Investigation of Potential Industrial Uses for Tools Assessing Saliency and Clutter of Design Features

by

Tanya S. Goldhaber

SUBMITTED TO THE DEPARTMENT OF MECHANICAL ENGINEERING IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

BACHELOR OF SCIENCE IN MECHANICAL ENGINEERING
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2010
[June 2010]

Signature of Author: _____

Tanya S. Goldhaber
Department of Mechanical Engineering
May 7th, 2010

Certified by: _____

Daniel Frey
Professor of Mechanical Engineering and Engineering Systems
Thesis Supervisor

Accepted by: _____

John H. Lienhard V
Collins Professor of Mechanical Engineering
Chairman, Undergraduate Thesis Committee

Investigation of Potential Industrial Uses for Tools Assessing Saliency and Clutter of Design Features

by

Tanya Goldhaber

Submitted to the department of Mechanical Engineering
on May 7th, 2010 in partial fulfillment of the
requirements for the degree of Bachelor of Science in
Mechanical Engineering

ABSTRACT

As human interaction with digital displays becomes an indispensable part of everyday life, user Interface (UI) design is becoming an increasingly important field. There is a great demand in industry for tools to aid designers in UI design, and in response to this need, a perceptual tool, DesignEye, has been developed. DesignEye creates maps of saliency and clutter within an image, which can be used by designers to find problem areas in a design. The experiment described here tested how subjects differ in their analysis of existing UI designs when they have also been given access to maps from DesignEye. Subjects were asked to evaluate existing designs in Ford vehicles for three conditions: (i) while being given no assistance, (ii) while being asked to use a design technique like squinting, and (iii) while being asked to use DesignEye output. It was found that subjects did not substantially differ in their analyses when given a perceptual tool. However, due to the backgrounds of the subjects tested and the experimental setup and environment, further testing is necessary to determine how DesignEye might change the way designers analyze designs, build consensus within teams, and objectively rate potential design options.

Thesis Supervisor: Daniel Frey

Title: Robert Noyce Career Development Associate Professor of Mechanical Engineering and Engineering Systems

ACKNOWLEDGEMENTS

The entirety of the research described here was carried out in the MIT Brain and Cognitive Science department in the lab of Professor Edward Adelson and Dr. Ruth Rosenholtz, and the author would like to thank Dr. Rosenholtz for her excellent supervision and mentorship during the course of this project. In addition, Alvin Raj helped with software development and image analysis, which was invaluable in preparing the experiment. The author would also like to thank Professor Daniel Frey, her thesis supervisor, for his guidance and feedback.

This research would not have been possible without support from Ford, who provided the images and user feedback used in the experiments and analysis. Dr. Dev Kochhar, from Ford Research & Advanced Engineering, provided invaluable information and materials, as well as helping to define the goals of the project.

Finally, the author would like to thank her friends and colleagues who gave feedback on experimental design or volunteered to be experimental subjects.

BIOGRAPHY

Tanya Goldhaber is class of 2010 at the Massachusetts Institute of Technology. She majored in Mechanical Engineering and had minors in Cognitive Science and Music. She was a participant in the pilot year of the Gordon-MIT Engineering Leadership (GEL) Program, and spent the summer after her junior year in England working for British Telecom (BT) in collaboration with the GEL program. At BT she worked on UI design, an experience that inspired her pursuit of visual cognition and design as a thesis topic. As the recipient of a 2010 Marshall Scholarship, Tanya will return to England in the fall of 2010 to pursue a PhD at the Engineering Design Centre at the University of Cambridge.

While at MIT, Tanya participated in a UROP at the Kanwisher Lab in the Brain and Cognitive Science (BCS) Department as part of her Cognitive Science minor. This experience fostered her love of psychology and cognitive science, a discipline, which she hopes to apply to engineering design. The work on this thesis was done as a collaboration between the MIT Mechanical Engineering and BCS Departments, a fascinating combination that Tanya plans to explore more in her PhD research and in her career.

Tanya is also an avid violinist and has been very active in the music department at MIT as an Emerson Scholar and member of both the MIT Symphony and the MIT Chamber Music Society. She has won multiple awards for her performances at MIT, including winning the MIT Concerto Competition in 2009. She plans to continue studying and performing on violin next year in Cambridge, UK.

TABLE OF CONTENTS

LIST OF FIGURES

Investigation of Potential Industrial Uses for Tools Assessing Saliency and Clutter of Design Features

1 Introduction

Whenever humans must interact with a product, the design of the User Interface (UI) determines the ease with which the product can be used. Electronic digital-display devices are in increasingly widespread use, but with their propagation comes an increase in variability and a resulting increase in probability for user confusion or error. Good visual UI design is therefore becoming an area of increasing study. While user testing often illuminates problematic parts of a UI, it is more difficult to come up with basic principles to guide UI design best practices.

UIs should be designed to minimize visual search time. In other words, important or commonly used components should be easily visible and should draw the eye more than parts that are not as critical to functionality. However, as anyone who has ever had trouble finding something on, for example, a website can attest, designers are not always adept at making these important features easy to find.

A key issue tends to be information overload. Although there are no longer many technical limitations on the amount of information that can be presented in high quality on something like a GPS display, the questions remains as to how much of this information *should* be presented. When someone is driving, do they need a detailed map or just a basic description of the route? How much functionality should be available when the car is moving? How much visual information is even available in a driver's peripheral vision? With more and more information available that designers can choose to display, and more and more functionality demanded by users, the issue of good visual design is increasingly complicated. There is a demand among companies that deal in visual design, like car companies whose cars come with GPS units, for example, for a tool or set of tools that can help designers understand what features will be useful and what features will be clutter or visual noise.

Two important features to consider when designing any component of a UI are saliency and clutter. "Saliency" refers to how much a component stands out from its surroundings, or how much it draws the eye relative to other components [Rosenholtz 1999; Rosenholtz 2001ab; Zhaoping, 2002; Rosenholtz et al, 2004]. The term can be applied to an object in almost any circumstance, and depends on properties of both the object and the surroundings. A ketchup stain is salient on the white shirt of the person next to you, but probably not on the white shirt of a person across the room from you. Within a UI, particularly one with which the user is not overly familiar, salient components draw the eye first, meaning they are the easiest components to find. Traditionally, saliency has been studied behaviorally using visual search tasks or eye-tracking apparatuses. "Clutter" in a display has to do with the density of relevant visual information in a particular region. Too much similar visual information in a region, for example, high-density text, leads to high clutter, which in turn correlates with difficulty in visual search [Rosenholtz et al 2005; Rosenholtz, Li, and Nakano 2007].

While expert UI designers are generally familiar with the concepts of saliency and clutter and their importance to good design, there are no standard tools to help them assess saliency and clutter in a design. Although saliency models do exist, most designers

either do not make use of them or do not know about them in the first place, and they are not packaged as a design tool. Currently, techniques like squinting at a design are taught as methods to help determine salient components of a design, but these methods have variable results. Rosenholtz et al. (2010) introduced a tool, "DesignEye," that can decompose an image into "maps" of saliency and clutter across the design, with image brightness at any point on the map corresponding to either high saliency or high clutter in that region. While this tool has been shown to accurately predict eye movements when looking at a design, the tool has not been tested for usefulness in an industrial design setting.

Rosenholtz et al. (2010) conducted interviews and observations with groups of industry professionals interested in DesignEye to gauge the interest in such a tool. The three main categories of interest were general design guidance, assigning objective ratings to design concepts in order to be able to assign economic value to those concepts, and decreasing the need for user studies, which are often expensive and time-consuming. It was also indicated that a tool like DesignEye could be used to build consensus more quickly within design teams. These initial studies also showed that the tool has promise in terms of getting designers to talk about design goals and introspect on design, in providing a "common language" for collaboration, and in finding a more objective way to compare and contrast designs. The study described below is in response to the obvious need to follow up on these initial observations with a more controlled, quantitative user study in order to gain insight into how the tool might actually be used.

It is not clear that DesignEye in its current form is able to aid industrial professionals in the ways they seek. Rosenholtz et al. conducted surveys with experimental users of DesignEye, and found that 75% of users found the tool potentially useful. There is a substantial gap, however, between an opinion of potential usefulness and a tangible demonstration of actual usefulness. The goal of this study is to determine how assessments of designs differ when subjects are given access to DesignEye. The experiment described in this paper aimed to test the usefulness and accuracy of DesignEye by seeing if subjects with varying amounts of design experience could better predict user feedback on automotive GPS designs if given this tool.

2 Methods

This experiment aimed at seeing how subjects differed in their assessments of automotive designs if given access to the DesignEye tool. Fifty designs of dashboards, control panels, and GPS systems currently in use in Ford vehicles were obtained from Ford, 22 of which were chosen to be presented to subjects in the experiment. Designs were first grouped into categories based on the function of the display (e.g. map, weather, radio, etc.), and then a subset of designs in each category was selected based on similarity across categories (e.g. daytime vs. nighttime displays) to comprise the final set of experimental stimuli. A full set of experimental stimuli is available in the appendix.

GPS systems in particular are an excellent UI to test, as they are often used "at a glance" and while the user is engaged in other activities. Increasing visual search time for GPS systems is not only frustrating for users, but also potentially dangerous, as many

people use a GPS while simultaneously driving. It is therefore critical to minimize visual search time in GPS units.

2.1 Experimental Conditions

Subjects were tested under one of three conditions, the "control" condition, the "no-tool" condition, and the "tool" condition. This was done to see if subjects thought about and subsequently rated the designs systematically differently when given access to DesignEye. Since each subject saw every stimulus, they could only be tested under one condition, so that responses for each stimulus in the three conditions could be compared across stimuli.

The three conditions are described below, followed by the formal instructions given to subjects in each condition.

1) Control Condition (Condition 1): Subjects were simply asked for feedback on the design, specifically for features they thought might make the design easy or hard to use. They were only provided with a picture of the design.
2) No-Tool Condition (Condition 2): Subjects were asked to use any tools or methods for design evaluation that they had been taught. Subjects were provided with an explanation of the "squinting" technique, which is taught in design classes and supposedly simulates early visual processing, so that it can be seen which contrasts are readily apparent, and told to use that method or a similar method to help them evaluate the designs. Subjects were only provided with a picture of the design.
3) Tool Condition (Condition 3): Subjects were provided with descriptions of saliency and clutter, along with saliency and clutter maps for each design. They were asked to use the maps to help them evaluate each design.

Instructions:

Condition 1

1) Look at the image of the design.
2) Determine features that you think are good about the design and write those down in the "Like" box.
3) Also determine features of the design that you think are bad or that need to be improved. Write these down in the "Don't Like" box.
4) Choose an overall rating for the design from the drop-down menu, 1 being very poor and 10 being excellent.
5) Even if you have seen a similar design, please make comments as thorough as possible, even if that means entering comments you have already made.
6) When you have done all of these things, press "Next" to continue. IMPORTANT: If you already are familiar with a design and/or the user feedback for the design, press "Know Feedback" instead to continue
7) Ask the experimenter if you have questions.

Condition 2

1) Look at the image of the design.
2) Use any design-evaluation techniques you know, such as squinting, to determine features that you think are good about the design and write these down in the "Like" box.
   -The squint test is a technique that simulates early visual processing so you can see whether the contrasts you've tried to establish are readily apparent. Close one eye and squint the other to disrupt your focus. Whatever distinctions you can still make out will be visible "at a glance."
3) Use these same design techniques to determine features of the design that you think are bad or that need to be improved. White these down in the "Don't Like" box.
4) Choose an overall rating for the design from the drop-down menu, 1 being very poor and 10 being excellent.
5) Even if you have seen a similar design, please make comments as thorough as possible, even if that means entering comments you have already made.
6) When you have done all of these things, press "Next" to continue. IMPORTANT: If you already are familiar with a design and/or the user feedback for the design, press "Know Feedback" instead to continue.
7) Ask the experimenter if you have questions.

Condition 3

1) Look at the original image on the left and its corresponding saliency and clutter maps on the right.
   -The saliency maps represent the saliency of each region of the design such that higher luminance means higher saliency. Higher saliency of a target correlates well with ease of searching for that target and is believed to be related to that target's ability to draw attention.
   -The clutter maps indicate the level of clutter for each region of the display. Clutter has to do with the density of relevant visual information in a particular region. A brighter region on the map corresponds to more clutter, and higher clutter values in turn correlate with difficulty in visual search.
2) Use these maps to determine features that you think are good about the design and write those down in the "Like" box.
3) Also use the maps to determine features of the design that you think are bad or that need to be improved, Write these down in the "Don't Like" box.
4) Choose an overall rating for the design from the drop-down menu, 1 being very poor and 10 being excellent.
5) Even if you have seen a similar design, please make comments as thorough as possible, even if that means entering comments you have already made.
6) When you have done all of these things, press "Next" to continue. IMPORTANT: If you already are familiar with a design and/or the user feedback for the design, press "Know Feedback" instead to continue.
7) Ask the experimenter if you have questions.

Each subject was tested on the same computer in a closed room in the MIT Brain and Cognitive Sciences building. Subjects were paid $10 for their participation in the experiment.

2.2 Controlling Across Subjects

Subjects were asked to fill out a questionnaire to determine both their real-world driving experience and their experience with product or UI design. Subjects were balanced across conditions for the following characteristics:

- Significant experience driving a Ford vehicle
- Driving frequently or having driven frequently in the past
- Product design experience
- UI design experience
- Past or present study of visual cognition

All subjects had normal or corrected-to-normal vision. Six men and nine women participated in the study, ranging in age from 18 to 58 years. The mean age was 27 overall (standard deviation 12 years), with mean ages 24.8, 26, and 30.4 for Conditions 1, 2, and 3 respectively. There were three women in Condition 1, four in Condition 2, and two in Condition 3.

2.3 Experimental Design

The GUIDE tool in MATLAB was used to create a basic GUI to present the stimuli to subjects. Per stimulus, all subjects were presented with a picture of the design with boxes on the bottom of the screen for them to enter what they felt were good and bad features of the design, along with an overall rating (1-10 scale) of the design. Subjects were also asked to indicate if they had seen or interacted with the design before in an actual vehicle. In the control and no-tool cases, subjects were only presented with a picture of the design. In the tool condition, however, subjects were additionally presented with saliency and clutter maps next to the design.

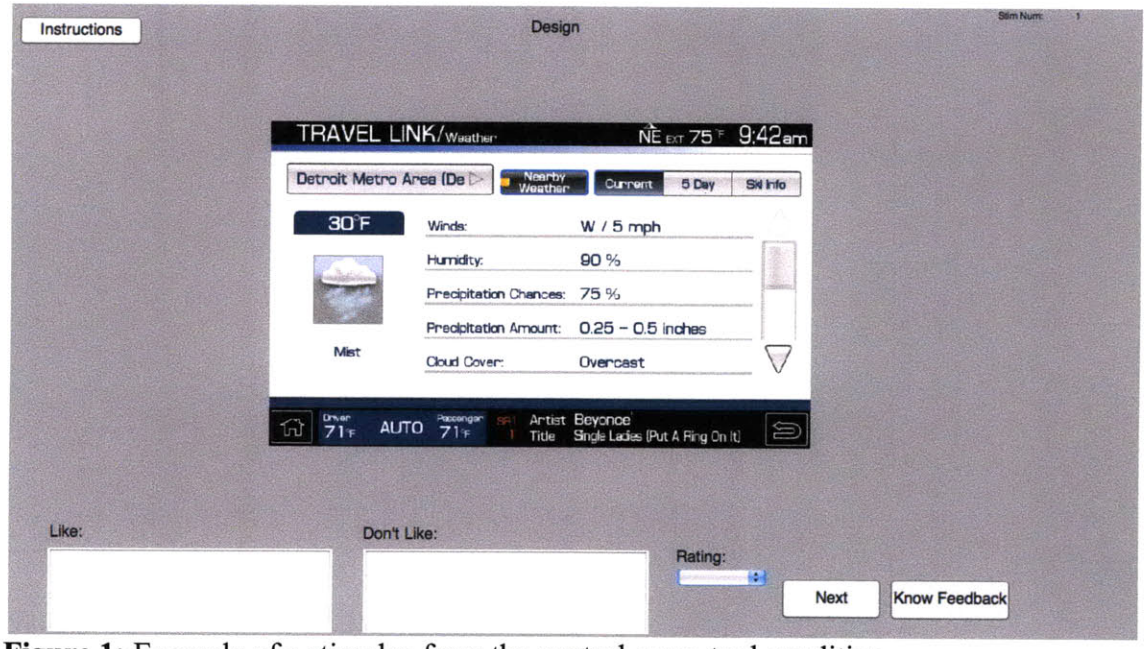


Figure 1: Example of a stimulus from the control or no-tool condition.

Figure 2: Example of a stimulus from the tool condition.

Saliency and clutter maps were generated for the tool condition by first normalizing all design images so that the GPS screen was approximately 500 pixels across. This was done because for images at too high a resolution, DesignEye's sensitivity to very high spatial frequencies gave an inaccurate map of what was actually salient or cluttered in the design (although future versions of DesignEye will be able to automatically adjust image size). In addition, this effect was negated by putting images through a Gaussian blur with

radius of one pixel. These images were then used to generate saliency and clutter maps in DesignEye. The saliency and clutter maps were normalized so that relative brightness was preserved across images and no image became saturated such that details of saliency and clutter could not be distinguished. These normalized saliency and clutter maps were presented alongside the original-resolution images in the tool condition.

The GUI permitted subjects to access instructions for the experiment at any time, as shown in Figure 3. For the no-tool condition, subjects were given a detailed description of the squinting method. For the tool condition, subjects were given a detailed description of the concepts of saliency and clutter (see section 2.1).

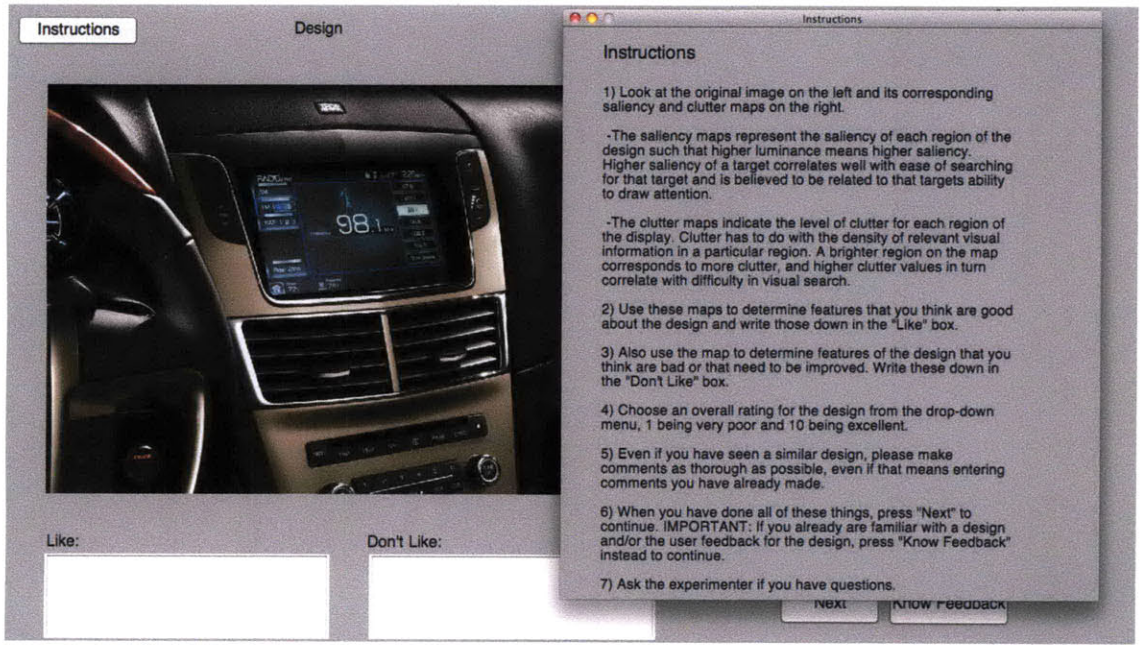


Figure 3: Stimulus while subject is viewing instructions.

All images were categorized before being put into the experiment (e.g. maps, information, etc.). Most categories only had one or two images, but one category had five images. Each condition in the experiment therefore had five subjects, all of whom saw these five images in a different order. The order was determined by Latin square design. Ordering of all other stimuli was randomized for each experiment.

2.4 Data Categorization

In addition to the experiment described in section 2.3, hereafter referred to as the "rating experiment," another experiment was run to code those subjects' responses in order to understand if and how subjects' responses were different across the three conditions. This second experiment will be referred to as the "categorization experiment." All the comments from the rating experiment were put into a spreadsheet, where no indication was given as to which design they had seen or what the features of that design were. The numerical rating was also removed. The comments were then investigated to

determine categories that the comments might fit into. Twenty categories were determined:

- References a specific design goal
- References saliency of a design feature
- References clutter in design
- Expresses that the design causes confusion
- References aesthetic features or personal preferences
- Talks about usefulness of information represented
- Expresses a concern about safety or safe use while driving
- References practicality or usefulness of design in general
- References visual search
- References specific design features (general)
- Says there is too much information or too many options
- Says there is not enough information or too few options
- References ease of use
- References use of space
- References text size or ease of reading
- References overall display or design quality
- References usefulness of color or color contrast
- References feature sizes
- References visual interference (e.g. icons covering one another up)
- References alignment/spacing of visual information

Subjects were then asked to look at a random subset of comments from the rating experiment and mark which category or categories the comments fit into for each response. For example, the comment "the display is cluttered because the icons overlap" would be marked as fitting into the categories "References clutter in design," "References visual interference," and "References alignment/spacing of visual information." Subjects were only told to mark which categories were referenced in a comment set for a stimulus, not how many times per comment each category was referenced.

Eight subjects participated in the categorization experiment. None of these subjects had participated in the rating experiment or seen the designs that the rating experiment participants had seen.

Once categorization results were obtained, they were looked over by the experimenter to check for understanding of the categories and to make sure there were no systematic omissions. Obvious systematic omissions or mistakes were corrected.

3. Results and Discussion

3.1 Analysis of comments-per-stimulus

As per the experimental design for the rating experiment, responses were separated into the categories "Like" and "Dislike," for features the subject thought were either useful or detrimental to the design respectively. The number of "Like" and "Don't Like" comments were tallied for each stimulus. Within comments, each statement was counted separately. For example, "I like that the play button is salient and that the album cover is easy to see" would be counted as two "like" comments, but "I like that the play button and album cover are easy to see" would be counted as one. In addition, some comments contained suggestions for how the design could be improved. The number of suggestions per stimulus was also counted.

The following figure shows the distribution of responses for each condition in the rating experiment:

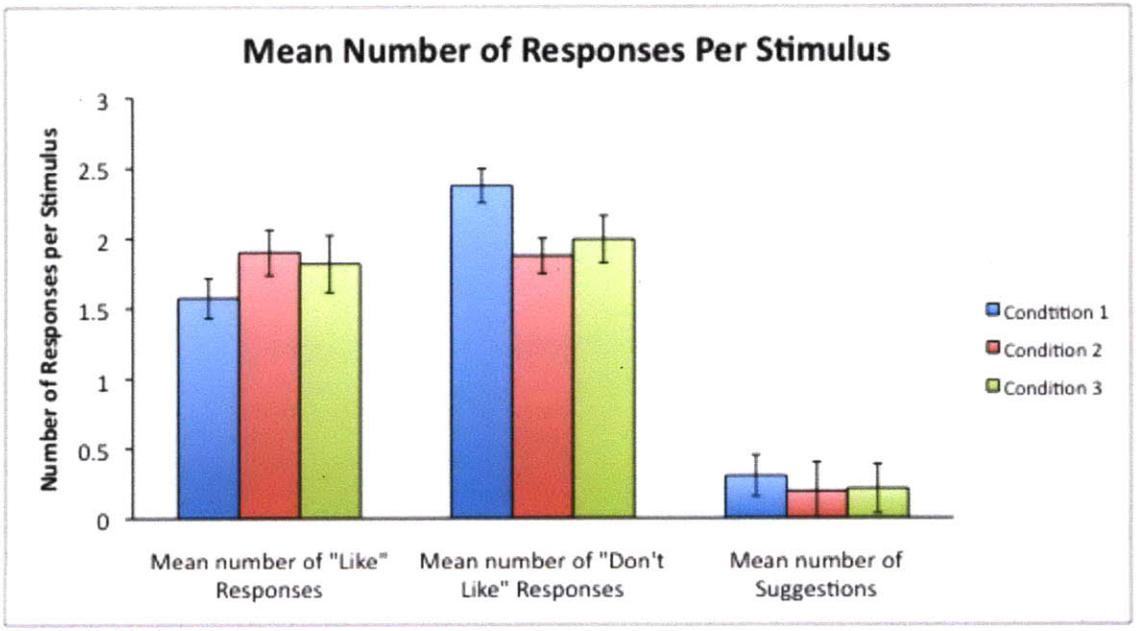


Figure 4: Distribution of the mean number of "Like" and "Don't Like" responses per stimulus, and mean number of suggestions per stimulus.

As is evident in Figure 4, there is not a significant variation in the mean number of "Like," "Don't Like," and suggestion responses across conditions. The only possibly significant trend is that the mean number of "Like" responses for Condition 1 is slightly lower than the other two conditions and that the mean number of "Don't Like" responses for Condition 1 is significantly higher. For all types of responses, there is no significant difference between Conditions 2 and 3, indicating that the introduction of some sort of anchor for assessing designs, whether it be a technique or a tool, does not significantly increase or decrease the number of responses subjects give.

3.2 Analysis of Response Categorization

Although giving subjects a perceptual tool to analyze designs did not seem to increase the number of comments they make, it is possible that it will affect the *kinds* of comments they make. As described in section 2.4, comments were categorized across 20 different categories to see what type of feedback subjects gave in all three conditions. Once the comments were divided into categories, the number of responses per category was summed separately for all three conditions.

Figure 5 shows the number of responses per category for the three conditions.

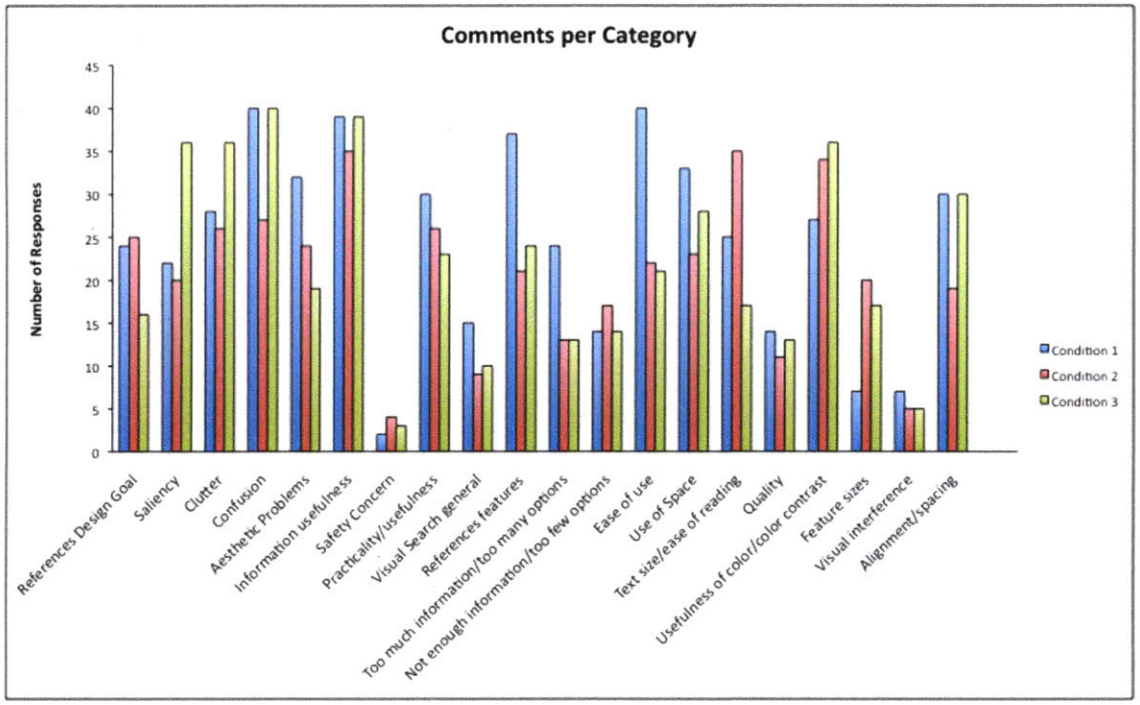


Figure 5: Total number of comments made per category in each condition.

A close investigation of Figure 5 yields the unsurprising observation that the number of comments that referenced saliency and clutter in the design significantly increased in Condition 3. This is not unexpected, as subjects in Condition 3 were given both instructions that specifically mentioned the terms saliency and clutter and a perceptual tool that pulled saliency and clutter information out of an image. The fact that references to saliency and clutter increased in Condition 3 does increase the confidence that most subjects were at least paying attention to the perceptual tool in the condition.

What is perhaps somewhat surprising is the fact that in no other category does Condition 3 really have a much higher number of comments than for *both* Categories 1 and 2. For example, in the "Alignment/Spacing" category (rightmost in Figure 5), Condition 3 has many more comments than Condition 2, but not a significant number

more than Condition 1. Similarly, for the "Usefulness of color/color contrast" category, Condition 3 stands out against Condition 1 but not Condition 2. The most common trend is for Conditions 1 and 3 to have the most comments, and for Condition 2 to have fewer.

What is surprising about this result is that Condition 1, where subjects had neither a technique nor a tool to aid design analysis, had just as many categories where it had the most comments as the other two conditions, and in fact stood out in a few categories as clearly having many more comments than the other two conditions. This goes counter to the hope that a perceptual tool could aid designers in improving UIs like the GPS systems studied in this experiment. Indeed, it is unclear from these results if the tool provided even a marginal benefit above the control conditions.

There are some other factors, however, that can potentially explain why subjects did not noticeably improve in Condition 3. One thing worth noting is that all comments were categorized by naïve third parties, most of whom had no experience in design, and none of whom had seen the designs or were familiar with the kinds of results that might be anticipated. While this was a good way to control for bias in data analysis, it is possible that a lack of clarity in instructions or variety in the way comments were categorized could call the analysis presented above into question. The obvious next step for future data analysis is to be able to confirm that raters have completely understood the instructions and what defines each category, and then have several participants rate each subsection of data to ensure that responses match up. It is also possible to have an experimenter rate the data, although this could lead to bias in data analysis.

An additional factor is that only five subjects participated per condition. This relatively low number of subjects might make it difficult to pick significant trends out of the data. Even doubling the number of subjects per condition might lead to a clearer picture of what kinds of comment changes subjects made in the different conditions.

Finally, it should be mentioned that the majority of the subjects (8 of 15) were students at MIT. Additionally, two subjects who were not MIT students had formally studied psychology, and three non-student subjects were industry professionals with some amount of design experience. For the purposes of studying systematic analytical behavior change due to the introduction of a tool, the population that participated in the current experiment was not really the ideal population to study, both lacking the necessary homogeneity and largely not being the target population of DesignEye. It is critical that in future experiments, DesignEye be tested in a similar fashion both on professional designers in a more realistic design setting and on completely naïve subjects with no design, psychology, or industry experience.

3.3 Analysis of Numerical Ratings

While no systematic differences appear among the different conditions while looking at comment categorization, it is plausible that the numeric ratings given to designs might show a different or more enlightening pattern. For each condition, each rating where the subject was familiar with the design was removed from analysis, and the ratings for each stimulus were then averaged. Figure 6 shows the average ratings per condition for each stimulus.

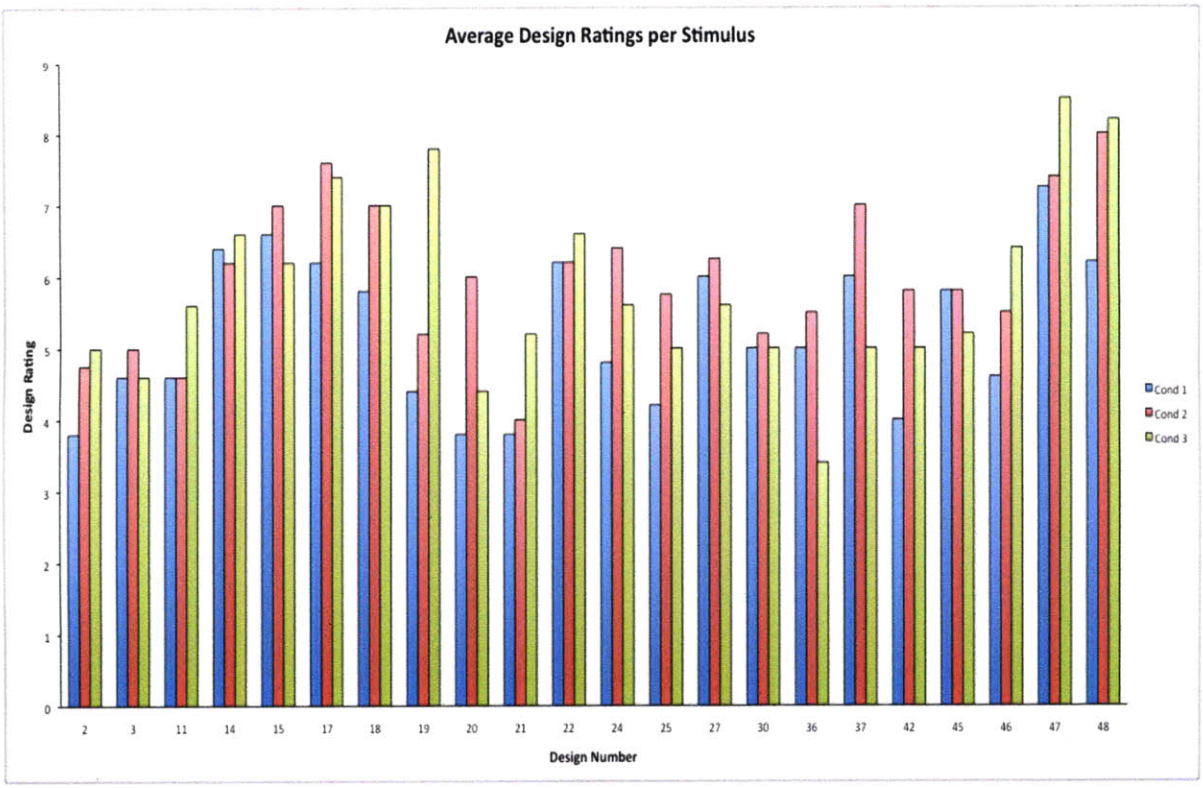


Figure 6: Averages of numeric assessments of design quality. Designs are shown and labeled by number in the appendix.

It appears from Figure 6 that there is not a consistent difference among conditions in terms of the numeric ratings, nor is the difference even very large for the majority of designs. This observation is made even clearer by the introduction of error bars (for a 95% confidence interval), as shown in Figure 7.

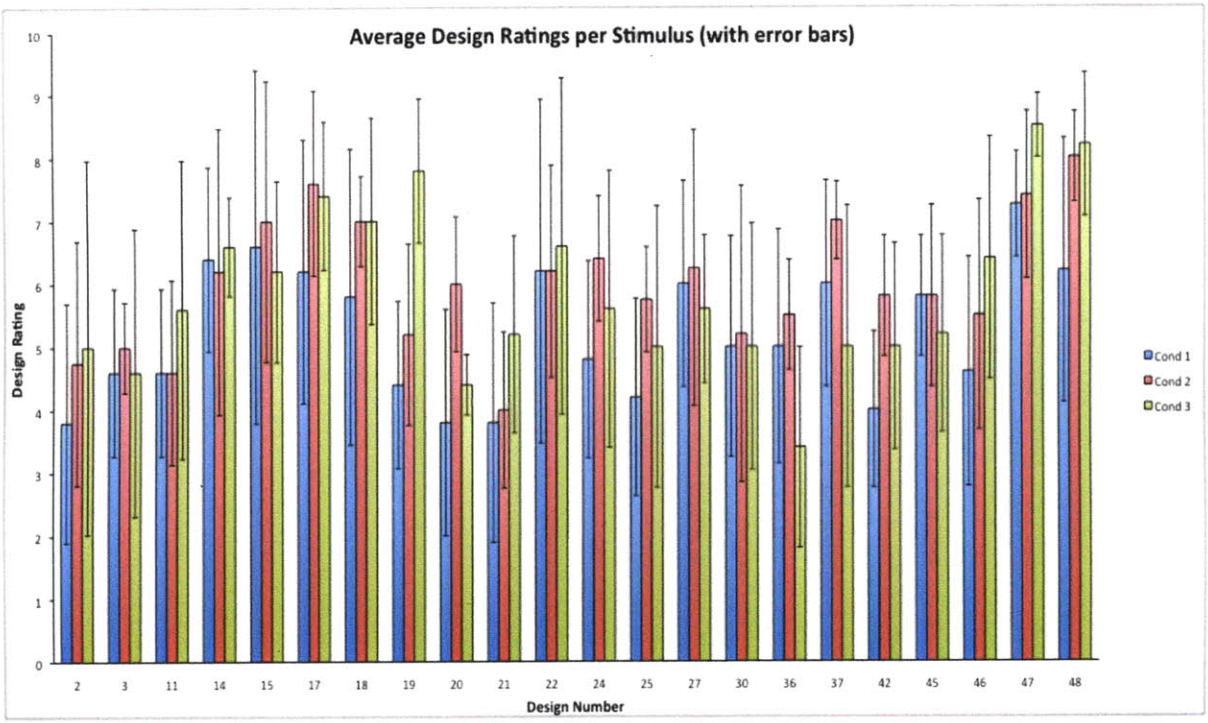


Figure 7: Averages of numeric assessments of design quality, with error bars shown.

As is apparent in Figure 7, only Image 19 has any significant difference among the three conditions. One problem is that the number of subjects was relatively small (only five subjects per condition), while the standard deviation in rating per stimulus was relatively large, ranging on average between 1.5 and 2, which is significant as the rating was out of ten. It is therefore hard to say with any confidence for most of the stimuli that any one condition led to a drastic difference among ratings.

Although the error analysis shown in Figure 7 would indicate that there was no significant difference between conditions for most of the stimuli, the figure nonetheless suggests that for many of the images, the introduction of a tool does make the rating significantly different that in the control and no-tool conditions. This is most apparent for designs 19, 46, and 47, where the tool condition had a much higher rating, and designs 36 and 37, where the tool condition had a much lower rating. The error in this analysis is so high because the number of subjects is so small (n = 5 per condition). A necessary follow-up experiment is to run a larger number of subjects on the same designs with the same procedure and re-analyze the number ratings. In order to halve the error, it would be necessary to run 20 subjects on each design. Therefore, an alterative experiment might be to present the images in the same manner, but collect only ratings and not comments, which would drastically cut down the experiment time and make it feasible to run that many subjects.

As mentioned earlier, one goal of DesignEye is to promote quick consensus-building within teams. Therefore, the average standard deviation of responses per stimulus was calculated for each condition, as shown in Figure 8.

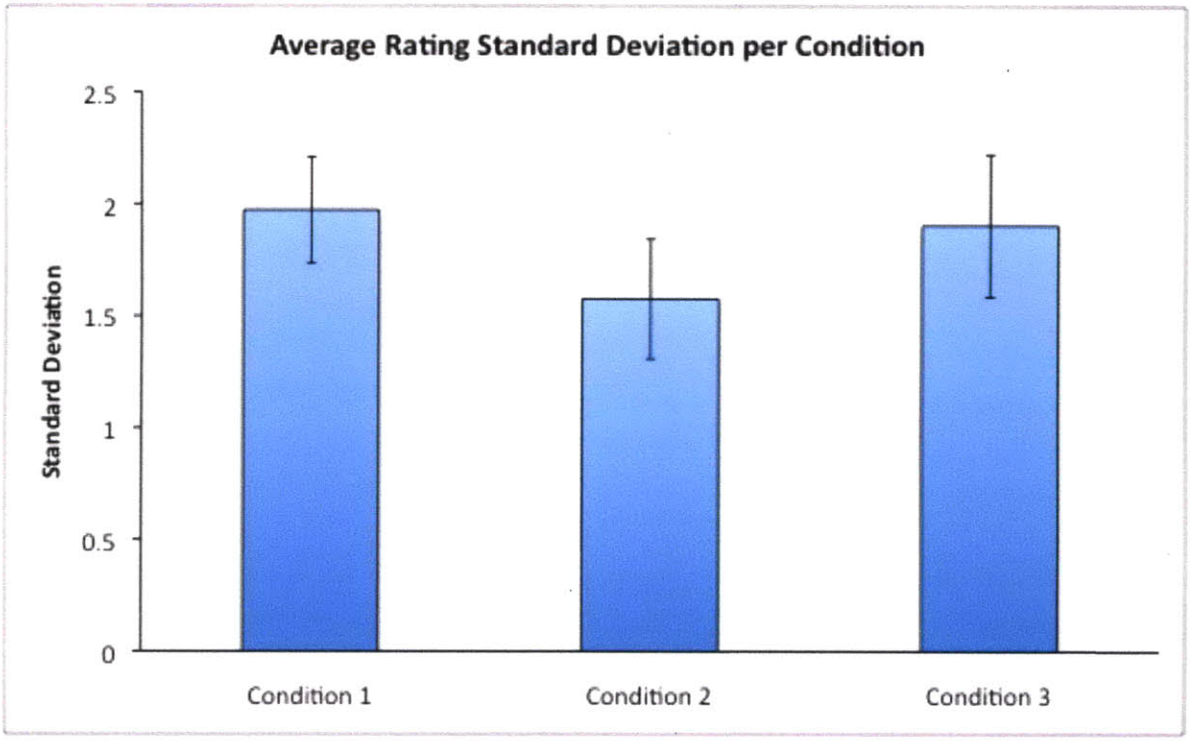


Figure 8: Average standard deviation for numeric responses for all three conditions.

As can be seen above, Condition 3, where DesignEye was used as a tool in assessing designs, did not lead to more "consensus" among subjects, as would have been indicated by a decrease in average standard deviation in Condition 3. Therefore, it cannot be assumed that the introduction of this tool unifies opinions of a design in and of itself. It is possible that the tool could be used to build consensus, but probably in the context of something like team meetings or design reviews.

3.4 General Discussion

It does not appear from initial experimental results that DesignEye makes a noticeable difference in terms of how subjects rate or analyze designs. However, this does not necessarily mean that DesignEye is not a useful tool, or cannot potentially be useful in an industrial context. If anything, these experiments provide several directions for future work. Since industrial contacts have expressed a desire for a perceptual tool such as DesignEye, it is critical to continue exploring possible uses and use cases for the tool so that it can be used most effectively.

One potential problem with the rating experiment was that subjects were not told *how* different distributions of saliency and clutter affect visual processing. Nor were subjects given any general principles for good and bad visual design and/or use of salient or cluttered information. Presumably designers in a real-world design situation would have access to all of this information and would be encouraged to use it. The next step in

assessing DesignEye is to utilize it in a more realistic design setting, testing actual designers and not students, and giving access to the kinds of information that real designers typically have. Additionally, each subject was only tested in one condition. An interesting follow-up would be to have subjects rate some subset of the stimuli in all three conditions and observe how their comments and ratings changed over the course of the experiment.

Finally, it is crucial to mention the initial experimental concept on which this experiment is based, and how critical it is that this experiment is eventually carried out. Initially, Ford was going to provide actual user feedback on the designs given to subjects in the rating experiment. Subjects were going to be asked to predict how users felt about the design, and the subjects' predictions would then be compared to the actual user feedback. However, until sufficient user feedback is obtained from Ford, the experimental procedures described above, with the modifications suggested, are the best option for acquiring the kind of data necessary to determine how DesignEye can be most effectively used in an industrial setting.

It should also be noted that, regardless of experimental results, the GUI used in the above experiments will serve as a platform for future experiments. The development of this interface for running experiments to test the use of DesignEye is an important step forward in reaching conclusions about the eventual industrial use of this tool.

4. Conclusions

The purpose of the above experiments was to explore how DesignEye could be used to help industrial UI designers more quickly, efficiently, and reliably create excellent interfaces for a wide variety of projects and applications. The experiments described here could not show that DesignEye significantly changed how people analyze UI designs both in terms of the kinds of comments they make about the designs and how they rate the designs. However, modifications to participant population, the experimental conditions and procedures, and the way the data is analyzed will quite likely lead to different results. In addition, an experiment testing if DesignEye can allow designers to more accurately predict user feedback on a design has the potential to be very enlightening.

A number of industry professionals from a variety of fields have indicated that a perceptual tool like DesignEye would be very useful in their respective design processes, so it is critical that it is understood how DesignEye is used, and how it could be modified to be used most effectively by designers. Follow-up experiments to those described above have the potential to suggest the kinds of changes to DesignEye that will eventually shape it into an essential industry tool.

References

ROSENHOLTZ, R. 1999. A simple saliency model predicts a number of motion popout phenomena. *Vision Research, 39,* 3157-3163.

ROSENHOLTZ, R, 2001a. Visual search for orientation among heterogeneous distractors: Experimental results and implications for signal detection theory models of search. *J. Experimental Psychology, 27*(4), 985-999.

ROSENHOLTZ, R. 2001b. Search asymmetries? What search asymmetries? *Perception & Psychophysics 63*(3): 476-489.

ROSENHOLTZ, R., NAGY, A.L. & BELL, N.R. 2004, The effect of background color on asymmetries in color search. *Journal of Vision, 4*(3), 224-240.

ROSENHOLTZ, R., & JIN, Z. 2005. A computational form of the statistical saliency model for visual search [Abstract]. *Journal of Vision 5*(8), 777a.

ROSENHOLTZ, R., LI, Y., MANSFIELD, J., & JIN, Z. 2005. Feature congestion, a measure of display clutter. *Proc. SIGCHI*, 761–770, 2005

ROSENHOLTZ, R., LI, Y., & NAKANO, L. 2007. Measuring visual clutter. *Journal of Vision 7*(2):17, 1-22. http://journalofvision.org/7/2/17/, doi:10.1167/7.2.17.

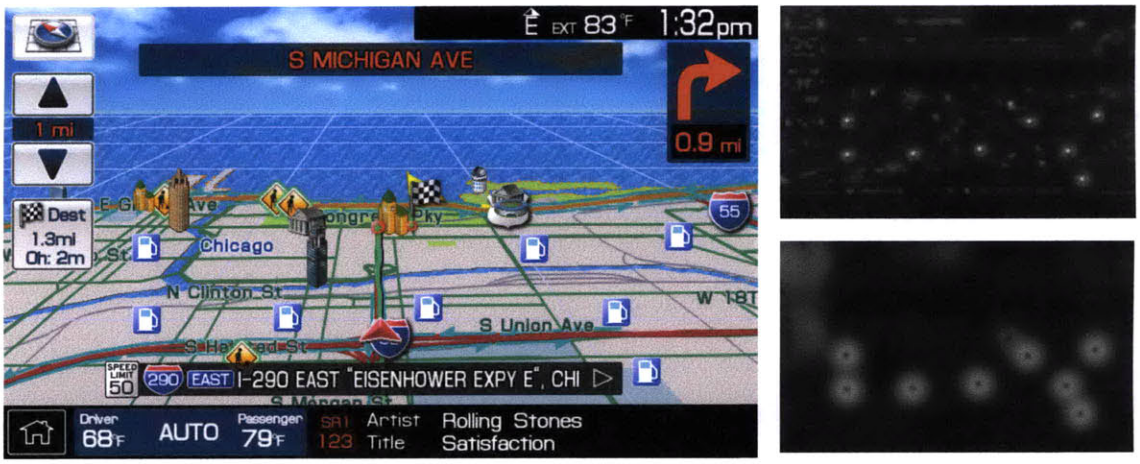
ROSENHOLTZ, R., DORAI, A., & FREEMAN, R. 2010. Do Predictions of Visual Perception Aid Design? *Transactions on Applied Perception* (in press).

ZHAOPING, L. 2002. A saliency map in primary visual cortex. *Trends in Cognitive Sciences, 6*(1), 9-16.

Appendix

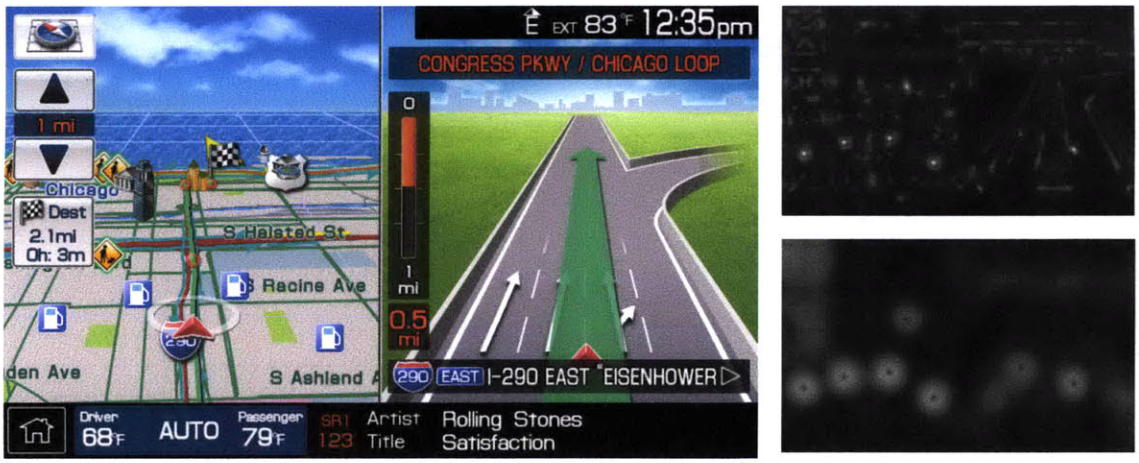
The following is a list of all the stimuli that appeared in the experiment along with corresponding image number, saliency map (top), clutter map (bottom), and sample responses from each condition. All images were taken from GPS systems in Ford vehicles.

Image 2:



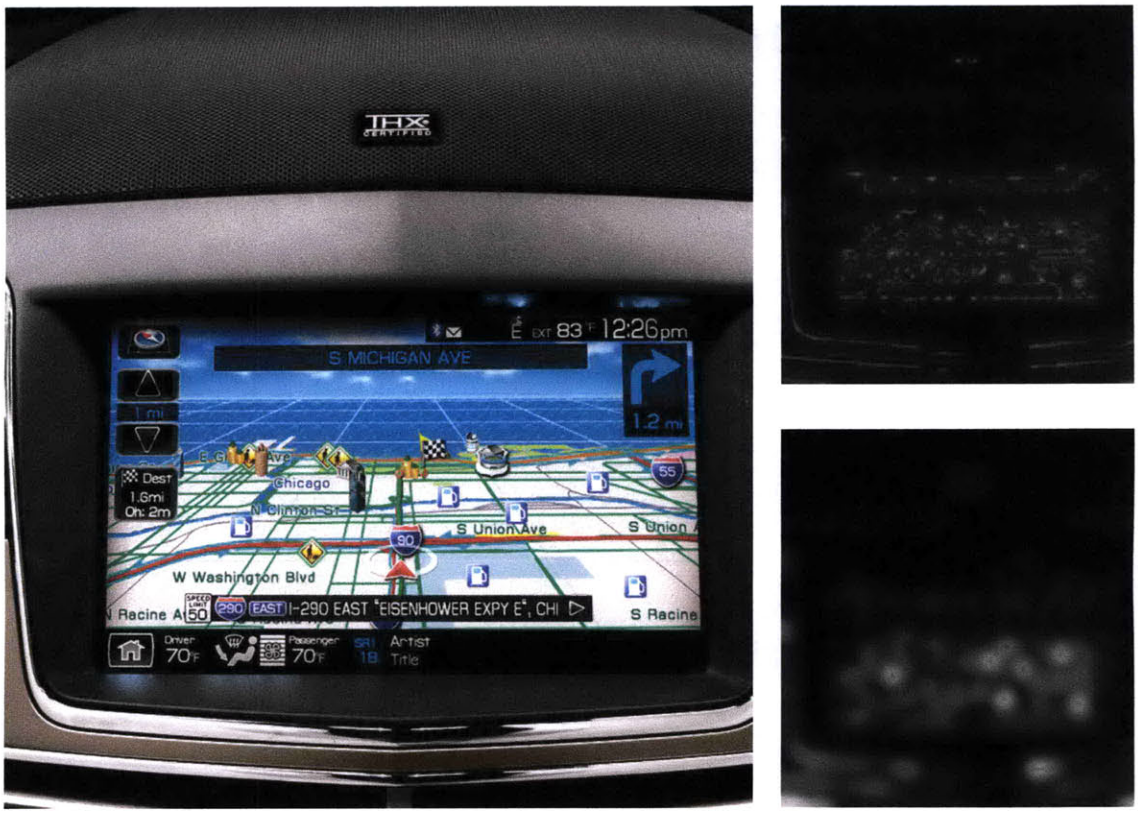
Condition 1: "If I'm not interested in gas stations, I would like to be able to remove those"
Condition 2: "Too much info on the screen at once."
Condition 3: "Too many icons make it hard to see the whole map."

Image 3:



Condition 1: "What is the relationship between the two pictures?"
Condition 2: "Left screen is TERRIBLE. WAY too cluttered."
Condition 3: "There are too many icons on the left hand side of the screen."

Image 11:



Condition 1: "Like that traffic is displayed on map."
Condition 2: "Not sure what road I'm on."
Condition 3: "Too many icons block the details of the map and could be simplified."

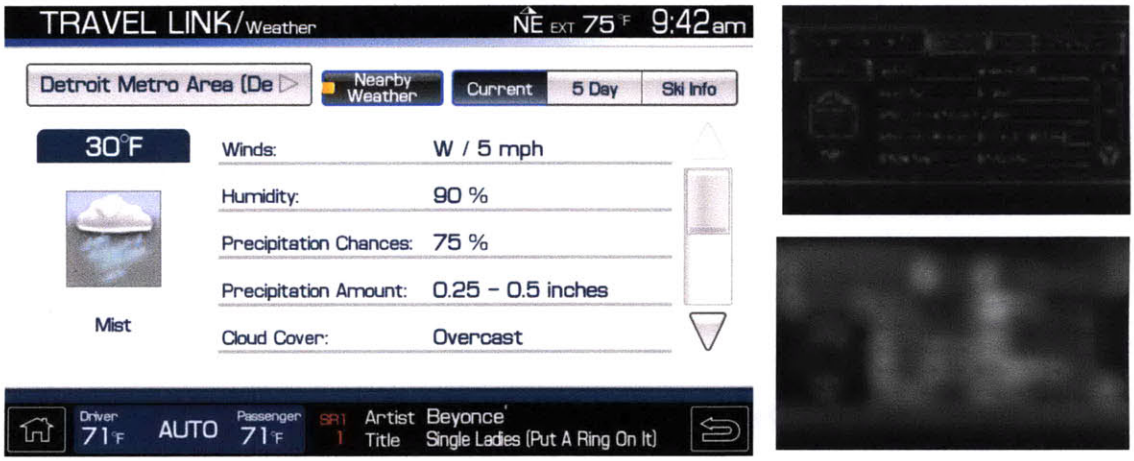
Image 14:

Condition 1: "Has easy access to presets."
Condition 2: "Center box seems like a lot of wasted space."
Condition 3: "Clear and easy to read."

Image 15:

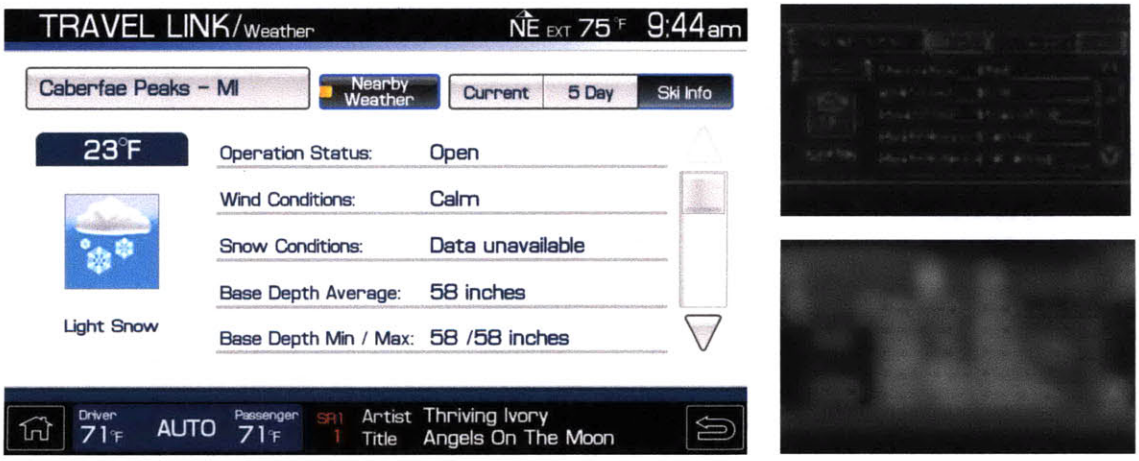
Condition 1: "I like the red things that show you when you are getting too close."
Condition 2: "Good view of where you are aiming."
Condition 3: "(Like) color coding of regions on ground."

Image 17:



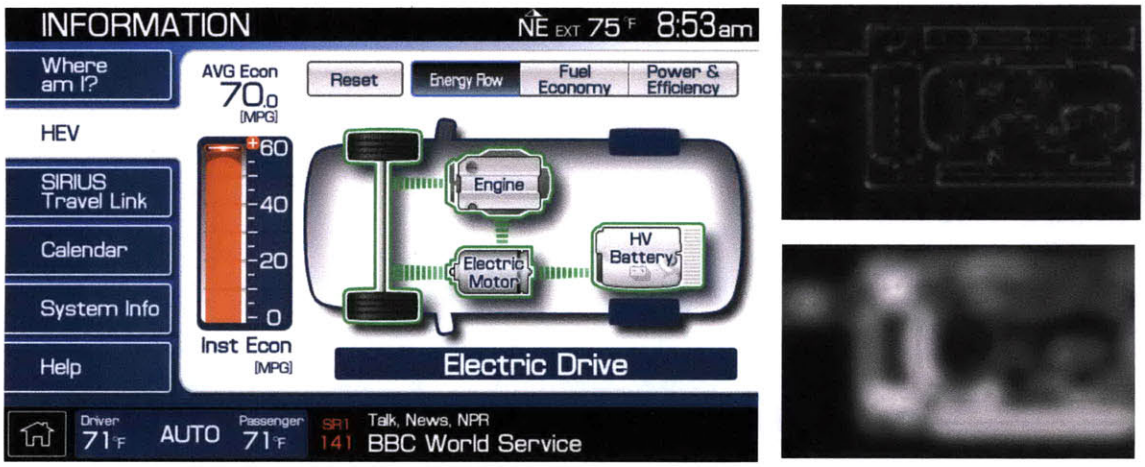
Condition 1: "Easy to navigate away from."
Condition 2: "Well organized and easy to read."
Condition 3: "All the information is easy to take in 'at a glance.'"

Image 18:



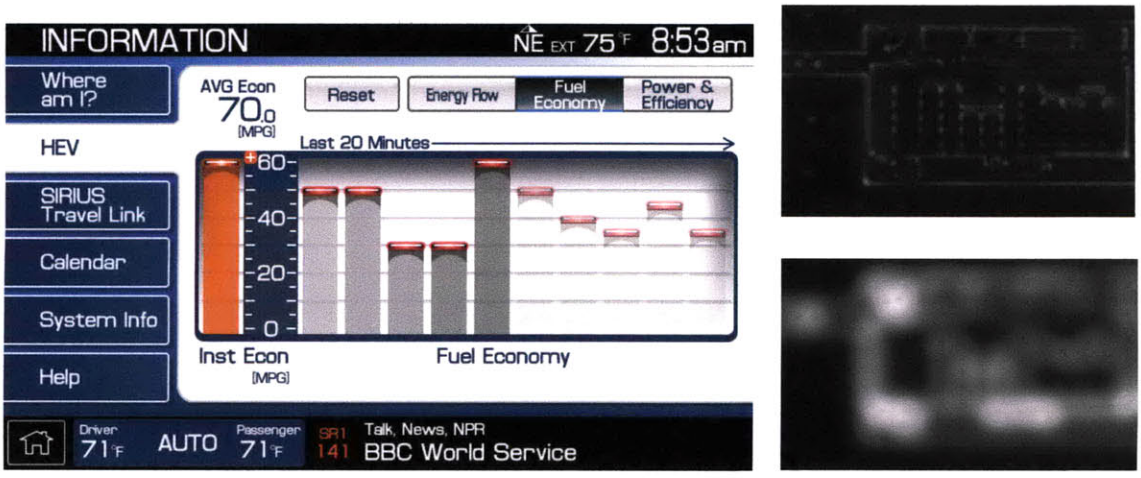
Condition 1: "Nearby Weather button is the wrong color and just looks strange."
Condition 2: "Everything is a good size and easy to read/understand."
Condition 3: "The grey lines should be more salient to make each line pop out."

Image 19:



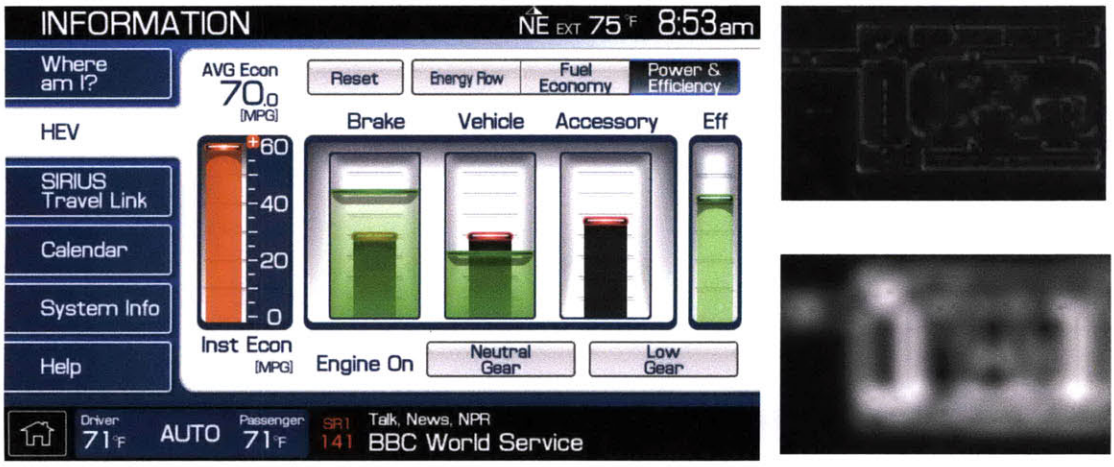
Condition 1: "Laid out clearly."
Condition 2: "Navigation buttons are a good size."
Condition 3: "The additional climate information and radio station details at the bottom are helpful and don't make the page feel cluttered."

Image 20:



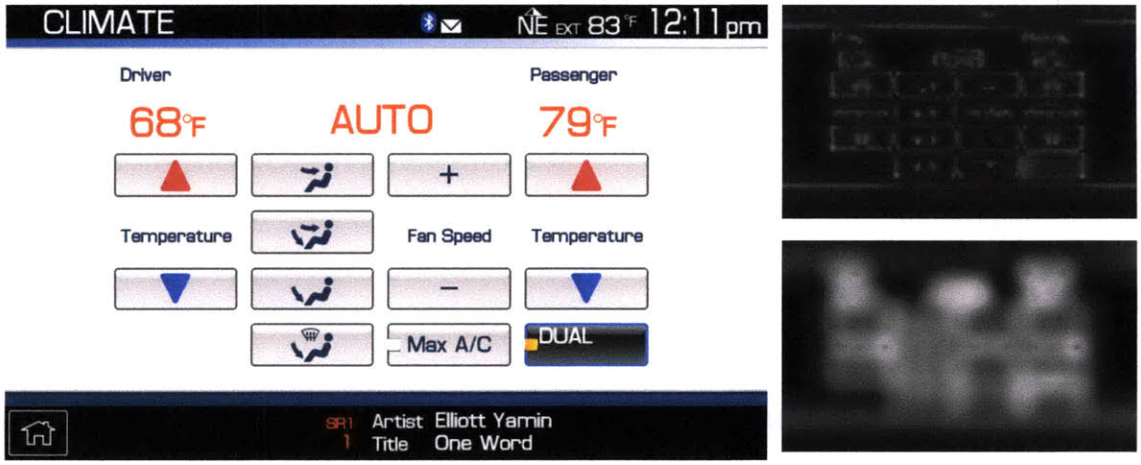
Condition 1: "Reset button should be separate."
Condition 2: "Fuel economy number still small."
Condition 3: "Grey makes me think it is not important."

Image 21:



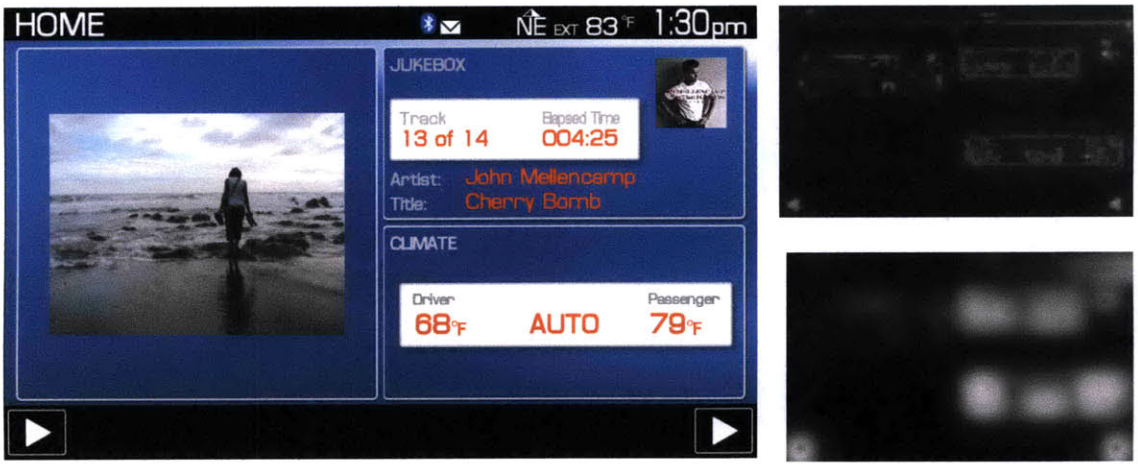
Condition 1: "Not sure what the red bars are compared to the green bars."
Condition 2: "Nice use of color."
Condition 3: "There is more clutter in the middle."

Image 22:



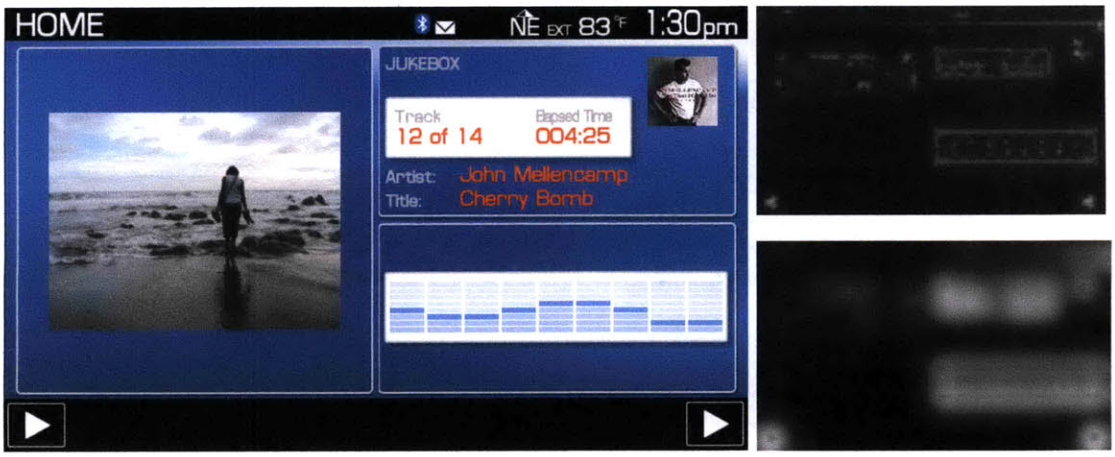
Condition 1: "The screen looks too cluttered."
Condition 2: "Artist/title seems like an afterthought."
Condition 3: "Took a little longer to understand vertical organization of the buttons."

Image 24:



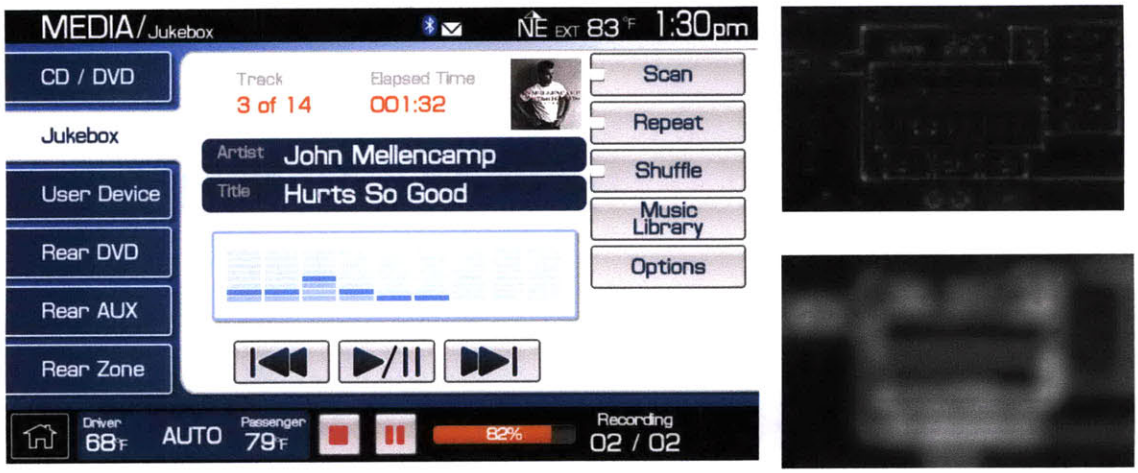
Condition 1: "Confused why there are two play buttons."
Condition 2: "Colors are terrible."
Condition 3: "Track and elapsed time are low contrast."

Image 25:



Condition 1: "You can't stare at album art while you drive."
Condition 2: "Text large enough to read."
Condition 3: "Low contrast for track and elapsed time."

Image 27:

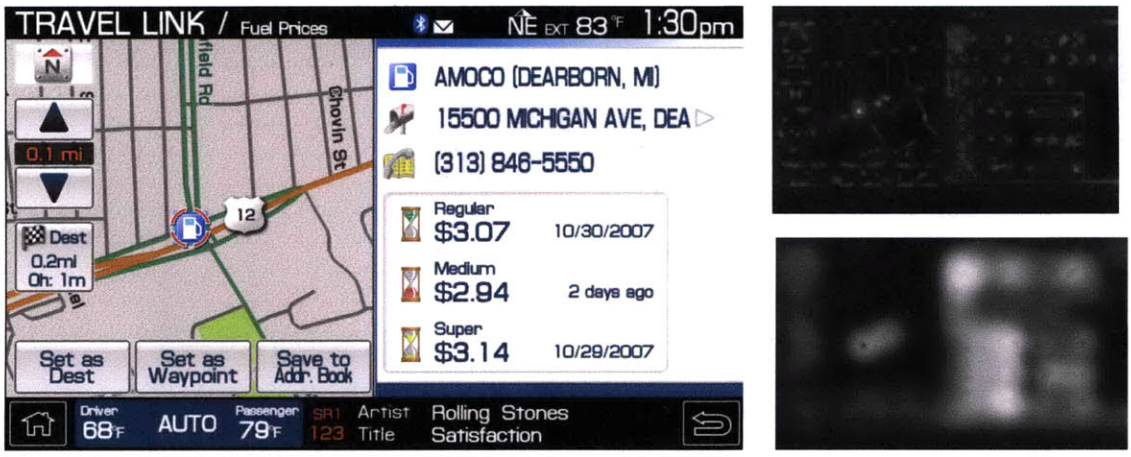


Condition 1: "Most of the interface makes sense."
Condition 2: "Text all the way across makes for a cluttered visual."
Condition 3: "Clutter seems a little scrunched up in the center."

Image 30:

Condition 1: "Screen is getting busy with all the buttons on the right."
Condition 2: "Too many options to interact with on the screen at once."
Condition 3: "The graphic equalizer is just clutter when you are choosing songs."

Image 36:



Condition 1: "The icons for phone # and address are nice and intuitive."
Condition 2: "Too much info on screen."
Condition 3: "Too many icons that clutter up the information and make it hard to get the details 'at a glance.'"

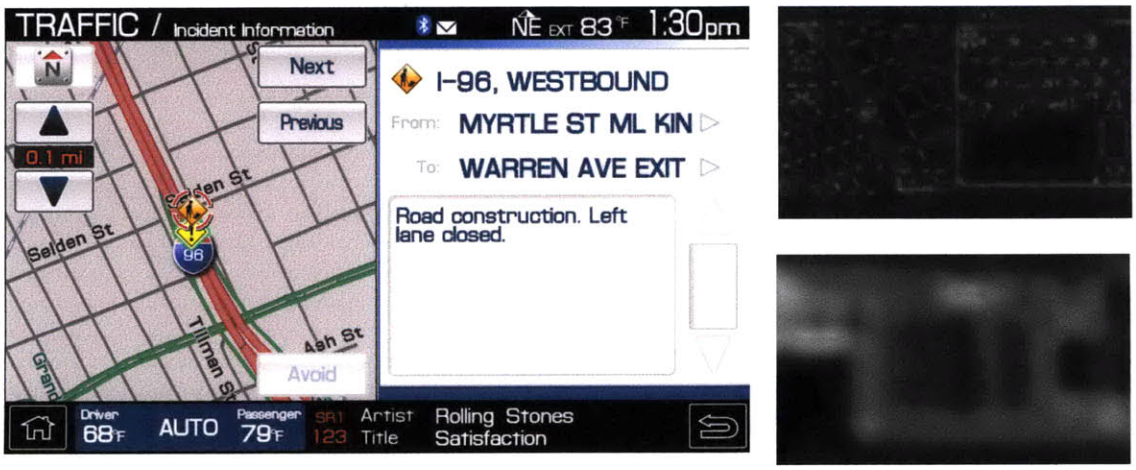
Image 37:

Condition 1: "This layout makes more sense for gas stations than movie theaters."
Condition 2: "Arrows make it clear which direction the gas station is."
Condition 3: "Lots of information on one page."

Image 42:

Condition 1: "Icons are redundant (and distracting)."
Condition 2: "Want an option/ability to see time till arrival."
Condition 3: "Why is the select below grayed out?"

Image 45:



Condition 1: "I don't like how all of the icons are stacked on top of each other on the map."
Condition 2: "Warnings piled on top of each other confusing - obscures street name."
Condition 3: "The wording of the streets is also cluttered and the additional 'ml kin' makes no sense."

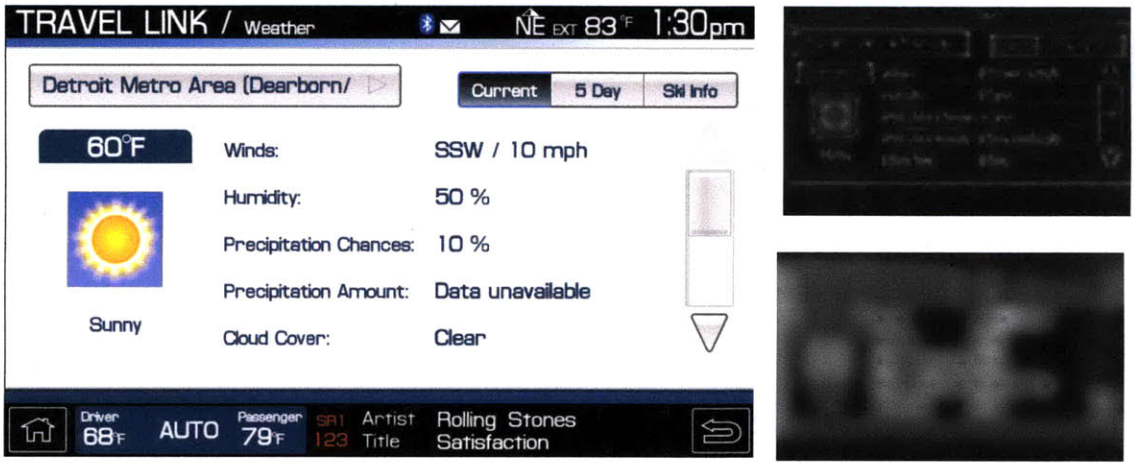
Image 46:

Condition 1: "Layout is uncluttered."
Condition 2: "Easy to read words."
Condition 3: "Main section is very clean."

Image 47:

Condition 1: "Pictures are clear and layout is uncluttered."
Condition 2: "Good sized weather icons."
Condition 3: "The graphics are clear and similar to other familiar graphical representations of the weather."

Image 48:



Condition 1: "Too much info at top and bottom of screen that isn't differentiated by size or color."
Condition 2: "Text too small and cluttered."
Condition 3: "The amount of white space seems rather bright."