Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Decoupling of Intuitions and Performance in the Use of Complex Visual Displays

Permalink

https://escholarship.org/uc/item/8hv9x7xc

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 30(30)

ISSN

1069-7977

Authors

Hegarty, Mary
Smallman, Harvey S.
Stull, Andrew T.

Publication Date

2008

Decoupling of Intuitions and Performance in the Use of Complex Visual Displays

Mary Hegarty[1] (hegarty@psych.ucsb.edu)
Harvey S. Smallman[2] (smallman@pacific-science.com)
Andrew T. Stull[1] (stull@psych.ucsb.edu)
[1]Department of Psychology, University of California, Santa Barbara, CA 93106, USA
[2] Pacific Science and Engineering Group, San Diego, CA 92121, USA

Abstract

Interactive display systems give users flexibility to tailor visual displays to different tasks and situations. But this flexibility can only be beneficial if users have meta-knowledge of the types of displays that are effective for different purposes. In previous studies with both Navy forecasters and undergraduate students, we found that users often prefer maps that display extraneous variables, especially those that add realism to the display, even when they are task-irrelevant. In the current study, we tested undergraduate students on a simple read-off and comparison task with weather maps and measured the effect of these extraneous variables on accuracy, response times, eye fixations, and intuitions about display effectiveness. Extraneous realism slowed response time and lead to more eye fixations on both task-relevant and task-irrelevant regions of the displays. In spite of these decrements in performance, and the fact that realism added no task-relevant information, about a third of participants persisted in favoring these realistic displays over non-realistic maps.

Keywords: Visual Displays, Naïve Realism, Eye Fixations.

Introduction

Research on comprehension and reasoning with visual displays has made it clear that the effectiveness of a visual display depends on the task to be performed with that display (Bertin, 1983; Cheng, 2002; Larkin & Simon, 1987; Shah, Freedman & Vekiri, 2005). The same task may be performed more or less efficiently with different types of displays, and a display that is effective for one task may not be useful for another.

Given the task-dependent nature of display effectiveness, how can cognitive scientists and designers best support information processing with visual displays? One traditional human factors approach is to analyze the information needs and tasks to be accomplished by particular users and design displays that best support these tasks. This is a good approach for bounded task domains with stable information needs over time. But for users who perform many diverse tasks with different information demands, this may not be feasible.

An alternative solution for less stable domains is to give the user control over parameters of the display, so that he or she can customize displays at will. However, it is important to realize that this solution puts the burden of the design process on the user. When users are allowed to customize their own displays, effective performance relies more on their knowledge and intuitions about which displays are most effective for different tasks.

Worryingly, past research suggests that users often have poor intuitions about what makes effective graphical displays (Smallman & St. John, 2005). In particular, users have misplaced faith in realistic representations and their ability to extract information from them–they mistake familiarity for effectiveness. Smallman and St. John term this misplaced faith *Naïve Realism*. For example, Navy users prefer spatially realistic 3D icons of ships and planes on their displays but these features lead to slow, error-prone identification. Similarly, users predict they will need high fidelity realistic 3D displays to lay routes across terrain when they actually perform the task better with lower fidelity displays that unmask the valleys and other avenues through the terrain necessary to successfully perform the task (Smallman, Cook, Manes, & Cowen, 2007).

Meteorology offers a rich domain in which to study these issues. Forecasters use weather maps for a variety of tasks, including reconciling model data with observations, generating forecasts for different client needs, and issuing warnings of severe weather events. The displays that they use while performing these tasks typically show a variety of different meteorological variables (pressure, wind, temperature, etc.) superimposed on the map. Existing display systems give forecasters a great deal of flexibility and tailorability in terms of what variables are shown (Hoffman, Detweiler, Conway, & Lipton, 1993). Given the flexibility they have, how well do users tailor their displays to best show the information they need?

In a recent naturalistic observation of twenty one Navy weather forecasters (Smallman & Hegarty, 2007), we found that when performing a forecasting task, participants accessed weather maps that were more complex than they needed, displaying variables that were extraneous to their task. This effect was exacerbated with forecasters of lower spatial ability. That is, low-spatial forecasters put more extraneous variables in their displays and their forecasts were somewhat less accurate.

In his cognitive analysis of principles of graphics design, Kosslyn (1989) states as a cardinal rule that "no more or less information should be provided than is needed by the user:" (p. 211). Tufte (1983) also cautions against including extraneous information in visual displays, calling this information "chartjunk". In adding extraneous variables to their displays, these meteorologists are violating a basic principle of effective graphics. The question we ask here is to what extent these extraneous weather map variables actually impair performance.

In a preliminary laboratory study on this question Canham, Hegarty, and Smallman (2007) had participants perform simple comparison tasks with weather maps that displayed different numbers of extraneous variables. The extraneous variables were either off-task meteorological variables (e.g., adding temperature information to a display when the only variable to be compared was pressure) or realism (completely task-irrelevant terrain features and state boundaries). We also measured participants' intuitions about displays by asking them to choose from among several possibilities, the map that they would prefer to use when comparing weather variables in different regions of the map.

Performance was very accurate (over 95%). However, as Figure 1 shows adding realism to weather maps added over half a second to average response times, and as the number of displayed meteorological variables increased, time to answer questions increased significantly. In spite of these decrements in performance, fully a third of participant responses were in favor of showing realistic over non-realistic maps. Clutter from just one extraneous variable penalized performance, yet participants harbored misplaced faith in realistic representation.

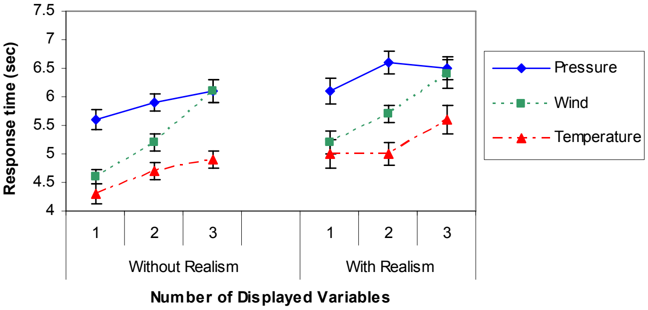


Figure 1. Mean response times observed by Canham, Hegarty, and Smallman (2007) (error bars show SEMs).

In the present study, we further investigate *why* individuals prefer realistic over non-realistic maps. In our previous study (Canham et al, 2007) participants were asked which of the maps they would *prefer* to use to perform a task. It is possible that their responses to this question were based on aesthetics rather than intuitions about their performance efficiency. We addressed this concern in the present study by asking an additional question that focused more on efficiency ("with which map would you perform *fastest*?"). If participants were responding on the basis of aesthetics in the previous study, they should be less likely to choose maps with extraneous realism when asked this efficiency question.

We also considered alternative accounts of *how* extraneous realism on a map affects performance on our weather comparison task. One possibility is that it takes more time to encode the relevant weather variables on these maps, because these variables are masked by the irrelevant realism or because encoding of the irrelevant information clutters the mental representation and has to be suppressed. Another possibility is that irrelevant realism on a map produces more eye fixations outside the task-relevant areas, for example, because users have to sample from a wider area to see the relevant patterns in the data, or because the irrelevant realism produces salient areas outside the relevant areas that capture users' attention.

To examine these possibilities, we monitored participants' eye fixations while they performed an attribute comparison task. If irrelevant variables in the regions to be compared mask the relevant variables, so that they are simply harder to see, time in the regions to be compared should increase. Additional time for the realistic maps might be due to longer fixations, more fixations, or both. If users have to sample from a wider area on realistic maps or the extraneous variables distract users' attention from the task relevant regions, then we should observe more fixations and time spent fixating regions of the display outside the relevant regions.

Method

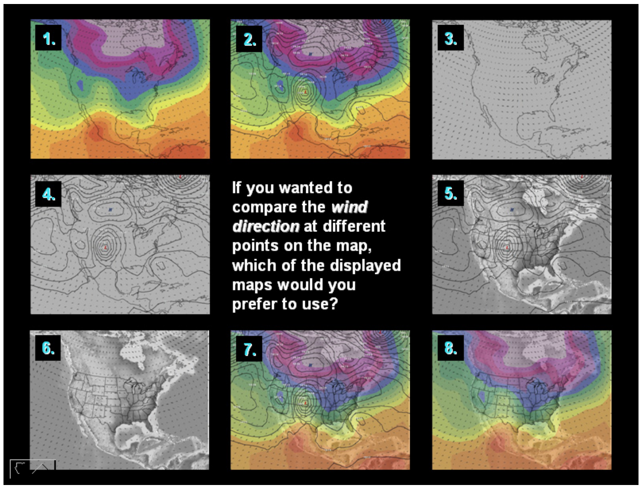
Participants

Twenty-six students at the University of California, Santa Barbara participated in the study, in return for either course credit or payment.

Materials

Maps. We created weather maps that depicted one of three weather variables on a map of the North American continent. The variables were surface temperature (depicted by color scale), surface air pressure (depicted by isobar contours), and surface wind (depicted by arrows of different sizes, such that the direction of the arrow indicated wind direction, and the size indicated wind speed). The coding schemes used were quite conventional. The maps were created using ArcGIS, based on data obtained from the national weather archive regarding the actual weather conditions on 12 different dates in the last 5 years. We created realistic and non-realistic versions of each weather map. For the realistic maps, the weather variables were superimposed on a map showing the surface topography of the North American continent and ocean floor (in light achromatic shaded relief), and the state and national boundaries (as dark lines). The non-realistic maps showed these variables on a map displaying only the shorelines of the North American continent (see Figure 2).

Intuition Trials. There were 18 intuition trials in the experiment. On each trial, participants were shown eight maps and were asked which map they would choose if they had to compare different points on the map in terms of a single meteorological variable (pressure, temperature or wind). On half of the trials they were asked to choose which map they would *prefer* to use to accomplish the task (see the example in Figure 2). On the other half they were asked with which of the eight displayed maps they would answer *fastest*. For example the set of maps shown in Figure 2 might be accompanied by the question "If you had to compare the wind direction at different areas of the map, with which map would you answer fastest".



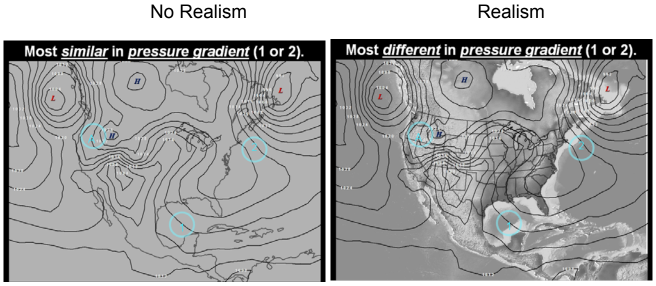
If you wanted to compare the *wind direction* at different points on the map, which of the displayed maps would you prefer to use?

Figure 2. Example of a preference trial.

Comparison Trials. On each trial of the comparison task, participants were shown a single weather map with three locations marked with the letter A and the numbers 1 and 2, respectively (see sample trials in Figure 3). The region marked A was always located within the borders of the United States. The task was to choose which of the two numbered regions (1 or 2) was either most similar or most different to the region marked A with respect to a single meteorological variable (pressure differential, wind direction, or temperature). A line of text above the map on each trial indicated whether participants were to choose the region that was most similar or most different, and indicated which weather variable was to be considered for that trial. This phase of the experiment had a 3 (meteorological variable to be compared) X 2 (realism) within participants design with 6 trials in each condition of the design for a total of 36 trials. Figure 3 shows examples of non realistic and realistic trials in the comparison phase in which the variable to be compared was pressure.

Apparatus

Eye movements were monitored by an SMI EyeLink head mounted eye tracking system which sampled participants' eye fixations at a rate of 250 Hertz. The aggregation software was set to detect saccades with an amplitude of 0.05 degrees or greater, an acceleration threshold of 9,500 degrees/second^2 and a velocity threshold of 30 degrees /second. Participants viewed images presented on a computer monitor while resting their chins on a chin rest, set 30 inches from a 15 by 11.5 viewing screen (21.7 by 28.1 degrees of viewing angle). The resolution of the screen was 800 by 600, with a refresh rate of 75 Hertz.



Most *similar* in *pressure gradient* (1 or 2).

Most *different* in *pressure gradient* (1 or 2).

Figure 3. Examples of comparison trials showing a non-realistic and a realistic map

Procedure

There were four phases in the experiment. First, participants were familiarized with the graphical conventions of the maps (the use of color to show temperature, isobar lines to show pressure, etc). Second they completed the 18 intuition trials. Third, the eyetracker was calibrated and they completed the 36 comparison trials while their eyes were tracked. Finally, participants completed the 18 intuition trials again.

Results

Intuition Trials

In the intuition trials, the optimal map choice (in terms of minimizing visual clutter) is the map that showed only the meteorological variable asked about and no extraneous realism. We categorized participants' map choices as (a) optimal if they chose this map, (b) "extraneous realism" if they chose the map showing the correct meteorological variable plus realism, (c) "extraneous weather information" if they chose a non-realistic map with additional meteorological variables and (d) "extraneous weather variables plus realism" if they chose a map that added both additional meteorological variable(s) and realism.

Table 1. Mean percentage of times that map types were chosen before and after performing the comparison task.

	Prefer?		Fastest?	
	Before	After	Before	After
Optimal Map (Chance: 12.5%)	56.5	73.9	60.4	83.5
Extraneous Realism (Chance: 12.5%)	34.3	26.5	33.5	16.1
Extraneous Weather Variables (Chance: 37.5%)	3.5	0.9	4.8	1.7
Extraneous Weather Vars. + Realism (Chance: 37.5%)	5.7	1.0	5.0	1.0

Table 1 shows the mean percentage of times that participants chose the different categories of maps and (in parentheses) the percentage of times they would have chosen these categories by chance had they picked randomly. Maps with extraneous variables were chosen almost one third of the time and participants frequently chose the map with the correct meteorological information but with extraneous realism. One sample t-tests comparing observed to chance values indicated that participants were significantly more likely than chance to choose both the optimal map (p <.001 in all cases), and the map that showed the optimal variables with a realistic rendering ($p < .05$ in all cases). They were much less likely than chance to choose maps with extraneous meteorological information ($p < .001$ in all cases).

Participants were more likely to choose the optimal map after performing the comparison task compared to before, $F(1, 25) = 8.04$, $p = .009$, $\eta_p^2 = .24$. They were marginally more likely to chose the optimal map when they were asked the efficiency question (with which map would you answer fastest?) than when asked the preference question (which map would you prefer to use?), $F(1, 25) = 3.76$, p=.06, η_p^2 = .13, suggesting that they based their answers to these two questions on somewhat different criteria (such as aesthetics versus efficiency).

Comparison Task

Accuracy on the comparison task was high (*Mean Proportion Correct* = .91, *SD* = 0.05), so our analyses instead focused on response times and eye fixations. Individual trials with response times of more than 3 standard deviations from the mean (2.9% of trials) were not included in these analyses.

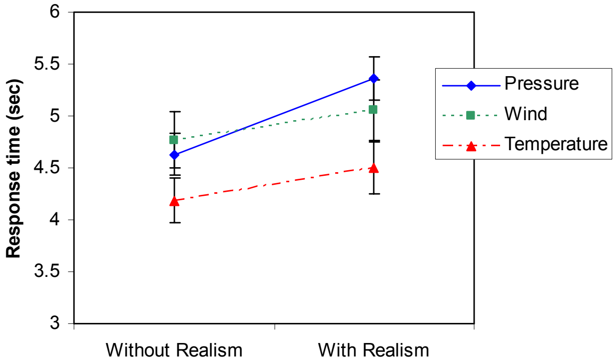


Figure 4. Mean response times for the comparison trials. (Error bars show standard errors)

Response Times. Figure 4 shows the mean response times for the different conditions in our design. There was a significant effect of realism on response time, $F(1, 25) = 24.58$, $p < .001$, $\eta_p^2 = .50$, with responses taking about a half a second longer on average on the realistic maps (4.97 seconds, *SD* = 1.17) than on the non-realistic maps (4.54 seconds, *SD* = 1.03), similar in magnitude to the "realism cost" in the earlier Canham et al. (2007) study.

Response time was also affected by the meteorological variable to be compared, $F(2, 50) = 11.37$, $p < .001$, η_p^2 = .31, with questions about pressure taking longest (*M* = 5.01 sec *SD* = 0.99) followed by wind (*M* = 4.91 sec, *SD* = 1.34) and temperature (*M* = 4.34 sec, *SD* = 1.15). The interaction between realism and the meteorological variable to be compared was not significant, $F(2, 50) = 2.07$, $p = .14$.

Eye Fixations. Figure 5 shows the fixations made by all 26 participants while solving a pressure question without (top) and with (bottom) realism. To analyze the eye fixations, we defined areas of interest on the maps corresponding to the three regions to be compared (labeled A, 1, and 2) and the text specifying the comparison to be made. Each of the 3 areas to be compared had a diameter of 90 pixels or about 3.2 degrees of visual angle and corresponded to 1.3% of the area of the display. The text (above the map, specifying the comparison to be made) was a rectangular area of 100 by 60 pixels and took up 7.5% of the screen. We derived three types of measures from the eye fixation data; total fixation duration (the sum of the durations of all fixations) on each area of interest, the number of fixations in each area of interest and the average length of fixations. Total fixation duration was highly correlated with number of fixations and showed the same pattern, so we only report number of fixations here.

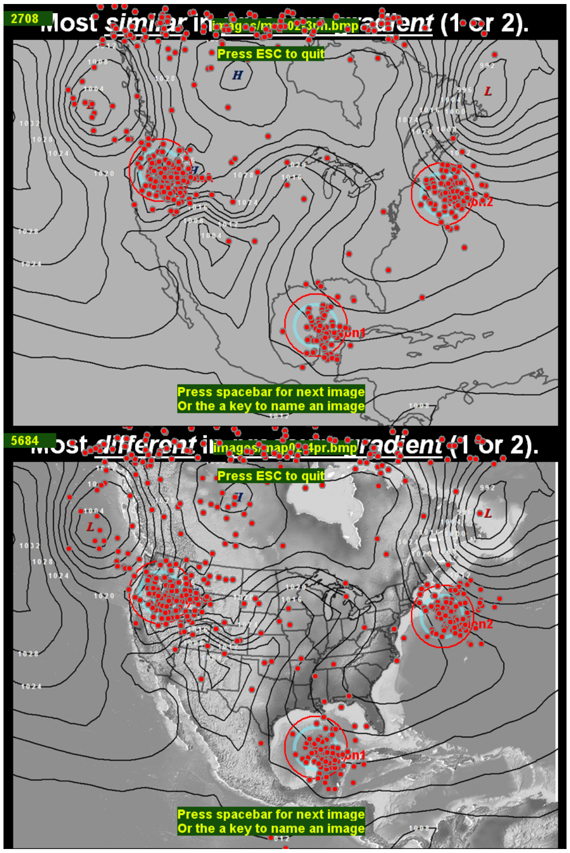


Figure 5. Plots of the fixations of all 26 participants during a pressure trial with a realistic and a non-realistic map.

The eye fixation analyses indicated that the longer response times with realistic maps reflected both longer fixations, and a greater number of fixations on realistic maps compared to non-realistic maps. Average fixation length increased from 223 msec. (*SD* = 27.6) for trials without realism to 233 msec. (*SD* = 27.6) on trials with realism ($F(1, 25) = 19.66$, $p < .001$, $\eta_p^2 = .44$).

Table 2 reports the average number of fixations in each of the areas of interest and on "other" areas of the display, that is, fixations that fell outside the regions marked A, 1, 2 or the text. The number of fixations on the region marked A was affected by the variable to be compared ($F(2, 50) = 13.10$, $p < .001$, $\eta_p^2 = .34$ and there was a significant interaction of the variable to be compared with realism, F(2, 50) = 13.10, $p < .001$, $\eta_p^2 = .34$. Analysis of simple effects indicated that participants fixated this region more often for the realistic trials compared to the non-realistic trials when the variable to be compared was pressure.

The number of fixations on the combined regions marked 1 and 2 (which the participants had to compare to A) showed a main effect of realism. Participants fixated these regions more often on realistic trials than on non-realistic trials, $F(1, 25) = 4.38$, $p < .001$, $\eta_p^2 = .15$. The variable to be compared (temperature, pressure, wind) also affected fixations to these regions ($F(2, 50) = 14.27$, $p < .001$, $\eta_p^2 = .36$. Fixations on the text were unaffected by realism, although this variable was also influenced by the variable to be compared, $F(2, 50) = 11.16$, $p < .001$, $\eta_p^2 = .31$.

Table 2. Mean Number of fixations in the different regions of interest as a function of trial type (SD in parentheses).

	Area of Interest			
	A	1&2	Text	Other
Temperature No Realism	3.24 (1.11)	4.03 (1.50)	3.87 (1.42)	3.91 (1.84)
Temperature + Realism	3.36 (1.38)	4.22 (1.74)	3.55 (1.53)	4.48 (2.31)
Pressure No Realism	3.59 (1.10)	4.47 (1.49)	4.13 (1.44)	4.39 (2.10)
Pressure + Realism	4.14 (1.10)	5.01 (1.32)	4.44 (1.47)	4.74 (2.19)
Wind No Realism	4.28 (1.44)	5.01 (1.60)	3.41 (1.04)	3.47 (2.02)
Wind + Realism	4.04 (1.48)	5.16 (1.87)	3.56 (1.15	3.96 (2.38)

Finally, there was a main effect of realism on the number of fixations made to "other regions," that is, outside the regions to be compared. These were fixated more often on realistic trials than on non-realistic trials, $F(1, 25) = 6.66$, $p < .05$, $\eta_p^2 = .21$. This result is illustrated by the example in Figure 5 which shows that whereas most fixations were clustered around the three regions to be compared, the fixations were more dispersed in the case of the more realistic map and there were more fixations in between the areas of interest. The number of fixations on "other regions" was also affected by the variable to be compared, $F(2, 50) = 3.65$, $p < .05$, $\eta_p^2 = .31$.

Discussion

The results of this experiment are consistent with our earlier research on intuitions and performance with weather maps by Navy meteorologists and undergraduate students (Smallman & Hegarty, 2007; Canham, et al., 2007). About one third of participants chose realistic maps over non-realistic maps, in spite of the fact that adding realism to the maps slowed their response times by about half a second on a simple read-off and comparison task. In addition, new data provided by eye fixations in this experiment revealed that when viewing realistic maps, participants made longer fixations, more fixations in the regions to be compared, and more fixations outside these regions.

The patterns of eye fixations allow us to speculate how the extra clutter imposed by realism on a map affects performance. Eye fixations were more dispersed around the regions to be compared on realistic maps than on the simpler non-realistic maps, suggesting that participants may need to sample from a wider area on these maps to encode the relevant patterns. Participants also had to re-fixate the relevant areas more often on realistic maps, suggesting that they may have had difficulty retaining information in working memory about the weather patterns on these maps. For example, if users tried to retain and compare visual images of regions of the maps, it is likely that the extraneous visual information on realistic maps overloaded working memory. Finally, participants' fixations were somewhat longer on realistic maps, suggesting that it may have taken more time to see the relevant information when it was partially masked by extraneous variables (cf. van Diepen & d'Ydewalle, 2003) .

While several theorists have made prescriptions for the construction of effective graphics, (Bertin, 1983; Kosslyn, 1989; Tufte, 1983) there has been little empirical validation of these principles. Our research provides new evidence for the basic principle that graphics should not provide more information than is needed by the user. Our finding that task performance is slowed by the addition of irrelevant variables to maps is also consistent with research showing that speed of visual search on maps is related to the amount visual clutter (Rosenholtz, Li, & Nakano, 2007). Similarly, Lohse (1993) found that the best predictor of response time to extract information from graphs was the number of non-target distractor objects in the task- relevant regions of the graph. It is clear that extraneous information in visual displays can hurt performance.

Despite the performance decrement from the extraneous information that they provided, participants preferred realistic maps over non-realistic maps about one third of the time. Our research validates Smallman and St. John's (2005) work on Naïve Realism and extends it to the meteorology domain. Furthermore, it shows that participants' choice of realism does not just reflect aesthetic preferences, as there was only a small increase in choice of the optimal map when participant were asked a question that focused on efficiency rather than preference. More positively, it shows that participants' choice of the optimal map increased after performing the comparison trials, suggesting that participants had some metacognitive awareness of the added difficulty produced by the realistic maps and learned to distinguish somewhat between maps that they might prefer and those with which they performed more efficiently.

The fact that some participants still choose realistic maps, even after experiencing the performance decrements with these maps, suggests that display designers should be cautious in giving users the ability to customize their own information displays. It has long been widely appreciated in the human factors domain that users' intuitions and performance can decorrelate (e.g., Andre & Wickens, 1995). But what our work highlights is just how systematic and predictable can be the metacognitive lapses responsible, predictable enough to begin to accommodate in display design. Alternative approaches might be to recommend display designs for different tasks, or to explicitly teach users principles of effective graphical displays. Without some kind of guidance, however, our research suggests users' intuitions may lead to graphical display configurations that actually impair their performance.

Of course, this research is limited in that we examined performance of naïve students in a simple read-off and comparison task. It is important to examine how display clutter affects performance in more complex naturalistic decision tasks. In current research we are examining how extraneous display information affects performance in inference tasks, for example situations in which participants have to predict future weather conditions based on knowledge of meteorological principles. We are also examining how display clutter affects performance of more expert individuals, to examine the extent to which expertise moderates the effects of clutter, if at all.

Acknowledgments

We thank Ryan Najima for help with running the experiment and Jerry Tietz for technical support. This research is funded by a grant from the Office of Naval Research.

References

Andre, A. D., & Wickens, C. D. (1995). When users want what's NOT best for them. *Ergonomics in Design*, *3*, 10-14.

Bertin, J. (1983). *Semiology of Graphics* (translated by William J. Berg). Wisconsin: University of Wisconsin Press.

Canham, M., Hegarty, M., & Smallman, H. (2007). Using complex visual displays: When users want more than is good for them. *Proceedings of the Eighth International Naturalistic Decision Making conference, Pacific Grove, CA,* June 1st -4th.

Cheng, P. C. (2002). Electrifying diagrams for learning: Principles for complex representational systems. *Cognitive Science, 26*, 685-736.

Hoffman, R., Detweiler, M., Conway, J., & Lipton, K. (1993). Some Considerations in Using Color in Meteorological Displays. *Weather and Forecasting*, *8*, 505-518.

Larkin, J., & Simon, H. (1987).Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*, 65-99.

Lohse, G. L. (1993). A cognitive model for understanding graphical perception. *Human Computer Interaction, 8*, 353-388.

Kosslyn (1989). Understanding charts & graphs. *Applied Cognitive Psychology, 3,* 185-226.

Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision, 7*, 1-22.

Smallman, H. S., Cook, M. B., Manes, D. I., & Cowen, M. B. (2007). Naïve realism in terrain appreciation. In *Proceedings of the 51st Annual Meeting of the Human Factors and Ergonomics Society* (pp. 1317 - 1321). Baltimore, MD, October 1st-5th.

Smallman, H. S., & Hegarty, M. (2007). Expertise, spatial ability and intuition in the use of complex displays. In *Proceedings of the 51st Annual Conference of the Human Factors and Ergonomics Society* (pp.200-204), Baltimore, MD, October 1st-5th.

Smallman, H. S., & St. John, M. (2005). Naïve realism: Misplaced faith in realistic displays. *Ergonomics in Design, 13*, 14-19.

Shah, P., Freedman, E. G., & Vekiri, I. (2005). The comprehension of quantitative information in graphical displays. In P. Shah & A. Miyake (Eds.) *The Cambridge Handbook of Visual Spatial Thinking*. Cambridge University Press, New York, NY.

Tufte, E. T. (1983). *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press.

van Diepen, P. M. J., & d'Ydewalle, G. (2003). Early peripheral and foveal processing in fixations during scene perception. *Visual Cognition, 10*, 79-100.