2015

Manipulating and Controlling for Personality Effects on Visualization Tasks

Alvitta Ottley
Tufts University

R. Jordan Crouser
Tufts University, jcrouser@smith.edu

Caroline Ziemkiewicz
Tufts University

Remco Chang
Tufts University

Manipulating and Controlling for Personality Effects on Visualization Tasks

Author
Institution
email@institution.edu

October 28, 2013

Abstract

Researchers in human-computer interaction and visualization have recently been challenged to develop a better understanding of users' underlying cognitive processes in order to improve system design and evaluation. While existing studies lay a critical foundation for understanding the role of cognitive processes and individual differences in visualization, concretizing the intuition that each user experiences a visual interface through an individual cognitive lens is only half the battle. In this paper, we investigate the impact of manipulating users' personality on observed behavior when using a visualization. In a targeted study, we demonstrate that personality priming can result in changes in behavior when interacting with visualizations. We then discuss how this and similar techniques could be used to control for personality effects when designing and evaluating visualizations systems.

1 Introduction

Recent research indicates that personality plays an important role in performance when using a visualization system [7, 19] and that measures of some personality traits can serve as predictors of a user's willingness to adapt their mental model to various visual metaphors [19]. What these findings underscore is the importance of considering individual differences such as users' personality when designing new visualizations. Given the increasing number of known individual differences that affect performance [7, 8, 15, 19], it is likely that one or more can cause a user to perform less optimally on a visual design. The findings also highlight how nuances in personality can affect evaluation, indicating that traditional measures of efficiency (speed and accuracy) may not fully capture the impact of a visualization design. Now armed with a better understanding of how individual differences impact performance with various visualizations, the question arises: Can we prime users to control for personality effects?

Researchers in HCI and psychology have investigated the efficacy of priming an individual's personality with the intent of temporarily influencing behavior. Studies include using emotionally-charged visual stimuli to inspire creativity [15] and eliciting varying levels of conformity by having users read words with positive or negative connotations [4]. Results indicate that noninvasive priming tasks can result in significant behavioral changes. In light of these findings, we posit that through priming, we may also be able to elicit performance changes when using visualization systems, specifically changes in speed and accuracy.

We focus our study on the effects of priming a well-established personality trait known as *locus of control* (LOC). LOC measures the extent to which an individual believes that events are determined by their actions or by external forces. Persons with a more *internal* LOC feel strongly that events are contingent upon their own actions, and those with a more *external* LOC believe that events are shaped more by outside forces than by their own influence. Research in personality psychology suggests that an individual's LOC may vary over time, and may even be intentionally manipulated [6, 11]. Additionally, a recent study by Ziemkiewicz et al. showed a strong correlation between LOC score and users' speed and accuracy on visualizations employing varying visual metaphors [19].

Inspired by these preliminary findings, we hypothesize that **we can significantly influence a user's speed and accuracy on visualization tasks by using priming techniques.** Specifically, we expect that prompting an average user to be more internal will make them exhibit the behavior of internal participants from Ziemkiewicz et al., and we expect a reverse effect when prompting an average user to be more external. Similarly, we posit that prompting internal participants to be more external and external participants to be more internal will lead to a reduction of differences between the groups, if not a full reversal of the original effect.

To test our hypothesis, we replicate the experimental design by Ziemkiewicz et al. [19] using existing priming techniques to manipulate LOC [6]. After priming, we measure users' performance as they complete search and inferential tasks on two hierarchical visualizations. We recruited three hundred online subjects with varying LOC scores via Amazon's Mechanical Turk and demonstrate that a small priming task is sufficient to significantly impact performance when performing complex tasks.

2 Background

There exist many well-vetted techniques for priming personality factors. For example, Epley et al. used nonconscious priming to elicit conformity to social pressures [4]. Participants were given a scrambled sentence task containing words related to either conformity or rebellion, and were told that they were performing a pilot test. They found that this priming task was sufficient to elicit conformity in a subsequent social scenario. Chalfoun et al. introduced subliminal cues to enhance users' learning capabilities when using tutoring systems [3]. They found that by including positive cues such as hints to encourage inductive thinking, they were able to increase reasoning ability and improve decision making when solving logic problems.

Lewis et al. studied the effect of affective priming on creativity by having participants choose images for their background when using drawing applications [15]. Participants were asked to select an image from one of the following randomly-assigned sets of priming stimuli: positive (e.g. smiling babies), neutral (e.g. hammers), negative (e.g. images of dead bodies) and control (no image). They found that individuals who were positively primed produced the most creative drawings as well as the highest quantity of drawings, providing evidence that affective priming can have a positive influence on productivity and creativity.

Harrison et al. extended their study of affective priming by measuring its impact on performance using information visualization [8]. Participants were asked to read a short story designed to elicit either positive or negative emotional responses. They were then asked to perform standard graphical perception tests. The results indicated that positive affective priming can be used to improve accuracy when interpreting certain visualizations [8].

2.1 Priming LOC

Similar priming techniques have been used to manipulate the personality trait *locus of control* (LOC). As previously stated, LOC measurers the extent to which a person believes that external events are influenced by their own behavior. Using the Rotter construct [18], individuals are scored on a 23-point scale where the extreme high and low ends of the scale are categorized as *internal* and *external*, respectively. Persons who score higher on the scale (*internals*) believe that events are contingent upon their own actions, while persons who score lower on the scale (*externals*) believe that events are controlled by external forces or supernatural beings.

Research in psychology has shown that higher LOC scores correlate with increased effectiveness at work [12], better academic performance [5] and greater ability to cope with stress [1]. A person's orientation on the locus of control scale also affects their learning style. Cassidy and Eachus [2] showed that internals are more likely to practice deep learning, while externals are more likely to practice surface learning. This implies that there is a relationship between locus of control and general problem solving techniques, which suggests a potential effect of locus of control on problem solving using visualizations.

Fisher et al. [6] used priming to investigate the effects of psychological intervention targeting LOC in persons with disabilities. In this study, patients with chronic lower back pain were randomly primed to score higher (measure more internal) or lower (measure more external) on the LOC scale. Researchers primed participants' LOC by asking experience recall questions: participants were asked to describe either times when they felt in control (thereby increasing their LOC score) or times they did not feel in control (thereby decreasing their LOC score). They found a significant difference in LOC scores before and after the application of this priming technique. They then assessed patients' perceived and actual physical ability using a lifting task. They reported that patients who were primed to be more internal spent more time on average performing the lifting task and selected heavier weights than the externally-primed group. Patients primed to be more external were significantly more likely to decline to participate in the lifting task. These results indicate that LOC can be primed using experience recall, and also demonstrates that priming LOC can influence behavior.

2.2 LOC and Performance on Visualization Systems

A 2010 study by Green et al. found a correlation between LOC and performance on two hierarchical visualizations [7]. They reported that LOC can be used to predict completion times as well as insight when using the two visualization tools. However, the source of the distinction between the two interfaces was ambiguous due to significant design differences between the two visualization systems.

Ziemkiewicz et al. extended this work by simplifying the visualizations studied, in order to isolate the variable of layout design [19]. They replicated Green et al.'s original study using four visualizations with visual metaphors that gradually shift from list-like to containment-like (see Figure 1). Their findings suggest that the observations reported by Green et al. were likely an interaction between LOC and the visualizations' *layout style*. As the visualizations became more container-like, participants with an internal LOC performed more poorly on inferential questions, while those with an external or average LOC performed equally well across all visual layouts. They hypothesized that an internal LOC may be associated with greater difficulty at adapting to visualizations with a strong containment metaphor, and that these participants may have difficulty with novel visual layouts.

While these studies demonstrate a relationship between LOC and performance with various visual layouts, a user's LOC is not always stable [6, 11], making the prospect of designing or evaluating based on user personality more complicated. However, we demonstrate that this instability may in fact be advantageous. In this work, we show that we can use priming to manipulate LOC in order to achieve performance changes when using a visualization system and discuss how this can be used to control for personality effects when designing and evaluating visualizations systems.

3 Experiment

To test our hypotheses we replicated the study by Ziemkiewicz et al. [19], holding constant the views, datasets, and questions to enable us to make accurate comparisons between the two results. We conducted a targeted study extending the prior work by applying the experience recall techniques introduced by Fisher et al. [6] to manipulate participants' LOC. We measured participants' baseline LOC prior to the main task using a 23-point Rotter LOC Scale [18] and used this score to assign priming groups:

- Participants who scored **higher than 15** (designated *internal*) were given a task designed to **decrease** LOC score and assigned to group $I^{\rightarrow E}$. Participants of this group were expected to exhibit performance measures that are similar to the average participants reported by Ziemkiewicz et al. [19].

- Participants who scored **lower than 10** (designated *external*) were given a task designed to **increase** LOC score and assigned to the group $E^{\rightarrow I}$. These participants were also expected to exhibit performance measures that are similar to the average participants of the previous study.

- Participants who scored **between 10 and 15** (designated *average*) were randomly given a priming task and assigned to the appropriate group, either $A^{\rightarrow I}$ or $A^{\rightarrow E}$. We anticipated that average users who were primed internal or external, would exhibit performance measures similar to the internal users of the previous study, while those who were primed external would exhibit performance measures similar to the external users.

(a) V1: Basic Tree View (b) V2: Nested Boxes View

Figure 1: The two visualization used in the current study. They were two of the four visualizations designed by Ziemkiewicz et al [19]. V1 was designed to have a list-like metaphor while v2 was designed to be container-like.

For simplicity, we used only the most extreme views (see Figure 1) presented by Ziemkiewicz et al. [19], as their results were most compelling for these views. The order in which the views were presented was randomized, and participants were asked to complete a search and an inferential task for each view.

3.1 Participants

We recruited 300 participants via Amazon's Mechanical Turk service. Mechanical Turk is an online market place that allows individuals to be recruited to complete small tasks for remuneration. Though the service is increasingly being used as a research tool because a large number of diverse participants can be recruited in a short period of time [10, 9], it is not without reservations. There are indeed some factors that may affect the validity of the data gathered from online resources such as vote flooding and lack of incentive for completion [14], but the absence of these factors have heightened the appeal for the use of Mechanical Turk in Human-Computer Interaction and Visualization research. It is especially useful for studies such as this, where there is a ground truth for evaluating the results [13] and incentives can be given for correct responses [14].

Of the 300 recruited, we discarded the results of 71 participants for failure to complete the task as required. Participants' data were also discarded if their interaction times were impractical (less than 10 seconds) and they also had no correct responses. The average LOC score was 12 on a 23-point scale ($\sigma = 4.49$) and there were 59 externals, 106 users with average LOC and 36 internals.

3.2 Materials

3.2.1 Locus Of Control Survey

Prior to beginning the experiment, each participant was given the a 29-question LOC personality survey. This survey is adapted from Rotter [18] and is comprised of a group of 23 forced-choice questions such that participants must choose either the external or internal response which best describes their belief. The remaining 6 questions are filler questions designed to disguise the purpose of the test. Once complete, a user was scored by counting the number of internal statements selected, with score of 0 meaning extreme external and a score of 23 meaning extreme internal [1]. The same locus of control test was administered at the end of the experiment to maintain consistency between the scoring.

3.2.2 Priming

Participants were assigned to one of two priming tasks which were slight modifications of those used

[1]This is reverse scored from the original Rotter survey which counts the number of external responses instead of internal responses.

by Fisher et al. [6]. In the first condition, participants were asked to describe times when they felt in control, which was designed to increase their LOC score (shifting their LOC toward *internal*). In the second condition, participants were asked to recall times when they felt they were not in control, which was designed to lower their LOC score (shifting their LOC toward *external*). Because our study was conducted via Amazon's Mechanical Turk, users completed priming tasks by entering at least three free-text examples of 100 words each to ensure effective priming. Below are the two priming stimuli used for this study.

Priming Question 1 (Increase Locus of Control) "We know that one of the things that influence how well you can do everyday tasks is your sense of control over problems you face. The more control you believe you have, the better you will succeed at the things you try and do. If you feel optimistic and able to make the best of your situations, you will do very well. In the spaces provided below, give 3 examples of times when you have felt in control and achieved things well. Each example must be at least 100 words long."

Priming Question 2 (Reduce Locus of Control) "We know that one of the things that influence how well you can do everyday tasks is the number of obstacles you face on a daily basis. If you are having a particularly bad day today, you may not do as well as you might on a day when everything goes as planned. Variability is a normal part of life and you might think you can't do much about that aspect. In the spaces provided below, give 3 examples of times when you have felt out of control and unable to achieve something you set out to do. Each example must be at least 100 words long."

3.2.3 Views

Instead of using the four visualizations as described in Ziemkiewicz et al., we simplified our study and used only the two extreme visualizations, V1 and V2. V1 displays data in a list-like fashion similar to that of Internet Explorer where hierarchy relationship is represented by indentation, and V2 employs a *hierarchy as containment* [20] metaphor and uses nested boxes to represent relationship. Like Ziemkiewicz et al. we limited the exploration of the visualization so that only one sub-tree can be explored and if the user attempted to explore another, the previously explored sub-tree would collapse to its original position. This maintains consistency with respect to the maximum amount of information that is displayed by each view. We also maintained consistency between the visualizations by keeping icons, text sizes and interaction styles constant for the two views.

3.2.4 Datasets

Dataset Name	Number of Species	Non-leaf Nodes
Amphibia	92	94
Aves	112	145
Eutheria	92	94
Lepidosauria	99	126

Table 1: The size of the four datasets used in the study.

The datasets presented in this study are also the same as the ones used by Ziemkiewicz et al. The previous work adapted four subsets of the full taxonomic tree from the National Center for Biotechnology Information's Genome database [16]. For every leaf, they presented information for that species such as the date when the entry was last updated and the number of proteins and genes stored in the database. Each dataset consisted of a phylogenetic tree where each leaf node is a species. The four datasets had an average of 98.75 leaf nodes (individual species) and 114.75 non-leaf nodes (Table 1).

3.2.5 Tasks

Consistent with the previous study, participants were then asked to complete a search task and an inferential task using each view. For each task, users were presented with a question and were expected to explore the visualization to retrieve the answer (Table 2). The search task was a simple look-up task whereby the user was asked to find information about a single species within a specific category. For example,

> Under the classification "Falco," find the species with a "Length" value over 18000.

Table 2: The eight task questions seen in the study.

Dataset	Question Type	Question
Amphibia	search	Within the classification "Batrachuperus," which species was most recently updated?
	inferential	Under "Anura," find the classification "Bufo" and note the subclasses it contains. There is another classification under "Mesobatrachia" that has something notable in common with "Bufo." Find that classification.
Aves	search	Under the classification "Falco," find the species with a "Length" value over 18000.
	inferential	Looking in "Sphenisciformes," find the classification "Eudyptula" and note the species under it. Now look in "Threskiornithidae" for a classification that has something notable in common with "Eudyptula."
Eutheria	search	Within the classification "Tarsius," find the species which was most recently updated.
	inferential	Under "Caniformia," find the classification "Canis" and note the subclasses and species it cotains. Now find another classification under "Ursidae" that has something notable in common with "Canis."
Lepidosauria	search	Under the classification "Bipes," find the species with the lowest "Length" value.
	inferential	Within "Scincomorpha," find the classification "Lacerta" and note the species under it. Now look in "Crotalinae" for a classification which has something in common with "Lacerta."

The inferential task asked the user to first find a specific classification and then find another classification with similar properties, forming a more complex analytical task. For example:

> Looking in "Sphenisciformes," find the classification "Eudyptula" and note the species under it. Now look in "Threskiornithidae" for a classification that has something notable in common with "Eudyptula".

3.3 Procedure

Participants were first asked to complete the LOC survey. Once this was completed, their locus of control score was calculated and each participant was then issued one of the two experience recall questions. The main portion of the study consisted of two sessions, one with each view, and for each session, participants were presented with a search task then an inferential task. After completing both sessions, they once again completed the LOC survey for which the order of the questions were altered.

We measured each participant's training times, interaction times and times taken to record their responses. Additionally, we recorded their initial LOC score as well as their post survey LOC scores.

4 Results

4.1 Priming

For all but one group ($A^{\rightarrow I}$), the priming prompts were successful at influencing the participants' LOC scores in the desired direction. We ran t-tests on the pre-test and post-test LOC scores for each of the priming groups, and in each case the change was significant at a $p < .01$ level. The group $A^{\rightarrow I}$ was not significant with $p = .3$. Although we observed the statistically significant changes with the other groups, on average the mean difference was small in each case, with the mean magnitude of difference being $M = 1.69$ in the average group, $M = 1.45$ in the internal group, and $M = 1.37$ in the external group.

4.2 Impact on Response Time

Although the change in LOC scores was not dramatic, our analysis revealed strong evidence that priming successfully caused performance changes on the visualization tasks. Consistent with findings by Ziemkiewicz et al. [19], we found significant differences in response time for inferential questions but not for search questions. Therefore, in the following analyses, we refer only responses to inferential questions unless otherwise indicated.

We found a high degree of variability among our subjects in terms of overall response time to inferential questions ($M = 263.5s, \sigma = 240.8$). For this reason, we chose to analyze our primary hypothesis using a repeated measures design. To test our main hypothesis, we performed a repeated

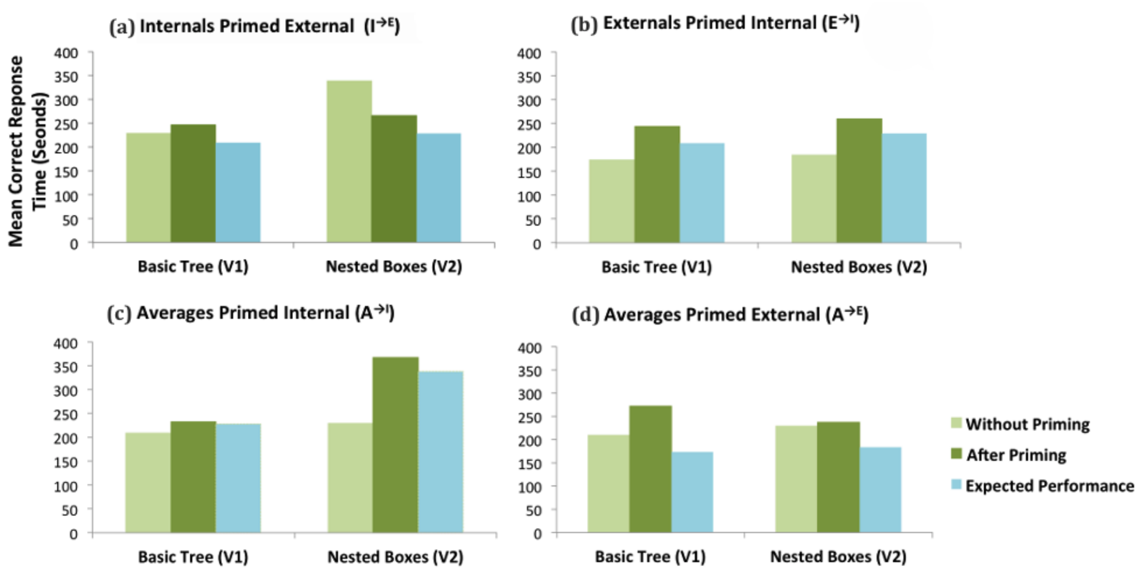

Figure 2: Mean correct response times on inferential task questions across the two views for each of the four priming groups. The average participants were successfully primed to behave as internal participants, while the internal and external participants were successfully primed to be more average.

measures ANOVA on Correct Response Time using a 2×4 mixed design of Visual Layout (within-subjects) by Priming Group (between-subjects). The ANOVA uncovered no significant main effect of Visual Layout, perhaps because the differences in performance across priming groups counteracted the overall differences in the effectiveness of each layout type. However, this test revealed a significant interaction between Visual Layout and Priming Group, $F(3, 57) = 2.85, p < .05$. This finding indicates that different priming groups showed significantly different patterns of performance between the two views.

Analysis indicates that introducing a priming stimulus was generally successful in affecting participants' performance. For group $I^{\to E}$, the response time difference between V1 and V2 was much smaller than those reported by Ziemkiewicz et al. [19], which indicates that participants with an internal baseline LOC were successfully primed to elicit the behavior of average users reported in the previous study (see Figure 2(a)). Likewise, priming the group $E^{\to I}$, successfully elicited performance measures similar to the average users of the previous study (see Figure 2(b)).

Priming was equally effective at influencing the performance of participants with an average baseline LOC. Ziemkiewicz et al. reported that average-scoring participants showed no significant difference in response time or accuracy across the four visualization views [19]. As described in the previous section, we primed half of these participants to be more external and half of them to be more internal. The response time results for both groups indicate that priming was indeed successful in causing performance differences:

- Group $A^{\to I}$ performed much more slowly on V2, with $M = 456.2s$ and $\sigma = 541.4$, as compared to V1, with $M = 215.5s$ and $\sigma = 177.4$ (see Figure 2(c))
- Group $A^{\to E}$ was conversely faster with V2, with $M = 231.7s$ and $\sigma = 137.1$ than with V1, with $M = 348.8s$ and $\sigma = 438.1$, though to a lesser degree (see Figure 2(d))

The behavior of group $A^{\to I}$ mimics the results for highly internal participants reported by Ziemkiewicz et al. [19]. While the behavior of group $A^{\to E}$ is clearly distinct from the average-scoring participants reported by Ziemkiewicz et al. [19], it does not replicate the results for their highly externals. Instead of showing no difference in response time between V1 and V2, ($t(46) = 2.01, p = .68$).

4.3 Impact on Accuracy

In addition to speed, we also examined participants' accuracy levels across views and priming groups. While we found no significant effects directly related to our initial hypothesis, we did find an unexpected result: when counts of accurate responses were aggregated across question type and visual layout, we found a significant effect of priming group on accuracy, $\chi^2(3, N = 916) = 19.94, p < .001$. Upon closer analysis, it is clear that this effect stems from the fact that participants in group $E^{\to I}$ were far more accurate overall with 63.5% correct responses. Differences in accuracy across the other three priming groups were not statistically significant: $I^{\to E}$ 49.2%; $A^{\to E}$, 45.1%; and $A^{\to I}$, 46.5%. These are comparable to the overall accuracy reported by Ziemkiewicz et al. [19]. While the previous study found that external participants were slightly more accurate overall than other participants, the magnitude of the difference was not comparable to that found in the current study.

Perhaps since the priming effect size was small on average, there was no significant correlation between amount of priming and performance difference. One notable finding is that those participants who did show a priming effect, i.e. those whose post-test score showed a difference in the direction predicted by the priming stimulus, were significantly more accurate overall. We counted accurate responses across all four questions (including search questions) and compared this accuracy count between the group that demonstrated an expected score change and the group that did not. The result was significant, $t(227) = 2.26, p < .05$. It is possible that this indicates that these participants were paying more attention to all parts of the study, affecting both the degree of their priming and their performance on the visualization tasks.

4.4 Other Findings

Along with our primary hypotheses, we also examined whether other factors had an effect on performance between the two views. While these are not directly related to our main hypothesis, they do provide context for our findings. One question we studied was whether the base LOC reported by the participant at the beginning of the study affected their performance, as it did in the study by Ziemkiewicz et al. We found weak evidence that this may be the case. To test this question, we studied the correlation between pre-test LOC score and the extent of inferential response time difference between the tree view and the nested boxes view. We named this calculated variable "tree advantage", and note that it is a similar measure to the differences tested by our repeated measures ANOVA. Positive values of tree advantage correspond to faster performance on the basic tree view (V1) thus signifying a behavior like internals of previous studies. Negative values correspond to faster performance on the nested boxes view (V2) and a value of 0 means there was not performance difference between the two views (Figure 3).

We found no direct correlation between base LOC and tree advantage. However, the effect was less ambiguous when we split the participants by priming direction. Among participants who were primed to be more internal, the correlation was positive and approached significance ($r(35) = .31, p = .07$), and the same was true among participants who were primed to be more external ($r(26) = .35, p = .08$). A higher score on the LOC scale corresponds to a more internal locus. Therefore, a potential positive correlation suggests that within each group, there may be an effect such that the more internal a participant is, the more of a performance advantage the basic tree view offers, and vice versa. This suggests that users who measure high on the LOC scale (extreme internals) show little or no performance change after being primed.

One reason this effect doesn't appear when the two groups are aggregated is that, while the two trends are parallel, they are offset in such a way that the trend disappears when they are combined. While this effect is not statistically significant in the existing data, it does raise the possibility that, even in the presence of perceived control priming, innate LOC does still affect visualization performance difference. Taken together with the fact that the difference between pre- and post- LOC scores were small, there is a possibility that the performance difference seen may have been due to some other factor. A closer examination of the priming responses revealed that majority of the participants responded to the stimuli with emotionally charged life stories, and it is also possible that performance differences

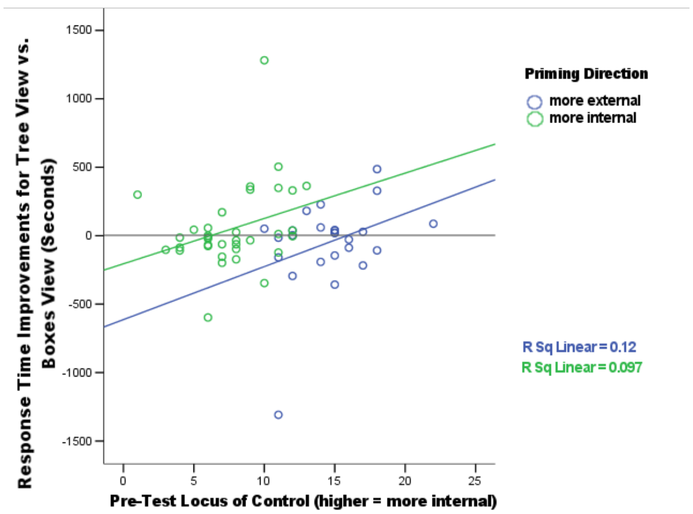

Figure 3: When users are divided by the direction in which they were primed, a correlation approaching (but not reaching) significance was found between pre-test LOC and the amount by which a participant performed faster with the basic tree view (V1). A higher score on the LOC scale indicates a more internal LOC. A positive value on the Y-axis indicates that the user was faster at answering inferential questions with the basic tree view, while a negative value means that the user was faster at answering inferential questions with the nested boxes view. Within each priming condition, there was a near-significant effect such that a more internal LOC was associated with a greater performance advantage for the basic tree view.

could somewhat be due to changes in participants' emotional state. A more focused experiment would be needed to verify this claim.

5 Discussion

Our results confirm that we can successfully use priming to alter users' performance during complex tasks. We were able to both reduce differences between extreme user groups as well as create differences between average users using priming. It is important to note that the change in performance is both *statistically* and *practically* significant. By asking internal users to recall times when they did not feel in control we are able to effect a remarkable improvement in performance: the response gap between the two views for internal users were reduced from 110 seconds (roughly two minutes) to just 20 seconds. Notably, we also found that priming affected a user's response time more than their accuracy. This is also true of previous work on metaphor priming and individual differences [21]. This is evidence that interactions between visual style and a user's frame of mind may be more relevant in situations where efficiency is important.

For all but one group ($A^{\rightarrow E}$) we observed the expected completion times. While the average completion time for the $A^{\rightarrow E}$ group was slower than the external users from Ziemkiewicz et al., the evaluation revealed no significant difference between their performance on the two views . This is still consistent with external users from Ziemkiewicz et al. Overall, the observed completion times for the current study were slightly slower than the those reported by Ziemkiewicz et al. One possible explanation could be differences between the time of day and the time of the week when the HITs were posted.

Although we reported significant changes in completion times, the mean change in F was small for groups that showed significant differences between pre-test and post-test LOC scores. Feedback from participants suggests this may be partly due to the fact that some users remembered their responses from earlier and tried to answer consistently. This could be an artifact of the Rotter LOC survey where for each question, the user chooses one of two statements which best describes them. For instance, one question asked users to choose between a statement that suggests that *there will always be someone who doesn't like you* and one which suggests that *if you can't get people to like you then you don't understand how to get along with others*. While we believe priming does affect perceived control of one's environment, the Rotter LOC may not be the appropriate tool for measuring small changes. Future work could investigate the use of other LOC surveys which uses a Likert scale instead of a forced-choice scale.

To some extent, these results complicate previous findings on personality effects in visualization. We found no evidence that we are able to prime users with an extreme baseline LOC to adopt the behaviors of the other extreme using this technique. However, we were able to successfully prompt extreme users to exhibit behavior similar to averages. It is possible that users with a strong tendency in a certain personality trait can only be coaxed out of that tendency to a limited degree. This indicates that while volatile individual differences are still important factors to consider during design and evaluation, priming is not a panacea. In addition to furthering our understanding of the role of individual differences in future applications, these findings also shed light on the results of previous studies. In the next section, we discuss how this study relates to the design and evaluation of visualizations.

6 Implications

This and previous studies underscore that evaluating tools to help people think is a complicated endeavor. Our results suggest that traditional efficiency measures of speed and accuracy may not capture all of what we value in a visualization. While accuracy alone may not reflect the actual difficulty of a task, interaction time proves to be far too sensitive to minor changes in user inclination to provide generalizable information about a system. Evaluation must therefore go beyond simply analyzing the efficiency of a visual design but should also include methods that analyzes the user's cognitive factors.

6.1 Evaluation

The way people think and solve problem is often situation-dependent. It is entirely possible that subtle aspects of user study procedures, task question design, and even a researcher's behavior can initiate unintentional cognitive priming and contribute a participant feeling more or less in control. If task performance can be affected by a user's cognitive state, this kind of unintentional priming could harm the validity of evaluation results. This recalls Ziemkiewicz and Kosara's finding [20] that metaphors used in the wording of task questions can interact with visualization layout in an evaluation setting.

Priming can also be intentional. While we only focused on LOC for this study, previous work highlights how other cognitive states can also affect performance on visualization systems [8, 22]. In some cases, the interaction of cognitive states can negatively affect both a user's speed and accuracy and therefore negatively affect evaluation. One practical application of priming is that we may be able to negate these disadvantages. Before evaluating visualizations, researchers can explicitly or subconsciously nudge the user into a specific frame of mind. By subtly presenting a positive news article as participants wait to begin the experiment or displaying a positive picture, researchers could affectively prime participants, making them better suited to perform certain tasks. Researchers can also administer an "unrelated" pre-task to disguise priming stimuli.

6.2 Design

One possibility for visualization design is to prompt a user into a certain frame of mind better suited for the tool at hand. For example, verbal or textual elements such as instructions could also be tuned to temporarily prime the user, improving their capacity for working with a specific interface type. A system could use language in its instruction texts that primes the user to adopt a different frame. It may even be possible to design elements in a more subtle way. Lewis et al. demonstrated how the use of images can influence a user's emotional state [15]. Subtly including images in an interface design may also improve the performance of someone who is having a bad day. Indeed, such prompting may be implicitly at work in existing designs and may affect other cognitive states. Understanding this process better may make it possible to automate some of this design work.

Ottley at al. proposed the use of priming for adaptive systems [17]. Future systems can be equipped with a better understanding of users' cognitive states and automatically nudge users into a specific frame of mind. Priming may be used to counteract biases and encourage a user to experience an interface with a new perspective. They also proposed the use of priming in the design of collaborative systems. While is it often advantageous to have different perspectives, it is sometimes necessary for collaborators to see an interface as everyone else does. Priming has the potential to unify their conflicting perspectives when they exist.

That said, individual differences research remains necessary. One of our findings is that some users are simply more susceptible to prompting than others. Identifying these less flexible user groups and how they respond to varying visualization designs is still important if we are to completely understand how priming techniques can be used to control personality effects. While we have focused on LOC in this work, similar priming techniques exist for other cognitive states. These may also provide opportunities for controlling for individual differences within evaluation studies. However, it is first necessary to determine whether these cognitive states affect performance on visualization tasks.

7 Conclusion

In this paper, we demonstrated that by manipulating a user's LOC, we can prompt them to exhibit significantly different behavioral patterns. Our results also highlighted the sensitivity of evaluation measures such as response time to small situational changes. We believe that these findings help build toward understanding personality factors can affect the ways humans solve problems with visualizations and contribute to the development of systems that are robust to the effects of individual differences. We believe that this research helps build toward a symbiosis between the system and the user, where not only do users adapt their systems to better suit their analytical needs, but systems can also encourage adaptation by the user to enhance performance.

8 Acknowledgements

This work was supported by the National Science Foundation [grant number xxx].

References

[1] C. R. Anderson. Locus of control, coping behaviors, and performance in a stress setting: A longitudinal study. *Journal of Applied Psychology*, 62:446–451, 1977.

[2] S. Cassidy and P. Eachus. Learning style, academic belief systems, self-report student proficiency and academic achievement in higher education. *Educational Psychology*, 20:307–320, 2000.

[3] P. Chalfoun and C. Frasson. Subliminal priming enhances learning in a distant virtual 3d intelligent tutoring system. *IEEE Multidisciplinary Engineering Education Magazine*, 3:125–130, 2008.

[4] N. Epley and T. Gilovich. Just going along: Nonconscious priming and conformity to social pressure. *Journal of Experimental Social Psychology*, 35(6):578–589, 1999.

[5] M. J. Findley and H. M. Cooper. Locus of control and academic achievement: A literature review. *Journal of Personality and Social Psychology*, 44:419–427, 1983.

[6] K. Fisher and M. Johnston. Experimental manipulation of perceived control and its effect on disability. *Psychology and Health*, 11:657–669, 1996.

[7] T. M. Green and B. Fisher. Towards the personal equation of interaction: The impact of personality factors on visual analytics interface interaction. In *IEEE Visual Analytics Science and Technology (VAST)*, 2010.

[8] L. Harrison, D. Skau, S. Franconeri, A. Lu, and R Chang. Influencing visual judgment through affective priming. In *To Appear: Proceedings of the 2013 annual conference on Human factors in computing systems (CHI)*, 2013.

[9] J. Heer and M. Bostock. Crowdsourcing graphical perception: Using Mechanical Turk to assess visualization design. In *Proceedings of the 2013 annual conference on Human factors in computing systems (CHI)*, pages 203–212, 2010.

[10] P.G. Ipeirotis. Analyzing the amazon mechanical turk marketplace. *XRDS: Crossroads, The ACM Magazine for Students*, 17(2):16–21, 2010.

[11] M. Johnston, P. Gilbert, C. Partridge, and J. Collins. Changing perceived control in patients with physical disabilities: An intervention study with patients receiving rehabilitation. *British Journal of Clinical Psychology*, 31:89–94, 1992.

[12] T. A. Judge and J. E. Bono. Relationship of core self-evaluations traits—self-esteem, generalized self-efficacy, locus of control, and emotional stability—with job satisfaction and job performance: A meta-analysis. *Journal of Applied Psychology*, 86:80–92, 2001.

[13] A. Kittur, E. H. Chi, and B. Suh. Crowdsourcing user studies with Mechanical Turk. In *Proceedings CHI*, pages 453–456, 2008.

[14] R. Kosara and C. Ziemkiewicz. Do Mechanical Turks dream of square pie charts? In *BELIV Workshop*, pages 373–382, 2010.

[15] S. Lewis, M. Dontcheva, and E. Gerber. Affective computational priming and creativity. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, pages 735–744. ACM, 2011.

[16] National Center for Biotechnology Information. NCBI genome database. http://www.ncbi.nlm.nih.gov/genome/. Accessed 11/18/2010.

[17] A. Ottley, E. M. Peck, L. Harrison, and R. Chang. The adaptive user: Priming to improve interaction. *CHI 2013 Workshop, Many People, Many Eyes: Aggregation Influences of Visual Perception on User Interface Design*, 2013.

[18] J. B. Rotter. Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs*, 80(609), 1966.

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

[20] C. Ziemkiewicz and R. Kosara. The shaping of information by visual metaphors. *IEEE Transactions on Visualization and Computer Graphics*, 14:1269–1276, 2008.

[21] C. Ziemkiewicz and R. Kosara. Preconceptions and individual differences in understanding visual metaphors. *Computer Graphics Forum*, 28:911–918, 2009.

[22] C. Ziemkiewicz, A. Ottley, R.J. Crouser, A.R. Yauilla, S.L. Su, W. Ribarsky, and R. Chang. How visualization layout relates to locus of control and other personality factors. *IEEE Transactions on Visualization and Computer Graphics*, page 1, 2012.