datasets that are widely used in visualization research [120, 5].

The Self-Assessment Manikin test, discussed in the background section, was given as a pre- and post- test to gauge the participant's emotional state. Each participant was randomly assigned to complete either the cereal or cars dataset tasks first. The tasks were given in increasing difficulty, and consisted of objectives such as identifying outliers or determining if a correlation exists between variables. More details on the

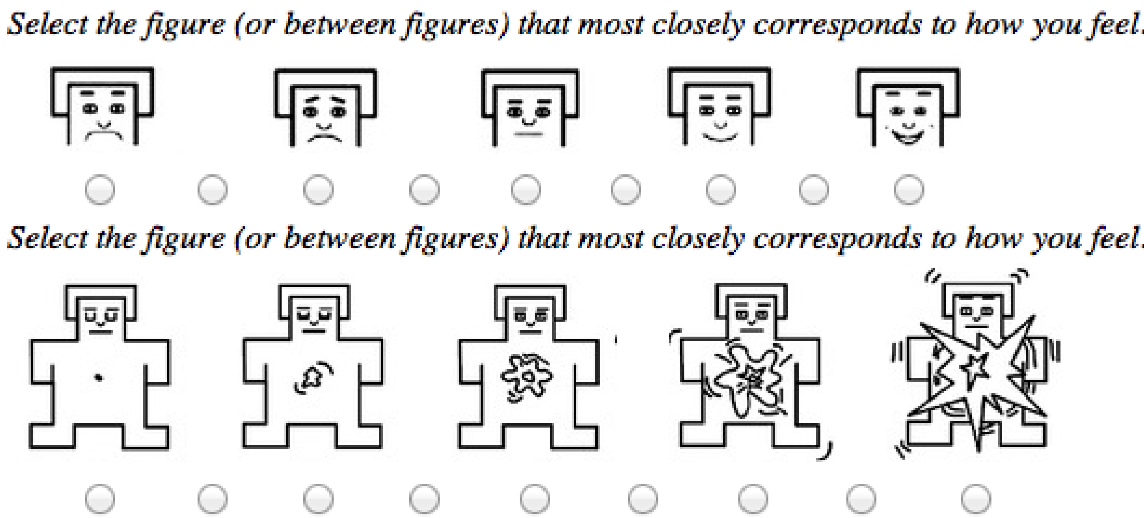

Figure 17: The Self-Assessment Manikin test allows users to report their valence (top) and arousal (bottom). [14]

tasks are given in 5.3.1.

While the participants were completing the tasks, several interaction events are logged. These can be found in table 1.

Video recording was used to capture the participant's face and screen to assist in building an emotion transition timeline after all tasks are completed. Using these emotion-transition timelines and interaction logs, we build a predictive model of user frustration. The results of this model and the Self-Assessment Manikin test will be discussed in the results section.

5.2.1 Hypotheses

For our experiment, we identified the following hypotheses to test:

- Frustration influences a user's interaction patterns with a visualization system.
- A user's affective state can be predicted from their interaction patterns.
- Within the experiment, interaction patterns that reflect emotional states are

generalizable across users.

- Within the experiment, interaction patterns that reflect emotional states are generalizable across the cereal and cars datasets.

5.2.2 Participants

We recruited 11 participants (4 female, 7 male) from visualization courses at our university. Participants age ranged from 20 to 35+ (as indicated on a scale). Participants were primarily computer science graduate students and most had taken at least one visualization course.

5.3 Procedure

The primary purpose of this experiment was to collect ground-truth data that can be used to build a predictive model. As such, the experiment can be divided into two portions: that of task-completion to collect interaction data and that of video mark-up to generate emotion transition timelines.

After participants completed consent and demographics forms, they were given an overview of the visualization system (see Figure 17) during which all of the possible interactions were explained. Additionally, participants were reminded about concepts such as correlations and outliers to better prepare them for the tasks.

After the overview, the participants were given the task sheets to look over and instructed to ask any questions they might have. Next, the participants were given the Self-Assessment Manikin (SAM) test. Participants were asked to use the scales in the SAM to report their perceived emotional state prior to performing the upcoming tasks. The SAM was also given as a post test.

Table 1: The ten interaction events logged by our system.

PC-Tooltip	Mouse-over to display of detailed information in parallel coordinates
SC-Tooltip	Mouse-over to display of detailed information in scatterplot
SC-Click	Highlighting an item in the scatterplot
PC-Click	Highlighting an item in parallel coordinates
Change SC Axes	Selecting new scatterplot axes
Change PC Axes	Rearranging axes in parallel coordinates
PC minbar moved	Moving an upper parallel coordinates slider
PC maxbar moved	Moving an bottom parallel coordinates slider
Opening WorkSpace	Opening a new display
Closing WorkSpace	Closing a display

To facilitate the participant's focus on the questions, they were not asked to report their emotional state during the tasks. When the participant finished the assigned tasks, the investigator displayed the recording of the participants face and screen. Then the investigator and the participant reviewed the videos simultaneously and identified points at which the participant transitioned from engaged or interested to frustrated or confused and vice versa. This process allowed us to build a timeline with two main threads: the interaction sequences and corresponding emotional states.

5.3.1 Tasks

The cereal and cars datasets, both commonly used in the visualization community, were used in this experiment. Furthermore, the tasks were designed to be presented in increasing difficulty and vagueness, so as to increase the probability of frustration. All tasks were designed to be completed in 5-15 minutes time. However, no time limit was enforced. The tasks are as follows:

Cars:

- In what year was the car with the highest MPG made?
- Is there a correlation between horsepower and displacement?
- What is the lightest car with the highest horsepower?

Cereal:

- What is the hot cereal with the highest fiber content?
- Is there a correlation between fat and calories?
- What manufacturer produces cereal with the highest average vitamin content?
- What manufacturer is most consistent in its nutritional content of its cereals?

5.3.2 Hidden Markov Models

Hidden Markov Models (HMMs) are used to statistically determine the probability that a system is in a given state based only on observation sequences [91]. In the case of our experiment, the user's affective state is hidden and only the interactions with the visualization system are observable.

Given training data containing an observation sequence O of length N ($O_0 = 0, O_1 = 1, \ldots, O_N = N$), and a state sequence S of length N ($S_0 = 0, S_1 = 1, \ldots, S_N = N$), we can calculate transition probabilities *tran* and emission probabilities *emis* that represent the probability of transitioning from state-to-state and of seeing a particular item of the observation possibilities for each state, respectively.

Probabilities are also produced for starting states based on their frequency of occurrence.

After these are calculated, when we encounter data for which only the observations are available and the states are hidden, we can use *tran* and *emis* along with the Viterbi algorithm [91], to calculate the most probable state sequence for the observed interaction sequence.

Given such a model, we can calculate predictive analytics metrics such as false-positive, true-negative, sensitivity, recall, and other metrics. The resulting metrics of our analysis and corresponding terminology definitions are given in the results section.

5.4 Results

The user study provided us with interaction logs for which each interaction event has a corresponding affective state. Using part of this data for training and part for testing, we trained Hidden Markov Models and tested their predictive capabilities on data in which we remove the participants' reported affective states. This was repeated to test each one of our hypotheses.

Training produces a HMM with probabilities for states given observations (interactions), for transitions between states, and for the initial state. The results are reported in the form of predictive analytics metrics [1]. One user's data was removed from the study due to a miscommunication regarding experiment instructions.

The average number of interactions logged per participant was 547. The average amount of interactions during which users reported being frustrated and engaged was 268 and 309, respectively. No correlation was found between number of transitions

Table 2: Results of leave-one-out cross validation across users

Metric	Cars	Cereal
Accuracy	67.71%	67.56%
Error	32.29%	32.44%
Sensitivity	61.28%	56.89%
Specificity	75.23%	80.44%
Avg true-positive	189.1	161.8
Avg true-negative	198.9	189.6
Avg false-positive	65.5	46.1
Avg false-negative	119.5	122.6

and accuracy.

Sensitivity is defined as the number of positive states correctly identified as such. For the purpose of this study, sensitivity denotes the classification rate of correctly identifying a user as engaged or interested. Similarly, specificity is the number of negative states correctly identified. In this study, specificity is equivalent to the model's rate of correctly identifying frustration.

5.4.1 Model accuracy across users

To evaluate the HMM across users, we used leave-one-out cross fold validation[1]. This evaluation method trains the HMM using $K - 1$ of the participants' data and tests the predictive capability of the HMM on the remaining data. This is repeated K times, once for each participant. The entire procedure is repeated for each dataset.

The results indicate that the interactions logged (see table 1) can be used to infer frustration or engagement across datasets with nearly 70% accuracy. Frustration was more accurately inferred than engagement, with engagement being classified as low as 57% and frustration as high as 80%. Refer to table 2 for additional results.

Table 3: Results of cross validation across datasets

Metric	Cars	Cereal
Accuracy	67.28%	67.47%
Error	32.72%	32.53%
Sensitivity	60.82%	56.79%
Specificity	74.81%	80.36%
Avg true-positive	187.7	161.5
Avg true-negative	197.8	189.4
Avg false-positive	66.6	46.3
Avg false-negative	120.9	122.9

5.4.2 Model accuracy across different datasets

To evaluate the HMM across datasets, we trained the HMM using the data from all users for the first dataset and evaluate the resulting HMM's predictive capability on each user's data from the second dataset in turn. This procedure is repeated after reversing the order of the datasets.

For predicting across datasets, the HMM performs at 67% for both the car and cereal datasets. Similar to the across-users evaluation, frustration was more accurately determined than engagement. Table 3 shows additional results.

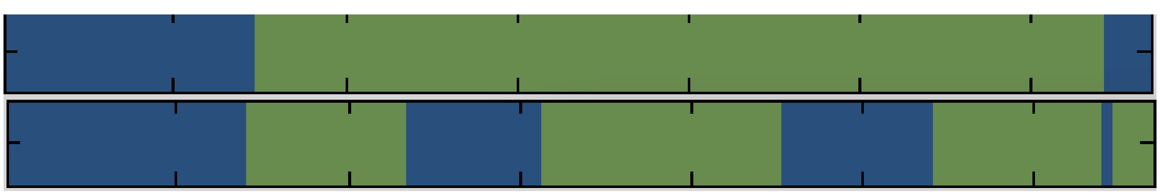

Figure 18: (Top) the predicted state transitions for a user and (bottom) the corresponding actual state transitions (green is frustrated, blue is engaged). The accuracy of these predictions was 70%

5.5 Discussion

We begin our discussion by addressing key questions based on our hypotheses:

Does frustration influence a user's interaction patterns with a visualization system?

Yes. Given the performance metrics in tables 2 and 3, we see that frustration is statistically differentiable and predictable. To illustrate this, see Figure 19 that visualizes a user's entire interaction sequence along with their corresponding reported states.

Can a user's affective state be predicted from their interaction patterns?

Yes, but only about 70% of the time. The results indicate that, given an interaction sequence, we can predict with whether a user is frustrated or engaged. Interestingly, it was easier for the model to predict frustration than engagement, as shown by the specificity and sensitivity results. Figure 18 compares the actual reported states to the best-fit estimate the model produced. Note that the model tends to miss short bursts of frustration, but seems to represent a smoothing of the actual states.

Are interaction patterns that reflect emotional states generalizable across users?

Yes. As it turns out, by leaving a user's data out of the training set and then using only their data for testing, the model tends to perform well. Note that, in figures 22 and 23, there are a few users for which the model does not perform well. However, we did not find any indication that the number of transitions influenced accuracy. Future work will explore additional reasons for why the model performed well on some user/dataset combinations and poorly on others.

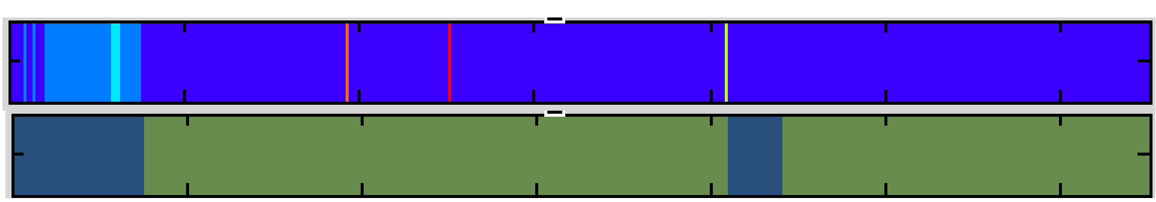

Figure 19: (Top) visualization of entire interaction sequence for a user and (bottom) their corresponding reported states

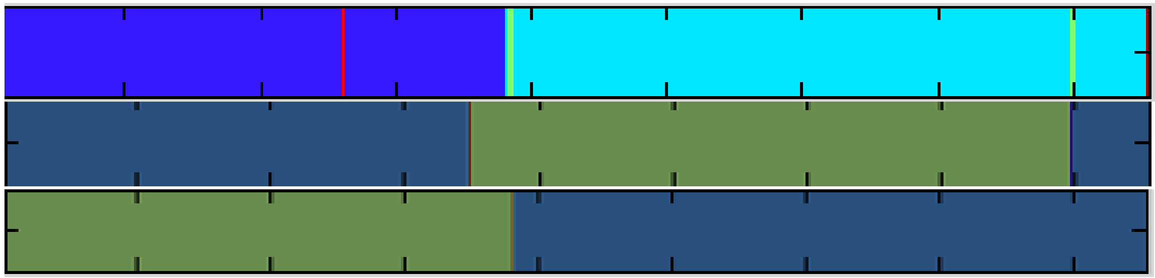

Figure 20: The (top) interaction sequence (middle) reported states and (bottom) estimated states for the worst performer (11% accuracy).

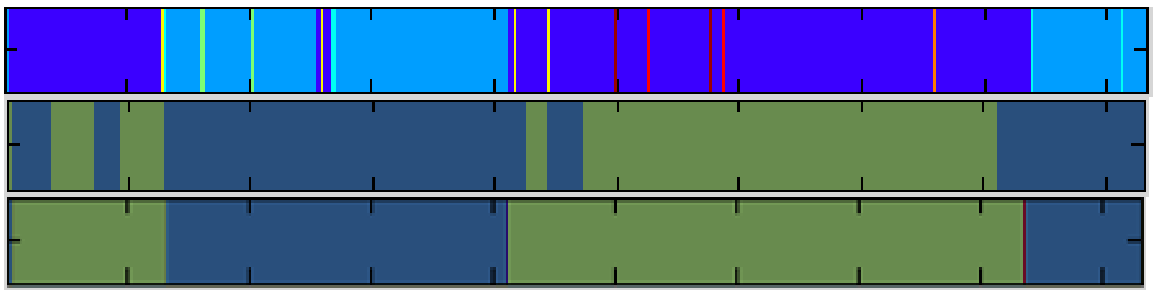

Figure 21: The (top) interaction sequence (middle) reported states and (bottom) estimated states one of the better performers (86% accuracy).

Are interaction patterns that reflect emotional states generalizable across datasets?

Yes. The results for predicting frustration across users and across datasets are similar. A possible reason for this is that the tasks for each dataset were designed to be similar. The first question asks the user to identify an extreme value, one asks them to identify a possible correlation, and the other asks them to find an outlier. The cereal dataset included one extra question to facilitate extra interaction. However, this had little influence on the accuracy and average number of interactions.

In figures 20 and 21, we see the interaction sequences, reported states, and predicted states of the worst and one of the most accurately predicted users. Both models were trained on the cars data and tested on the cereal data. Interestingly, there seems to be a discrepancy between a key interaction sequence change in the worst performer

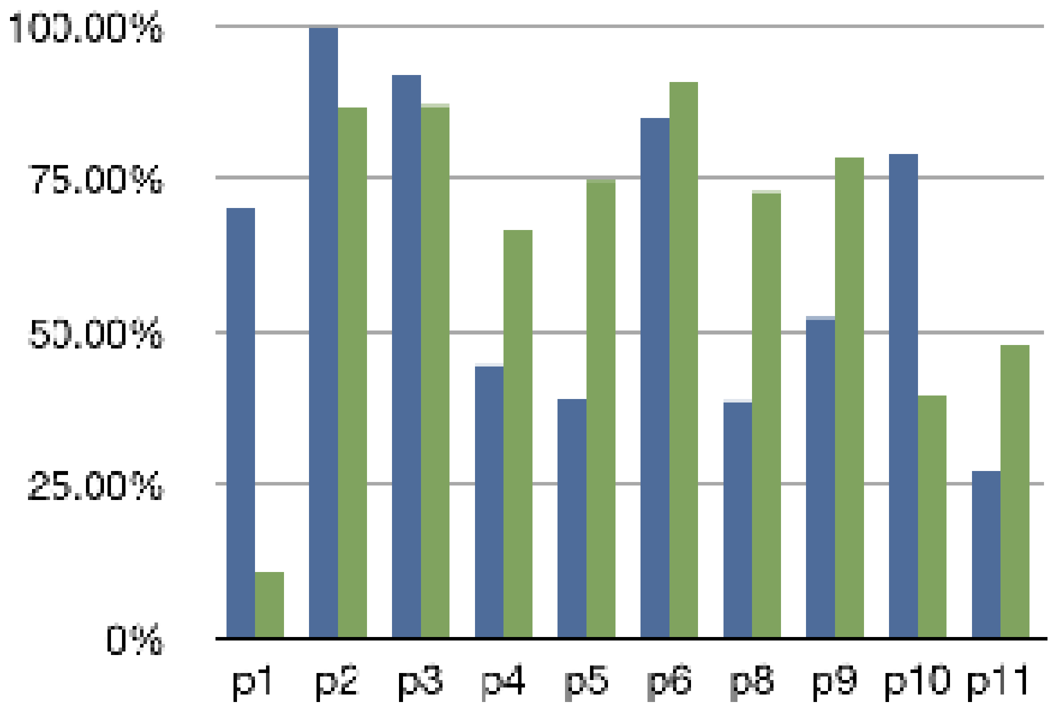

Figure 22: Accuracies for each participant across data (blue is cars, green is cereal)

and the point at which they become frustrated. In contrast, the better performer seems to have indicated several transitions in their timeline. One insight we can draw from this is that self-reporting is still not perfectly reliable. Also, since the model would have done better to reverse all predictions for the worst performer, this may indicate that an option to build personalized models would prove useful if emotional intelligence is ever integrated into visual analytics systems.

To explore this further, if we train using the second-worst performer's and test on the worst performer's data and vice versa, we get improvements of 11% to 89% and 48% to 53% accuracy, respectively. This finding raises the question that there could be clusters of users who exhibit frustration in interfaces in different ways.

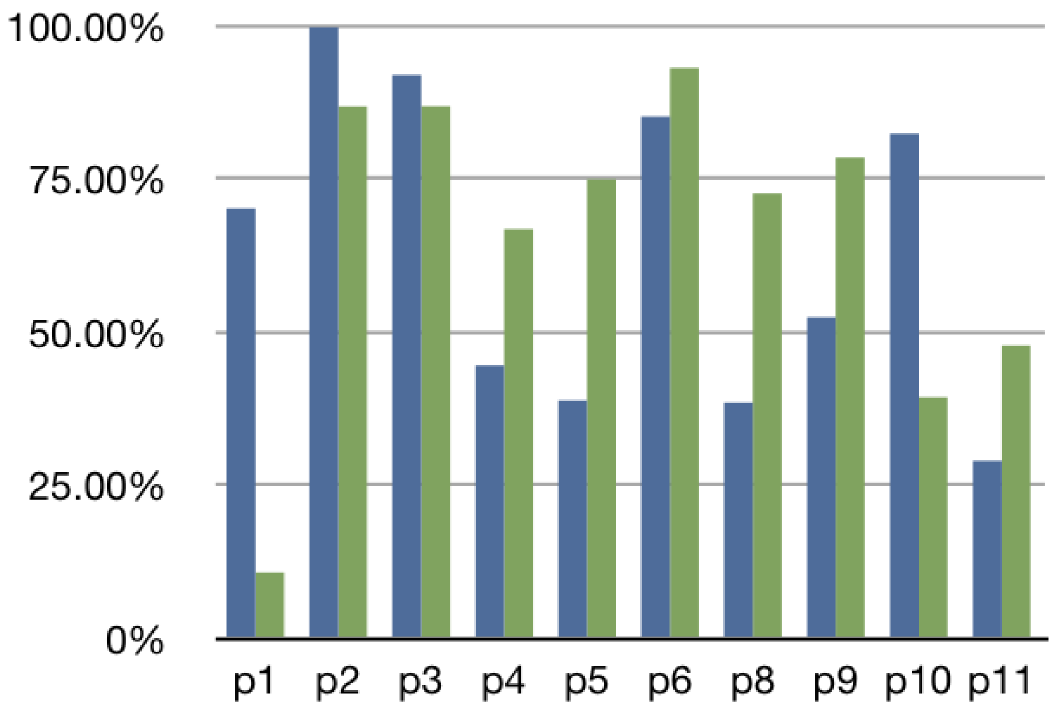

Figure 23: Accuracies for each participant across users (blue is cars, green is cereal)

Some users indicated extra information when asked to report frustration. A few participants indicated that the visualization display was confusing at times, that the interaction was unresponsive (some had trouble highlighting specific lines in parallel coordinates), or that they did not fully understand the tasks. Whether it is the data, the interaction methods, or the visualization that causes the emotional change is unknown and can be examined in future work. However, since the method of discourse from a human to a visualization system is through interaction, it is possible that no matter what the cause of the emotional change, it will be reflected in the user's interaction patterns.

We did not use the temporal relationships between events or information regarding what data is selected, but such information could help improve the accuracy of the models. While many more interaction events could be used, zooming or filter events, for instance, we are still able to achieve reasonable frustration detection results. Other possible indicators of frustration in interaction logs will be explored in future work. We expect that the use of more interaction possibilities and metrics will prove useful in building more robust predictive systems.

5.6 Experiment Summary

We show that interaction sequences can be used to predict frustration in visualization systems. To achieve this, we conducted an experiment in which participants performed basic visualization tasks using linked parallel coordinates and scatterplot displays while their interactions are logged and their face and screen are captured in video files. After participants completed the tasks, they viewed the aforementioned videos along with an investigator and created an emotion transition timeline. Then the interaction logs and emotion timelines were used in training and evaluating Hidden Markov Models. Predictive analytics metrics that show model performance across users and datasets are reported along with other results relating frustration and interaction.

CHAPTER 6: VISUALIZATION, EMOTION, AND COGNITIVE MODELS

The effects of individual differences on user interaction is a topic that has been explored for the last 25 years in HCI. Recently, the importance of this subject has been carried into the field of information visualization and consequently, there has been a wide range of research conducted in this area. However, there has been no consensus on which evaluation methods best answer the unique needs of information visualization. We propose the ICD^3 Model (Individual Cognitive Differences), whereby individual differences are evaluated in 3 dimensions: cognitive traits, cognitive states and experience/bias. Our proposed model systematically evaluates the effects of users' individual differences on information visualization and visual analytics, thereby responding to Yi's [118] call for "creating a standardized measurement tool for individual differences".

6.1 Introduction

In recent years, strides have been made toward understanding the impact of individual differences on performance when interacting with visual analytics systems. Research has shown that factors such as personality [42, 121], spatial ability [20], biases [76, 123, 124] and emotional state [4, 38, 59, 82, 96, 101] impact a user's performance. Though progress is undeniable, a common limitation is that every cognitive factor that affects visualization performance is not considered or properly controlled.

For instance, studies that focus on personality factors alone do not consider how differences in working memory, perceptual ability, and previous experience can also affect visualization performance. Indeed, as stated by Yi in his position statement in 2010, the visualization community has yet to employ a comprehensive and standardized model for measuring individual differences such that researchers can better understand how factors in individual differences interact with each other and with existing evaluation techniques [118].

We acknowledge that this work cannot solve all of the problems described above. However we propose a first-step towards a solution by introducing the ICD^3 Model (Individual Cognitive Differences) - a 3-dimensional model that encompasses the cognitive facets of individual indifferences. A necessary step in attempting to define a model of individual cognitive differences was to seek an underlying structure of previous research by identifying which factors are dependent and which are independent of one another. By surveying the existing literature, we propose that individual differences can be categorized into three orthogonal dimensions: cognitive traits, cognitive states and experience/bias.

Cognitive traits are user characteristics that remain constant during interaction with a visual analytic system. Factors such as personality, spatial visualization ability, and perceptual speed are all examples of cognitive traits. These have been shown to correlate with a user's ability to interact with a visualization [21, 27, 42, 110, 121] and can be generalized to predict the behavioral patterns of users with different cognitive profiles.

Cognitive states, on the other hand, are the aspects of the user that may change dur-

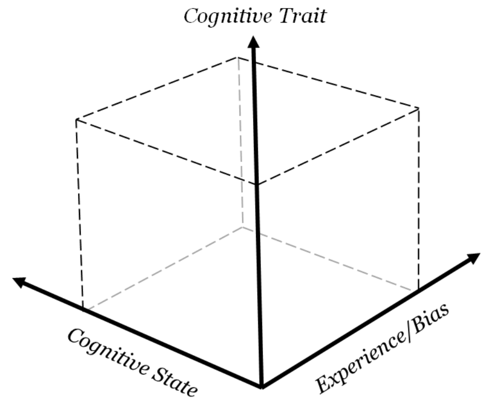

Figure 24: The ICD^3 categorizes individual cognitive differences in three orthogonal dimensions: Cognitive Traits, Cognitive States, and Experience/Bias

ing interaction and include situational and emotional states, among others. Research has shown that a user's performance can be altered by changes in their emotional state [4, 38, 59, 82, 96, 101], and the importance of combining workload with performance metrics has been noted for decades [51, 81, 117]. Although cognitive states are difficult to measure because of their volatility, they provide important contextual information about the factors affecting user performance that cannot be described through cognitive traits alone.

Cognitive states and traits can describe a significant portion of a user's cognitive process but they are not comprehensive; experience and biases can also affect cognition. Intuitively, we think of experience and bias separately, but they both describe learned experiences that can affect behaviour when familiar problems arise, and are therefore not orthogonal. Although there has been little work about the impact of experience/bias on interaction with visual analytics systems, previous studies have shown that learned behavior such as confirmation bias can affect performance and decision-making [49].

Taken together, these three dimensions can create a model that encapsulates the cognitive aspects of individual differences (Fig. 24). Similar to how analyzing state and trait alone would disregard potential performance gains from expertise, ignoring any one dimension of the model would also result in an incomplete description of performance. For example, analyzing only expertise and traits ignores changes that may be triggered by workload or emotions (cognitive state). Thus, the model is only complete if all three dimensions are considered. By using ICD^3, evaluators can identify what factors must be controlled in an experiment and which should be included as independent variables. The community can also begin to evaluate visualizations using this common platform and be able to better reproduce and extend each other's research.

6.2 Background

Evaluation has been an active area of research in visualization in recent years. Several researchers have worked toward developing standard evaluation methodologies and secondary measures to evaluate perception and comprehension [60, 78, 99, 100]. In this section, we focus on those that directly report or suggest the use of brain imaging or individual differences for evaluating visualizations.

Anderson et al. [2] demonstrated the use of EEG to measure the user's cognitive load when viewing different boxplot designs. In a position statement presented at BELIV 2010, Riche [94] proposed the use of multiple physiological measurements (heart rate, eye gaze, brain imaging, etc.) for evaluating visualizations. At the same BELIV workshop, Yi [118] proposed studying individual differences when evaluating visualizations. Yi argued that understanding how users differ in personality and cognitive factors is important in evaluating visualizations. In a follow-up research project, he demonstrated that there is a significant difference between novice and expert users when using a visualization to solve analytical tasks and pinpoints the importance of additional research in individual differences in visualization evaluation [61].

The emergence of this body of research ultimately highlights the need for better evaluation methods that address the unique needs of visualizations, but there is no consensus on which methods address these needs. What is clear, however, is that the field of visualization has yet to identify a systematic and objective way of measuring individual differences in user analysis of visualizations. We seek to address this by organizing the existing research into a cohesive structure: the ICD^3 model.

6.3 Dimensions of Individual Differences

In this section we discuss each of the three dimensions in the ICD^3 model: cognitive traits, cognitive states, and experience/bias. Figure 24 shows the three dimensions and how they could be represented graphically. Specifically, we illustrate how the components of these dimensions affect performance, and tie these to related experiments in visualization.

6.3.1 Cognitive Trait

Cognitive traits such as spatial ability, verbal ability and working memory capacity vary considerably among individuals and have been demonstrated to affect perception, learning and reasoning. Consequently, it has been shown that cognitive factors can affect a user's performance when using a visualization. We propose using these factors to measure the stable traits that make up a user's basic cognitive profile.

Several studies have demonstrated the effect of basic cognitive abilities on user performance in visualization tasks. For example, Conati and McLaren [27] found that perceptual speed, the speed at which users compare two figures, correlates with accuracy on an information retrieval task. Another commonly studied cognitive factor that has been shown to impact interaction in a visualization is spatial ability, and refers to the ability to reproduce and manipulate spatial configurations in working memory. Chen and Czerwinski [21] found correlations between spatial ability and the visual search strategies users employed in a network visualization. Participants with high spatial ability performed better in search tasks and navigated an interactive node-link visualization of a citation network more efficiently. Velez et al. [110] tested the correlation of speed and accuracy with a number of factors related to spatial ability, including spatial orientation, spatial visualization, visual memory and perceptual speed. These factors affected users' speed and accuracy in the comprehension of three-dimensional visualizations, similar to those found in scientific visualization applications. Similarly, Cohen and Hegarty [26] found that a user's spatial ability affects the degree to which interacting with an animated visualization helps when performing a mental rotation task.

An interesting aspect of these findings is that an individual's spatial ability not only affected performance, but also how they approached tasks. If people with varying cognitive abilities employ different strategies, an evaluation methodology will need to take these strategies into account to fully understand user behavior.

Perceptual and spatial abilities are not the only cognitive factors that have been shown to have an effect. Yi [118] proposed that one must investigate beyond a users' basic spatial ability to better understand the variability in visualization evaluation. Many personality factors relevant to visualization use are both reliably measurable and consistent over a user's lifetime, making them potential candidates for understanding a user's stable traits. For example, the Five Factor Model, a common model in personality psychology, categorizes personality traits on five dimensions: extraversion, neuroticism, openness to experience, conscientiousness and agreeableness. Green and Fisher [42] studied how varying scores on the Five Factor Model as well as locus of control impacts the way users interact with visualizations. Locus of control [95] is the degree to which a person feels in control of (internal locus of control) or controlled by (external locus of control) external events. The authors found individuals with an external locus of control performed better at complex inferential tasks when they used a visual analytics system than when they used a web-based interface with a list-like view. The study also revealed a correlation between neuroticism and task performance. Ziemkiewicz et al. [121] found that users with a more internal locus of control showed greater difficulty adapting to visual layouts with a strong metaphor of containment (i.e. a layout with many containers) versus a more traditional list-like menu.

The results of these studies suggest that cognitive traits may account for some of the observed individual variability in visualization use. Understanding this variability will help to improve the generalizability of evaluation findings. Therefore, it seems prudent to include this in a model of individual differences in user research.

6.3.2 Cognitive State

Cognitive state refers to the current condition of a person's mental processes. Unlike cognitive traits, cognitive state can change from moment to moment during interaction with a visualization, impacting performance, understanding, or retention.

Cognitive load is the most studied cognitive state in visualization evaluation, as it often has a direct impact on performance. In particular, working memory has been labelled as an information bottleneck in visualization because it is limited by both size and duration [70, 75, 87]. When multiple visual elements compete for space, there is a loss of information and often a decrease in performance. Speed and accuracy regularly suggest mismatches between visual design and perceptual affordances [15], and dual-task studies can be designed to evaluate mental demand through performance [73, 80].

Cognitive load theory breaks down this generic concept of workload into three more narrowly-defined categories [18]. Germane load describes the memory needed to the process and understand new schemas, intrinsic load refers to the amount of memory necessary for a given task (and cannot be modified by instructional design), and extraneous load is determined by the memory needed to absorb information and can be modified based on presentation. This last category is what researchers typically refer to when comparing the workload demands between visualizations.

Unfortunately, an increased load on working memory is not always reflected in behavioral metrics [117], and it is possible for one person to exert significantly more mental effort than another to achieve the same level of performance in a visualization [115]. Accordingly, researchers have suggested the integration of performance and mental demand during evaluation [51, 81, 117]. Paas and Merrienboer constructed a two-dimensional model of performance and mental effort to define cognitive efficiency [81], and Huang et al. tailored the model to visualization evaluation by adding a third dimension - response time [51]. However, this extra exertion is not necessarily an indication of poor design. Hullman et al. proposed that “visual difficulties” may introduce beneficial obstructions that aid information processing and engagement [52].

Moving away from cognitive load, emotional states triggered by visual imagery or from other external sources can also impact interaction with a visualization. Bateman et al. suggested that emotional responses to “chart-junk” may have favorably impacted the recall of information [6]. Previous work has shown positive emotional states to enhance attention regulation, working-memory performance, open-ended reasoning, creativity, and “big picture” understanding [4, 38, 59, 82, 96]. Conversely, negative emotional states, such as anxiety, can disrupt visuospatial working memory [101]. Finally, emotions have a strong link to decision-making and cost-benefit analysis [7]. Observing these subtle (or not so subtle) nudges to performance is necessary to fully describe the interaction between a person and a visualization.

These studies represent a small subset of work from the psychology literature that has addressed cognitive state and performance. For example, cognitive load is an umbrella term that needs to be narrowed in order to be predictive of performance (for example, visuospatial working memory v. verbal working memory). Additionally, the effect of emotional state on visualization performance has been largely unexplored. Considering these factors will help construct more accurate models of performance in visual analytic systems.

6.3.3 Experience and Bias

Whereas cognitive state refers to current mental processes, and cognitive trait to stable aspects such as personality, neither of these capture how experience and bias can affect visualization performance. Here we cover a sample of the extensive work on the effects of experience and bias on cognitive performance from the fields of psychology and decision science. We then relate them to recent work in visualization that has begun to explore the role of experience and bias in visualization.

Both experience and biases form through previous interactions with a given problem, and are often utilized when a similar problem is encountered later on [28, 106]. Although experience and bias could be discussed at length separately, here we discuss them together, since they are not orthogonal to each other [29]. For instance, while extensive experience assists with avoiding biases common to novices, experience has also been shown to introduce biases that novices do not encounter, such as the failure to appropriately weight information that contradicts previous findings [57].

Experience is associated with the formation of effective reasoning strategies for given problem types [36, 98], many of which are applicable to reasoning with visualizations. An experiment from Cox et al. [28] explored the relationship between experience and performance on a hypothetico-deductive task, and found that participants who had experience with similar problems were able to utilize previously formed reasoning strategies on the new task.

Such tasks parallel the hypothesis testing techniques described in Pirolli and Card's sensemaking model [88], which has been utilized widely in the design of visual analytics systems.

While the effect of experience on cognitive processes has been studied extensively, there is relatively little work in the visualization community which has explicitly examined and discussed how differences in experience affects performance in interactive visualizations [24]. Perhaps the first to address experience directly is Kwon et al. [24], who identify common roadblocks novice analysts face when using a complex visual analysis system. Other visualization work has explored experience's effects on visualization somewhat indirectly. For instance, Dou et al. [31] explore the how well novice users were able to infer the reasoning processes of expert analysts based on a visualization of the experts' interaction logs. Arias-Hernandez et al. describe pair analytics [3], an analysis process which pairs one analyst with visualization experience with another who has experience in the data domain.

Bias refers to a predisposition to behave a certain way for a given task [92, 106]. Similar to experience, there is little work in the visualization community that discusses the relationship between visual analysis and different types of cognitive biases. Notably, Zuk and Carpendale [124] discuss bias and uncertainty in depth, focusing on the many ways in which bias can affect reasoning with uncertain data and how visualization may aid users in debiasing. Another example of debiasing comes from Miller et al. [76], who describe an experiment in which a system consisting of a statistical model and corresponding visualization was used to assist users in avoiding regression bias. Their results showed that the visualization approach outperformed both no-visualization and algorithmic approaches, supporting the notion that visualization and interaction help users manage biases effectively. Ziemkiewicz and Kosara [123] found that visualizations can be subject to perceptual biases, which can adversely influence how users recall spatial relationships between visual elements. Many other types of cognitive biases exist which impact reasoning and task performance [29, 49], yet the relationships between these and visualization is largely unexplored.

The experiments described here underscore the argument that experience and bias can significantly influence visualization performance. Since cognitive states and traits also affect performance, however, it is imperative that we explore the relationship between these three dimensions.

6.4 Proposing the ICD^3 Model

In light of the three dimensions that we have discussed, we believe that a structured model would be useful in order to describe individual cognitive differences when users interact with visualizations.

We therefore propose a three dimensional model that is composed of cognitive traits, cognitive states, and experience/bias - the ICD^3 model. In this model, each orthogonal dimension would represent an individual difference of a user thereby allowing researchers to describe or perhaps even predict a user's ability to interact with a visualization, by knowing where that individual lies along the three different axes. This would allow for not just isolated cognitive factors, but also for the *interaction* of the user's different cognitive abilities.

Figure 25 gives a hypothetical example of a user looking at percentage judgments in treemaps. The cognitive state in this example is the user's workload, their cognitive

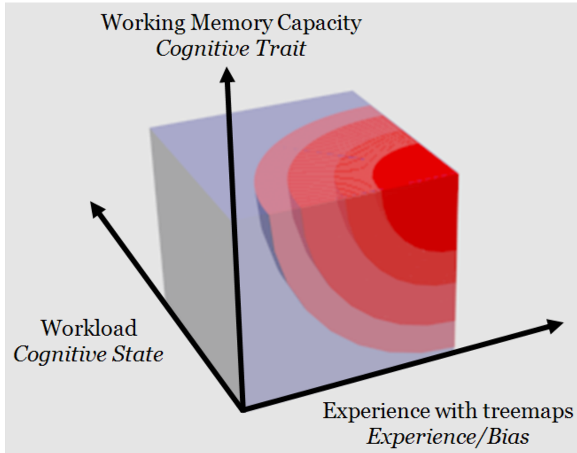

Figure 25: An example of how an ICD^3 model might be constructed. We map the interaction of workload, working memory capacity, and experience on performance of percentage judgments in treemaps. Darker red represents better performance.

trait is their working memory capacity, and their experience/bias is how experienced they are with treemaps. An ICD^3 model would show that if the user is an expert, has a low workload, and has a high working memory capacity, then they have higher performance and abilities with percentage judgment in treemaps. Conversely, if the user is overloaded with work, has a naturally low working memory capacity, and has no experience of treemaps, then they will be less effective in performing that task.

After defining the visualization, task, and cognitive factors, a set of experiments can then be run in which participant workload, working memory capacity, and experience is varied. For each interaction of factors, performance is recorded in the instance at the appropriate coordinates. Given enough data, we construct a descriptive topology of performance for a task and visualization.

Unfortunately, the interaction of cognitive facets is ordinarily much more nuanced than depicted in Figure 25. For the sake of simplicity, we chose working memory capacity, workload, and experience because their impact on performance is relatively straightforward. But in practice, we have little knowledge of how other combinations of state, trait, and experience/bias influence interaction with a visualization.

For example, some studies have shown that extraverts and introverts perform differently when they receive positive or negative feedback about a task, thus modifying their cognitive state [11]. Introverts tend to perform well when given positive feedback and worse when given negative feedback. Reciprocally, extraverts perform worse than intraverts given positive feedback, but their performance improves under negative feedback. This exemplifies why it is important to consider the interaction of state and trait.

However, other studies have suggested that people with an external locus of control (LOC), which is correlated with extraversion [79], perform better in visualizations where they have had no previous experience than people with an internal LOC [121]. This demonstrates how trait and experience can interact to influence performance.

Each of these examples provide a two dimensional snapshot of how cognitive dimensions can impact performance. But how do we combine the knowledge of these two studies? How would performance be impacted when an experienced intravert is given negative feedback, or an inexperienced extravert is given positive reinforcement during a task? Thus, a key attribute of the ICD^3 model is that limiting the scope of evaluation to any two of the three described dimensions leaves an incomplete and potentially misleading model of performance:

- Analyzing state and trait without experience ignores performance gains by expertise

- Analyzing state and experience without trait ignores interaction differences that are driven by personality or inherent cognitive strengths (e.g. spatial ability)
- Evaluating experience and trait without state disregards the moment-to-moment cognitive changes in the user that could be driven by emotion or workload

While instances of the ICD^3 model should be constructed for a explicit task and visualization, we imagine that the interaction of certain cognitive factors will be generalizable across visual forms (and tasks). In the next section, we explore the implications of the ICD^3 for design.

6.5 Implications for Individual Differences Models

One important advantage of understanding individual users' cognitive states, traits, and biases as a cohesive structure is that this opens up the possibility of developing adaptive, mixed-initiative visualization systems [104]. As noted by Thomas and Cook in *Illuminating the Path* [104], an important direction in advancing visual analytics research is the development of an automated, computational system that can assist a user in performing analytical tasks. However, with few exceptions, most visualization systems today are designed in a "one-size-fits-all" fashion without the ability to adapt to different users' analytical needs into the design.

There is mounting evidence that successful adaptive systems can significantly improve a user's ability in performing complex tasks. In the recent work by Solovey et al. [102], the authors show that with the use of a brain imaging technology (fNIRS) to detect a user's cognitive states, the system can adapt the amount of automation and notably improve the user's ability in performing a complicated robot navigation task. Ziemkiewicz et al. [121] demonstrate that the impact of locus of control (LOC) on visualization can be significant. When the user is given a hierarchical visualization that correlates with the user's LOC, a user's performance can be improved by up to 52% in task completion time, and 28% in accuracy.

It is clear that adaptive systems offer new possibilities for visualization research and development, but more work is necessary to model *how* and *when* a system should adapt to a user's needs. As noted earlier, only emphasizing one or two of the three proposed dimensions can lead to a system incorrectly assessing the user's analysis process and provide the wrong adaption. By examining all three dimensions in a cohesive fashion, it becomes possible for a system to predict a user's performance and realize the potentials of an adaptive, mixed-initiative system as proposed by Thomas and Cook.

6.6 Developing a Model of Individual Differences

Creating a precise model of individual differences is a daunting task. From the outside, it can appear that even the slightest deviations between people can influence performance in a visualization, whether it is as obvious as taking a formal course in visualization or as subtle as reading emotionally-charged news articles between analysis tasks. Cognitive states may interact with and manipulate each other - for example, emotional state has been shown to impact working memory - and people simultaneously bring many traits and experiences to the table each time they see a visualization. Furthermore, there are almost certainly cognitive traits, states, and experiences that impact interaction more than others.

While we don't believe that these problems impact the orthogonality of our proposed model, it illuminates the potential dependency of factors within each dimension, increasing the difficulty of predicting human interaction. We highlight at least two future areas of research that will be critical to addressing these challenges.

First, discovering new and unobtrusive methods to capture cognitive state, trait, and experience/bias will ultimately drive research in individual cognitive differences. For example, recent advances in non-intrusive physiological sensors that detect emotional states, such as the Affectiva Q-Sensor [89], will enable future studies into the impact of emotional state and visualization performance. In real-world scenarios, it is unrealistic to expect users to be subjected to numerous forms and intrusive monitoring equipment. The simple act of filling out personality surveys or applying brain sensing equipment is enough to potentially modify cognitive state (or introduce biases) before interaction. It should be a central goal to record as many cognitive factors as possible, in as little time as possible, with as little disruption as possible.

Second, finding dominant individual cognitive factors both within dimensions and between dimensions should limit the sheer volume of cognitive tests necessary to describe interaction. For example, if participants have a low working memory capacity, their locus of control might not matter given a certain task and a visualization. If this is true, then having a participant fill out a survey to determine locus of control may be unnecessary. Similarly, we suspect that a person's experiences and biases may impact performance more than many other cognitive traits and states. Thus, if we know a person is an expert at a simple task, emotional state might be irrelevant. Identifying these dominant factors should reduce the number of interactions between cognitive factors.

The generalizability of cognitive states, cognitive traits, experiences/biases on performance in visualization has yet to be seen. As a result, the ICD^3 model takes a conservative approach by specifying an exact set of cognitive factors and requiring tests to be performed on a fixed task and fixed visualization. By identifying important factors or important interactions between factors, we hope to construct new metrics in the future that are more predictive of interaction with a visualization.

6.7 Summary

We have made initial steps towards a model that captures the various cognitive aspects that affect visualization performance by dividing them into three dimensions: cognitive states, cognitive traits, and experience/bias. Furthermore, we have discussed how each of these dimensions are orthogonal to each other, meaning that during visualization interaction, a user may exhibit different values for states, traits, or experience/biases. Each of the dimensions are found to influence cognitive processes related to visualization, such as reasoning and perception. The ICD^3 model provides a sample space for experiments involving visualizations, so that we may form a better understanding of the cognitive underpinnings of visualization.

CHAPTER 7: TOWARDS A MODEL OF EMOTION IN VISUALIZATION

In this chapter, I discuss how emotion relates to existing cognitive and perceptual models for visualization, and propose a preliminary model for emotion in visualization.

To date, emotion has not been explicitly included in any cognitive models for visualization. This is not surprising, since much work in the psychology of emotion is relatively new, largely spurred by Antonio Damasio's influential book *Descartes Error* [30], which was published in 1994. Additionally, emotion is only a part of the emerging research area of individual cognitive differences in visualization, which we discussed in chapter 6. Even so, relating the results of this dissertation to well-established cognitive models in visualization is a necessary step towards establishing the role of emotion in visualization.

In the following section, I discuss how emotion relates to some of the popular cognitive and perceptual models in visualization, including van Wijk's visualization model [109], Green's human cognition model [44], Pirolli's sensemaking model [88]. Then, leveraging these existing models, I propose a preliminary model for emotion in visualization, that will serve as a basis for future experiments examining the relationship between emotion and visualization.

7.1 Existing Models in Visualization

One of the most recognized models for visualization is van Wijk's model, which focuses on the relationship between the knowledge the visualization can convey to its user. Van Wijk describes the act of the user interacting with the visualization as follows (emphasis mine):

> The amount of knowledge gained depends on the image, the current knowledge of the user, and the particular properties of the *perception and cognition* P of the user.

One result of this dissertation is that we now know that emotion relates directly to the P (perception and cognition of the user during visualization use) in van Wijk's model. This relationship is illustrated in Figure 26. Future experiments, however, should investigate the role of attentional scope in visualization further (which Healey recently called for in [46]), as well as the role of emotion, visualization, and other low-level components of perception and cognition, such as working memory.

Van Wijk's model also accounts for a user leveraging their own knowledge to guide exploration. Recent work in visualization has shown that cognitive traits such as locus of control can impact how a user interacts with a visualization [43, 121]. Additionally, the experiment described in chapter 5 showed that emotional states such as frustration and engagement can be inferred from interaction. It seems, then, that van Wijk's model could be modified to promote cognitive states and traits as having a more significant role in how a user interacts with a visualization. In other words, explicit knowledge is not the only factor that determines how a user interacts with a

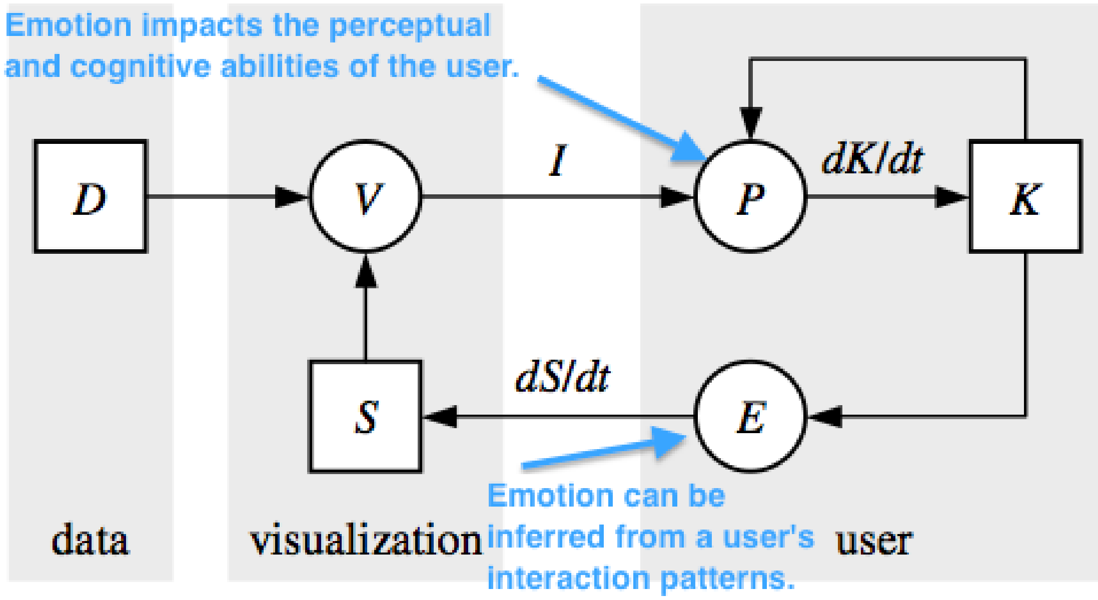

Figure 26: Relating emotion to van Wijk's visualization model.

visualization, feelings, intuition, and preference also guide interaction.

In contrast to van Wijk's model, the human cognition model (HCM) proposed by Green et al. [44] focuses on leveraging existing knowledge and models of higher-level cognition to create a model of human cognition for interactive visual analytics tools. Their view of interaction expresses similar concerns as to how van Wijk's model displays knowledge as the primary driver of interaction. For example, when advocating for multiple views visualizations, they state:

> Humans themselves do not interact with information in one dimension; humans are capable of multi-layered processing: perceptual, emotional, and higher-cognitive.

The HCM authors also cite Wilson and Schooler's study on introspection and decision quality, which refers to previous studies involving emotion to help explain their results how introspection impacts decision quality [116].

Since the HCM takes a broad view of cognition, it leaves room for the role of emotion. The HCM also includes existing work from cognitive psychology that takes emotion into account. Therefore, future work in visualization would benefit from exploring how emotion in visualization contexts can impact higher level cognitive factors, such as decision making and creativity.

Finally, we turn our attention to Pirolli's sensemaking model [88]. Throughout this influential work, the authors continually emphasize that attention and working memory are essential to effective hypothesis generation and testing:

> Time pressures and data overload work against the individual analysts ability to rigorously follow effective methods for generating, managing, and evaluating hypotheses.

It follows that impediments to working memory and attention, such as high cognitive load or negative emotion [22], should be considered when designing visualizations.

7.2 A Model for Emotion in Visualization

In the previous section, I briefly reviewed three well-known cognitive/perceptual models for visualization, and discussed how emotion relates to each. In this section, I propose a preliminary model for emotion in visualization, leveraging components of previous models for visualization and existing studies involving emotion.

Effective visualization use requires both cognitive (high level) and perceptual (low level) processes. At the same time, many cognitive and perceptual processes have been shown to be influenced by emotion. Therefore, the components of a model for

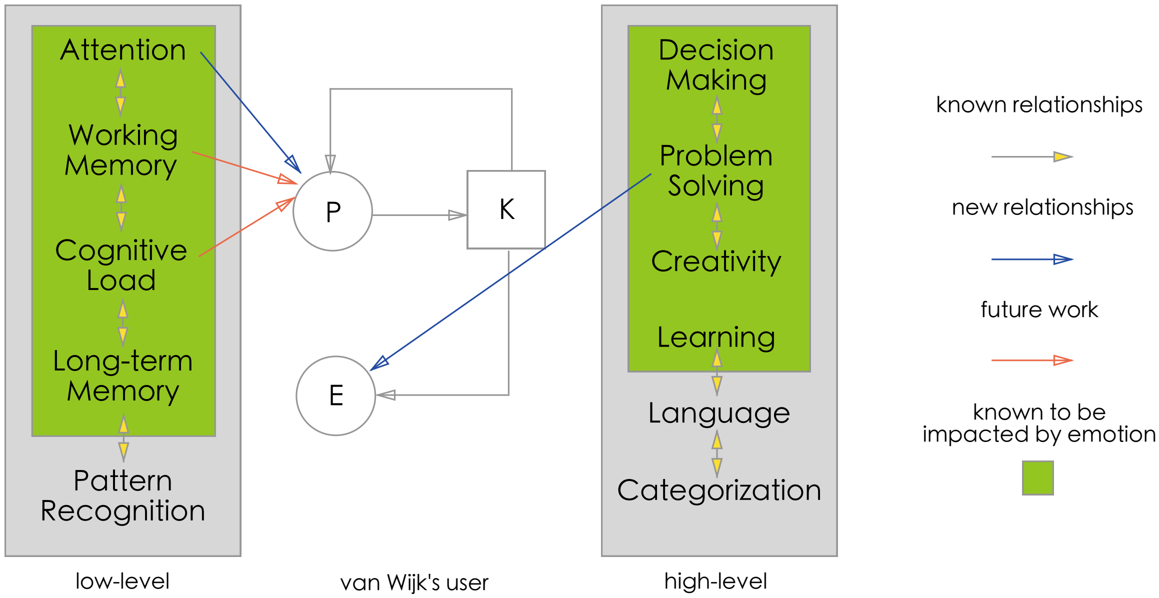

Figure 27: Initial model of emotion in visualization.

emotion in visualization should include elements of cognition and perception that have been a) explored in the visualization space and b) have empirical evidence as being influenced by emotion.

High level processes used in visualization, that also have been shown to be impacted by emotion, include decision-making [44], problem-solving [24], creativity [32], and learning [19]. Similarly, low level processes involved in visualization include attention, working memory, and long-term memory [46, 6]. These high level and low level components form the right and left of the model in Figure 27, respectively. User-focused visualization processes, borrowed from van Wijk's model [109], form the central portion of the model. Finally, several relationships are depicted between elements in the model: previously established relationships, relationships supported by the results of this dissertation, and relationships slated for future work.

Attention was identified as one of the primary factors involved in the effect seen in chapters 3 and 4: that positive emotion led to better performance on visual judgment tasks. Therefore, the model shows a link between attention and visualization perception, illustrating that we now know that emotion plays a role in this relationship.

Links representing future work are drawn between working memory and perception, as well as cognitive load and perception. This is a natural step from the experiments described in chapters 3 and 4, since those experiments primarily involved the attentional capacities of the user. At the same time, it is important to note that attention, working memory, and cognitive load are closely related [101, 46, 2]. Therefore, future experiments should focus on the relationship between these factors, and how emotion might play a role in modulating them.

The experiment in chapter 5 showed that changes in user interaction patterns during a problem solving task were related to periods of frustration or engagement. To show that emotion also plays a role here, a link is drawn between problem solving and interaction. Although recent work in visualization has begun to explore the cognitive factors behind interaction [62] and the role of decision-making in visualization [58], there is still much work in the problem solving and decision making space (e.g. the experiments in [116, 107, 92]), that could be adapted to better understand how users interact with a visualization.

This model is intended to serve as a roadmap, in part by collecting and enumerating the cognitive and perceptual factors often referenced in visualization research, but also by illustrating several remaining gaps in our knowledge of how we perceive and interact with visualizations.

CHAPTER 8: CONCLUSION

At the onset of this dissertation, the role of emotion in visualization was unknown. While psychologists have established that emotion is an essential element of human cognition and perception, a popular notion persists among scientists that emotion is an impediment to rational thinking. Yet a steady flow of experiments from psychology and cognitive science shows that this position is incorrect. On the contrary, emotion is an inseparable component of reasoning and decision making, and even plays a role in shaping how we perceive visual information.

In recent years, this revised perspective of emotion and cognition/perception has emerged in human-computer interaction research, leading to several useful and interesting findings. For instance, a person's immediate (less than 500 milliseconds) emotional reaction to a webpage is an accurate predictor of whether they will engage with a website, and how they perceive a website's usability. Another experiment demonstrated that simply adding a positively valenced image to an interface led users to produce more creative solutions to an open-ended problem. In light of these findings, some scientists are even revisiting older research and interface designs to retroactively evaluate the influence of emotion (Don Norman's influential book *Emotional Design*, is a prime example of this).

Although progress has been made in reconciling modern emotion research with human-computer interaction, there are aspects of information visualization that make understanding emotion even more important. Consider the role of visual perception in visualization. In visualization, accurate visual perception is essential. Therefore, findings in cognitive psychology that show how emotion influences how well we perceive visual information become directly relevant. Visualization systems also often include a) complex datasets and b) interactive processes designed to support reasoning, which make findings about the how emotion influences risk perception and problem-solving approaches applicable.

Taking these observations about emotion and visualization together, I have taken initial steps towards establishing the role of emotion in visualization:

To answer the question of whether emotion impacts low-level perception in a visualization setting, I replicated a classic visualization experiment, adapting it to include emotion priming techniques from psychology research. These experiments showed that emotion influences even the most basic chart tasks (chapters 3 and 4). In particular, participants exposed to positive stimuli performed better on visual comparison tasks across several charts and difficulty levels.

To explore the role of emotion in visualization interaction, I designed the experiment in chapter 5. During this experiment, participants completed several problem-solving tasks while the system logged their interactions. Afterwards, participants constructed a timeline of how frustrated/engaged they felt at each point during the tasks. Using this information, Hidden Markov Models were trained which correctly inferred a participant's emotional state from their interaction sequence 70% of the time.

Given this new information about how cognitive states like emotion can impact visualization use, the next chapter situates emotion within the existing body of individual differences research in visualization (chapter 6). In particular, a three dimensional model for individual differences was proposed, consisting of cognitive states, cognitive traits, and experience/bias. How this model can be applied to visualization research, and implications for evaluation are explored.

Finally, I propose a preliminary model of the role of emotion in visualization (chapter 7). To form this model, several existing cognitive/perceptual models in visualization were evaluated, with the best fit (van Wijk's model) forming the baseline. Next, the results from my experiments and relevant psychology/cognitive science literature to form the full model. This model shows several areas of cognition and perception that are used in visualization, and which of those areas are known to be influenced by emotion. In doing so, this model provides a roadmap for future experiments exploring the role of emotion in visualization.

While emotion can play a significant role in many areas where visualization is used, the area of health risk perception and decision making is one of the most promising for an immediate application of the results of this dissertation. Every day, thousands of patients receive positive test results for cancer, but a positive test result in many cases does not mean that the patient actually has cancer. Facilitating effective reasoning for Bayesian reasoning problems such as this is an active topic in the visualization community, yet accuracy rates remain low [74]. One proposed reason for this challenge is the emotional response to reasoning about cancer, which has been confirmed by presenting the same numerical problem using a different wording [108]. The results in this dissertation show that emotion impacts how we perceive visual data by influencing attention and working memory. Therefore, if we apply this line of thinking to the cancer test results problem, we will be able to uncover some of the underlying challenges patients face in Bayesian reasoning tasks.

Emotion is a pervasive element of the human experience. As data becomes more commonplace in our everyday activities, understanding how emotion shapes our perception of data becomes increasingly important. Though this dissertation has taken some initial steps in this area, our understanding of the nuances of human cognition and perception as they relate to data analysis are incomplete and must be explored further to better understand how to design effective visualizations for people. I hope the experiments and findings presented here serve as a suitable start towards this end.

REFERENCES

[1] E. AlPaydin. *Introduction to Machine Learning.* MIT Press, 2010.

[2] E. Anderson, K. Potter, L. Matzen, J. Shepherd, G. Preston, and C. Silva. A user study of visualization effectiveness using eeg and cognitive load. In *Computer Graphics Forum*, volume 30, pages 791–800. Wiley Online Library, 2011.

[3] R. Arias-Hernandez, L. T. Kaastra, T. Green, and B. Fisher. Pair analytics: Capturing reasoning processes in collaborative visual analytics. *System Sciences (HICSS)*, pages 1–10, 2011.

[4] F. G. Ashby, V. Valentin, and U. Turken. The effects of positive affect and arousal on working memory and executive attention. In S. Moore and M. Oaksford, editors, *Emotional Cognition: From Brain to Behavior*, pages 245–287. Amsterdam: John Benjamins, 2002.

[5] K. Bache and M. Lichman. UCI machine learning repository, 2013.

[6] S. Bateman, R. L. Mandryk, C. Gutwin, A. Genest, D. Mcdine, and C. Brooks. Useful junk? The effects of visual embellishment on comprehension and memorability of charts. *Proc. CHI '10*, pages 2573–2582, 2010.

[7] A. Bechara. The role of emotion in decision-making: Evidence from neurological patients with orbitofrontal damage. *Brain and Cognition*, 55(1):30–40, 2004.

[8] A. Bechara, H. Damasio, and A. R. Damasio. Emotion, decision making and the orbitofrontal cortex. *Cerebral Cortex*, 10(3):295–307, 2000.

[9] M. S. Bernstein, G. Little, R. C. Miller, B. Hartmann, M. S. Ackerman, D. R. Karger, D. Crowell, and K. Panovich. Soylent: A Word Processor with a Crowd Inside. *ACM CHI*, pages 313–322, 2010.

[10] J. Bertin. *Semiologie Graphique.* Gauthier-Villars, 1967.

[11] J. Boddy, A. Carver, and K. Rowley. Effects of positive and negative verbal reinforcement on performance as a function of extraversion-introversion: Some tests of Gray's theory. *Personality and Individual Differences*, 7(1):81–88, 1986.

[12] G. Bohm and H.-R. Pfister. Anticipated and experienced emotions in environmental risk perception. *Judgement and Decision Making*, 3:73–86, 2008.

[13] K. E. Boyer, E. Y. Ha, M. D. Wallis, R. Phillips, M. A. Vouk, and J. C. Lester. Discovering tutorial dialogue strategies with hidden markov models. In *Proceeding of the 2009 conference on Artificial Intelligence in Education: Building*

Learning Systems that Care: From Knowledge Representation to Affective Modelling, pages 141–148, Amsterdam, The Netherlands, The Netherlands, 2009. IOS Press.

[14] M. M. Bradley and P. J. Lang. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1):49 – 59, 1994.

[15] C. Casner and J. Larkin. Cognitive efficiency considerations for good graphic design. Technical report, 1989.

[16] C. F. Chabris and S. M. Kosslyn. How do the cerebral hemispheres contribute to encoding spatial relations? *Current Directions in Psychological Science*, 7(1):8–14, 1998.

[17] S. Chaiken. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, 39(5):752 – 766, 1980.

[18] P. Chandler and J. Sweller. Cognitive Load Theory and the Format of Instruction. *Cognition and Instruction*, 8(4):293–332, Dec. 1991.

[19] R. Chang, C. Ziemkiewicz, R. Pyzh, J. Kielman, and W. Ribarsky. Learning-based evaluation of visual analytic systems. In *BELIV '10*. ACM, 2010.

[20] C. Chen. Individual differences in a spatial-semantic virtual environment. *Journal of the American Society for Information Science*, 51(6):529–542, 2000.

[21] C. Chen and M. Czerwinski. Spatial ability and visual navigation: An empirical study. *New Review of Hypermedia and Multimedia*, 3(1):67–89, 1997.

[22] L. Chepenik, L. Cornew, and M. Farah. The influence of sad mood on cognition. *Emotion*, 7(4):802–811, 2007.

[23] S. Chong and A. Treisman. Attentional spread in the statistical processing of visual displays. *Attention, Perception, & Psychophysics*, 67(1):1–13, 2005.

[24] B. Chul Kwon, B. Fisher, and J. Yi. Visual analytic roadblocks for novice investigators. In *IEEE Visual Analytics Science and Technology*, pages 3–11. IEEE, 2011.

[25] W. S. Cleveland and R. McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387):531–554, 1984.

[26] C. Cohen and M. Hegarty. Individual differences in use of external visualisations to perform an internal visualisation task. *Applied Cognitive Psychology*, 21(6):701–711, 2007.

[27] C. Conati and H. Maclaren. Exploring the role of individual differences in information visualization. *Advanced Visual Interfaces (AVI)*, pages 199–206, 2008.

[28] J. R. Cox and R. A. Griggs. The effects of experience on performance in Wason's selection task. *Memory & Cognition*, 10(5):496–502, Sept. 1982.

[29] P. Croskerry. Achieving quality in clinical decision making: Cognitive strategies and detection of bias. *Academic Emergency Medicine*, 9(11):1184–1204, Nov. 2002.

[30] A. Damasio. *Descartes Error*. New York: Grosset/Putnam, 1994.

[31] W. Dou, D. H. Jeong, F. Stukes, W. Ribarsky, H. Lipford, and R. Chang. Recovering reasoning processes from user interactions. *Computer Graphics and Applications, IEEE*, 29(3):52 –61, 2009.

[32] W. Dou, C. Ziemkiewicz, L. Harrison, D. Jeong, R. Ryan, W. Ribarsky, X. Wang, and R. Chang. Comparing different levels of interaction constraints for deriving visual problem isomorphs. *IEEE VAST*, pages 194–202, 2010.

[33] L. Elting, C. Martin, S. Cantor, and E. Rubenstein. Influence of data display formats on physician investigators' decisions to stop clinical trials: prospective trial with repeated measures. *British Medical Journal*, 318(7197):1527–1531, 1999.

[34] C. Eriksen and J. St. James. Visual attention within and around the field of focal attention: A zoom lens model. *Attention, Perception, & Psychophysics*, 40(4):225–240, 1986.

[35] M. Eysenck and M. Calvo. Anxiety and performance: The processing efficiency theory. *Cognition & Emotion*, 6(6):409–434, 1992.

[36] C. W. Fisher, I. Chengalur-Smith, and D. P. Ballou. The impact of experience and time on the use of data quality information in decision making. *Information Systems*, 14(2):170–188, 2003.

[37] S. Franconeri, J. Scimeca, J. Roth, S. Helseth, and L. Kahn. Flexible visual processing of spatial relationships. *Cognition*, 122:210–227, 2012.

[38] B. Fredrickson. What good are positive emotions? *Review of General Psychology*, 2(3):300–319, 1998.

[39] B. Fredrickson and C. Branigan. Positive emotions broaden the scope of attention and thought-action repertoires. *Cognition and Emotion*, 19(3):313–332, 2005.

[40] K. Gasper and G. Clore. Attending to the big picture: Mood and global versus local processing of visual information. *Psychological Science*, 13(1):34–40, 2002.

[41] A. S. Goeritz. The induction of mood via the www. *Motivation and Emotion*, 31(1):35–47, 2007.

[42] T. Green and B. Fisher. The impact of personality factors on visual analytics interface interaction. *IEEE Visual Analytics Science and Technology*, pages 203–210, 2010.

[43] T. Green and B. Fisher. Towards the personal equation of interaction: The impact of personality factors on visual analytics interface interaction. *IEEE VAST*, pages 203–210, 2010.

[44] T. Green, W. Ribarsky, and B. Fisher. Building and applying a human cognition model for visual analytics. *Information Visualization*, 8(1):1, 2009.

[45] L. Harrison, R. Chang, and A. Lu. Exploring the impact of emotion on visual judgement. *IEEE VAST*, 2012.

[46] C. Healey and J. Enns. Attention and visual memory in visualization and computer graphics. *IEEE TVCG*, 8(7):1170–1188, 2012.

[47] J. Heer and M. Bostock. Crowdsourcing graphical perception: Using mechanical turk to assess visualization design. *ACM CHI*, pages 203–212, 2010.

[48] D. Hermans, J. D. Houwer, and P. Eelen. A time course analysis of the affective priming effect. *Cognition and Emotion*, 15(2):143–165, 2001.

[49] R. Heuer. *Psychology of intelligence analysis*. United States Govt Printing Office, 1999.

[50] J. Hollands and I. Spence. Judging proportion with graphs: The summation model. *Applied Cognitive Psychology*, 12(2):173–190, 1998.

[51] W. Huang, P. Eades, and S.-H. Hong. Measuring effectiveness of graph visualizations: A cognitive load perspective. *Information Visualization*, 8(3):139–152, Jan. 2009.

[52] J. Hullman, E. Adar, and P. Shah. Benefitting infovis with visual difficulties. *IEEE TVCG*, 17(12):2213–2222, 2011.

[53] J. Hullman, E. Adar, and P. Shah. The impact of social information on visual judgments. *ACM CHI*, pages 1461–1470, 2011.

[54] A. Isen. Positive affect facilitates creative problem solving. *Journal of Personality and Social Psychology*, 52(6):1122–1131, 1987.

[55] A. Isen. The influence of positive affect on clinical problem solving. *Medical Decision Making*, 11(3):221–227, 1991.

[56] P. Johnson-Laird and K. Oatley. Cognitive and social construction in emotion. *Handbook of Emotions, 2nd ed.*, page 458, 2000.

[57] R. Johnston. Analytic culture in the US intelligence community: An ethnographic study. *The Center for the Study of Intelligence*, 160:72, 2005.

[58] S.-H. Kim, Z. Dong, H. Xian, B. Upatising, and J. S. Yi. Does an eye tracker tell the truth about visualizations?: Findings while investigating visualizations for decision making. *IEEE TVCG*, 18(12):2421–2430, 2012.

[59] K. Klein and A. Boals. Expressive writing can increase working memory capacity. *Journal of Experimental Psychology*, 130(3):520–533, 2001.

[60] R. Kosara, F. Drury, L. Holmquist, and D. Laidlaw. Visualization criticism. *IEEE Computer Graphics and Applications*, 28(3):13–15, 2008.

[61] B. Kwon, B. Fisher, and J. Yi. Visual analytic roadblocks for novice investigators. *IEEE Visual Analytics Science and Technology*, pages 3–11, 2011.

[62] H. Lam. A framework of interaction costs in information visualization. In *IEEE InfoVis*, 2008.

[63] P. Lang, M. Bradley, and B. Cuthbert. International affective picture system: Affective ratings of pictures and instruction manual. *NIMH Center for the Study of Emotion & Attention*, 2008.

[64] P. Lang, M. Bradley, and B. Cuthbert. International affective picture system (iaps): Affective ratings of pictures and instruction manual. *Technical Report A-8. University of Florida, Gainesville, FL*, 2008.

[65] R. Larsen and B. Fredrickson. *Measurement issues in emotion research*, pages 40–60. 2003.

[66] S. Lewis, M. Dontcheva, and E. Gerber. Affective computational priming and creativity. *ACM CHI*, pages 735–744, 2011.

[67] H. Lipford, F. Stukes, W. Dou, M. Hawkins, and R. Chang. Helping users recall their reasoning process. In *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on*, pages 187 –194, 2010.

[68] I. Livingston, L. Nacke, and R. Mandryk. Influencing experience: the effects of reading game reviews on player experience. *Entertainment Computing–ICEC 2011*, pages 89–100, 2011.

[69] G. Loewenstein and J. S. Lerner. *The role of affect in decison making*, pages 619–642. Oxford University Press, 2003.

[70] G. L. Lohse. The role of working memory on graphical information processing. *Behaviour & Information Technology*, (16):297–308, 1997.

[71] J. Mackinlay, P. Hanrahan, and C. Stolte. Show me: Automatic presentation for visual analysis. *IEEE TVCG*, 13(6):1137–1144, 2007.

[72] W. Mason and S. Suri. Conducting behavioral research on amazons mechanical turk. *Behavior Research Methods*, 44(1):1–23, 2012.

[73] L. Matzen, L. McNamara, K. Cole, A. Bandlow, C. Dornburg, and T. Bauer. Proposed working memory measures for evaluating information visualization tools. In *BELIV '10*, 2010.

[74] L. Micallef, P. Dragicevic, and J.-D. Fekete. Assessing the effect of visualizations on bayesian reasoning through crowdsourcing. *IEEE TVCG*, 19(12):2536–2545, 2012.

[75] G. Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2):81–97, 1956.

[76] S. Miller, A. Kirlik, A. Kosorukoff, and J. Tsai. Supporting joint human-computer judgment under uncertainty. *Human Factors*, 52(4):408–412, 2008.

[77] R. Morris, M. Dontcheva, and E. Gerber. Priming for better performance in micro-task crowdsourcing environments. *IEEE Internet Computing*, 16(5):13–19, 2012.

[78] T. Munzner. A nested process model for visualization design and validation. *IEEE Visualization and Computer Graphics*, 15(6):921–928, 2009.

[79] D. Organ. Extraversion, locus of control, and individual differences in conditionability in organizations. *Journal of Applied Psychology*, 60(3):401, 1975.

[80] F. Paas, J. E. Tuovinen, H. Tabbers, and P. W. M. Van Gerven. Cognitive load measurement as a means to advance cognitive load theory. *Educational Technology*, 38(1):63–71, 2003.

[81] F. G. Paas and J. J. Van Merriënboer. The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors*, 35(4):737–743, 1993.

[82] E. Phelps and M. Carrasco. Emotion facilitates perception and potentiates the perceptual benefits of attention. *Psychological Science*, 17(4):292–299, 2006.

[83] R. Picard and S. Daily. Evaluating affective interactions: Alternatives to asking what users feel. *CHI Workshop on Evaluating Affective Interfaces: Innovative Approaches*, 2005.

[84] R. Picard and J. Scheirer. The galvactivator: A glove that senses and communicates skin conductivity. *ACM CHI*, pages 1538–1542, 2001.

[85] R. W. Picard, E. Vyzas, and J. Healey. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:1175–1191.

[86] W. Pike, J. Stasko, R. Chang, and T. OConnell. The science of interaction. *Information Visualization*, 2009.

[87] S. Pinker. A theory of graph comprehension. In *Artificial Intelligence and the Future of Testing*, pages 73–126. 1990.

[88] P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. *International Conference on Intelligence Analysis*, pages 2–4, 2005.

[89] M.-Z. Poh, N. C. Swenson, and R. W. Picard. A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. *IEEE Transactions on Biomedical Engineering*, 57(5):1243–1252.

[90] G. Pourtois, G. Thut, R. G. de Peralta, C. Michel, and P. Vuilleumier. Two electrophysiological stages of spatial orienting towards fearful faces: Early temporoparietal activation preceding gain control in extrastriate visual cortex. *Neuroimage*, 26(1):149–163, 2005.

[91] L. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257 –286, 1989.

[92] S. S. Reich and P. Ruth. Wason's selection task: Verification, falsification and matching. *British Journal of Psychology*, 73(3):395–405, Apr. 1982.

[93] J. Richard Heuer. *The Psychology of Intellligence Analysis*. 1999.

[94] N. Riche. Beyond system logging: Human logging for evaluating information visualization. BELIV '10, 2010.

[95] J. Rotter. Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and pplied*, 80(1):1, 1966.

[96] G. Rowe, J. Hirsh, and A. Anderson. Positive affect increases the breadth of attentional selection. *Proceedings of the National Academy of Sciences (PNAS)*, 104(1):383–388, 2007.

[97] J. A. Russell. Core affect and the psychological construction of emotion. *Psychological Review*, 110(1):145 – 172, 2003.

[98] T. A. Salthouse and D. R. Mitchell. Effects of age and naturally occurring experience on spatial visualization performance. *Developmental Psychology*, 26(5):845–854, 1990.

[99] J. Scholtz. Beyond usability: Evaluation aspects of visual analytic environments. In *IEEE Visual Analytics Science And Technology*, pages 145–150. IEEE, 2006.

[100] J. Scholtz. Progress and Challenges in Evaluating Tools for Sensemaking. In *CHI '08 Workshop on Sensemaking*, 2008.

[101] A. Shackman, I. Sarinopoulos, J. Maxwell, D. Pizzagalli, A. Lavric, and R. Davidson. Anxiety selectively disrupts visuospatial working memory. *Emotion*, 6(1):40–61, 2006.

[102] E. Solovey, P. Schermerhorn, M. Scheutz, A. Sassaroli, S. Fantini, and R. Jacob. Brainput: Enhancing interactive systems with streaming fNIRS brain input. In *Proc. CHI '12*, pages 2193–2202. ACM Press, 2012.

[103] W. Swinyard. The effects of mood, involvement, and quality of store experience on shopping intentions. *Journal of Consumer Research*, pages 271–280, 1993.

[104] J. Thomas and K. Cook. *Illuminating the path: The research and development agenda for visual analytics*, volume 54. IEEE, 2005.

[105] E. R. Thompson. Development and validation of an internationally reliable short-form of the positive and negative affect schedule (panas). *Journal of Cross-Cultural Psychology*, 38(2):227–242, 2007.

[106] A. Tversky and D. Kahneman. Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131, Sept. 1974.

[107] A. Tversky and D. Kahneman. The framing of decisions and the psychology of choice. *Science*, 211(4481):453–458, Jan. 1981.

[108] A. Tversky and D. Kahneman. Extentsional versus intuitive reasonong: The conjunction fallacy in probability judgment. *Psychological Review*, 90(4):293–315, 1983.

[109] J. van Wijk. The value of visualization. *IEEE Visualization*, pages 79–86, 2005.

[110] M. Velez, D. Silver, and M. Tremaine. Understanding visualization through spatial ability differences. In *IEEE Visualization*, pages 511–518. IEEE, 2005.

[111] C. Verheyen and A. Goeritz. Plain texts as an online mood-induction procedure. *Social Psychology*, 40(1):6–15, 2009.

[112] E. Vogel, A. McCollough, and M. Machizawa. Neural measures reveal individual differences in controlling access to working memory. *Nature*, 438(7067):500–503, 2005.

[113] P. Vuilleumier and J. Driver. Modulation of visual processing by attention and emotion: Windows on causal interactions between human brain regions. *Philosophical Transactions: Biological Sciences*, 362(1481):837–855, 2007.

[114] P. Vuilleumier and Y.-M. Huang. Emotional attention: Uncovering the mechanisms of affective biases in perception. *Current Directions in Psychological Science*, 18(3):148–152, 2009.

[115] C. Wickens and J. Hollands. *Engineering Psychology and Human Performance.* Prentice-Hall, Upper Saddle River, NJ, 1999.

[116] T. Wilson and J. Schooler. Thinking too much: introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology*, 60(2):181–192, 1991.

[117] Y.-y. Yeh and C. D. Wickens. Dissociation of performance and subjective measures of workload. *Human Factors*, 30(1):111–120, 1988.

[118] J. S. Yi. Implications of individual differences on evaluating information visualization techniques. BELIV '10, 2010.

[119] J. S. Yi, Y. ah Kang, J. Stasko, and J. Jacko. Toward a deeper understanding of the role of interaction in information visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1224 –1231, 2007.

[120] J. S. Yi, R. Melton, J. Stasko, and J. A. Jacko. Dust & magnet: multivariate information visualization using a magnet metaphor. *Information Visualization*, 4(4):239–256, 2005.

[121] C. Ziemkiewicz, R. Crouser, A. Yauilla, S. Su, W. Ribarsky, and R. Chang. How locus of control influences compatibility with visualization style. *IEEE Visual Analytics Science and Technology*, pages 81–90, 2011.

[122] C. Ziemkiewicz and R. Kosara. Preconceptions and Individual Differences in Understanding Visual Metaphors. *Computer Graphics Forum*, 28(3):911–918, 2009.

[123] C. Ziemkiewicz and R. Kosara. Laws of attraction: From perceptual forces to conceptual similarity. *IEEE Visualization and Computer Graphics*, 16(6):1009–1016, 2010.

[124] T. Zuk and S. Carpendale. Visualization of uncertainty and reasoning. *Smart Graphics*, pages 164–177, 2007.