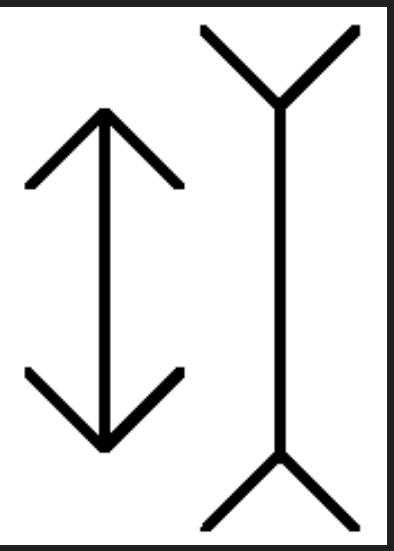
User Evaluation

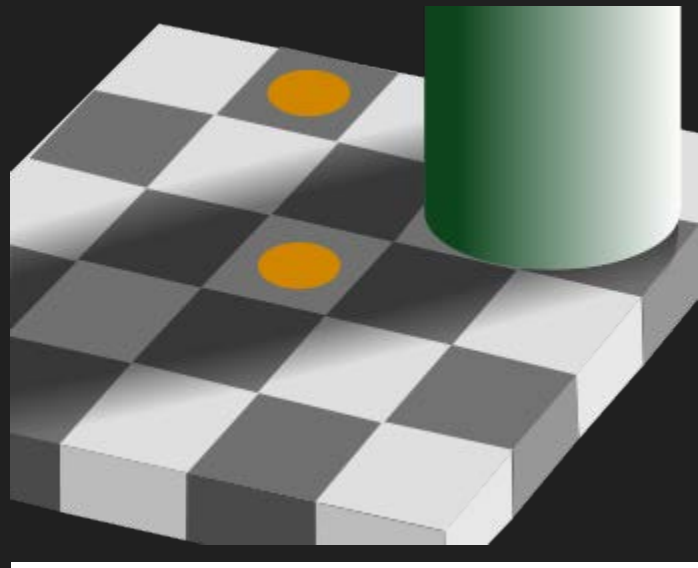
Alark Joshi

Optical Illusions

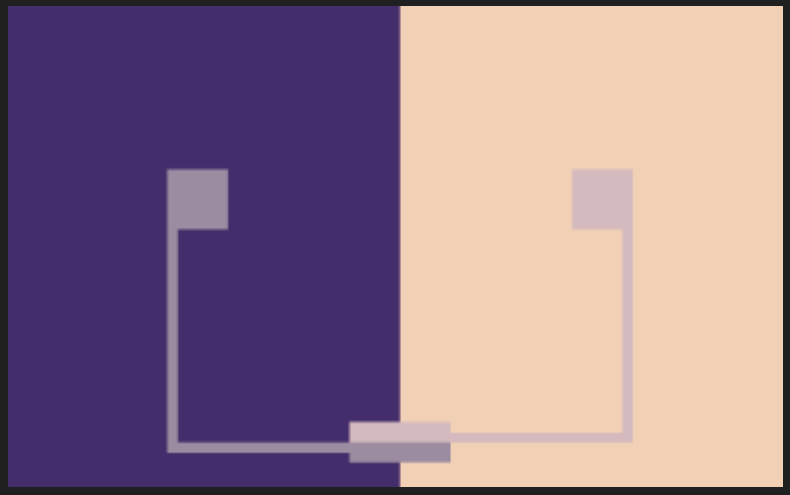


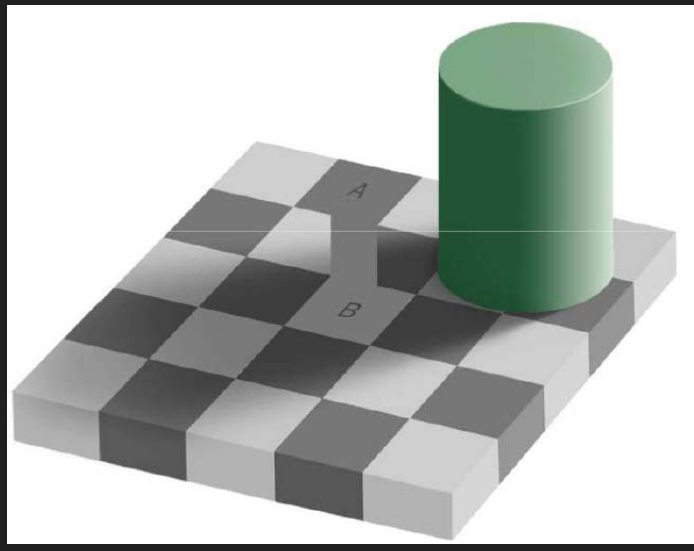
Same Length

Same Color perceived differently



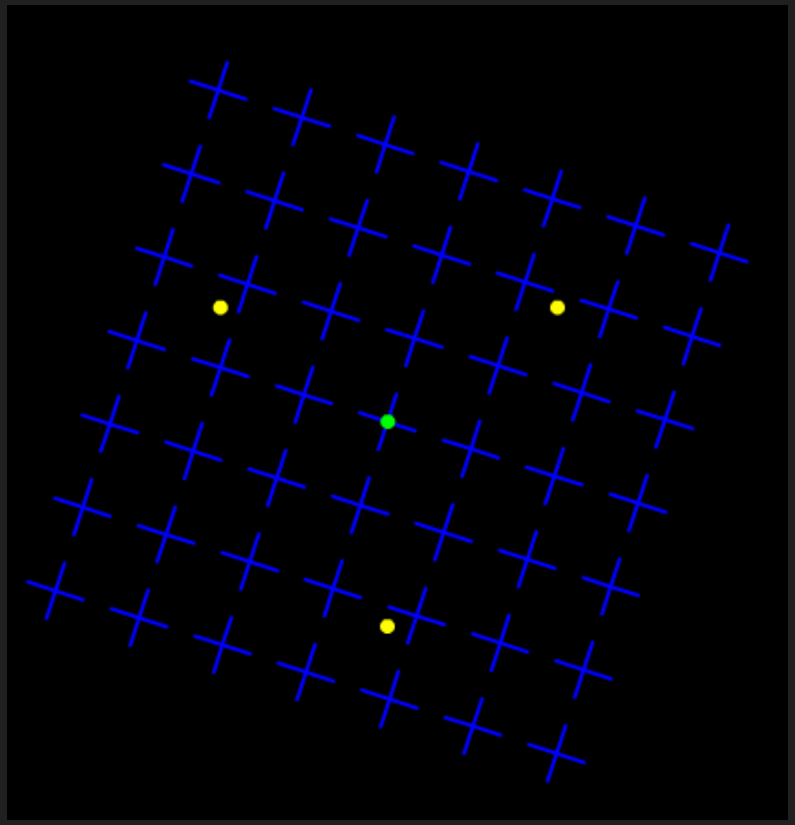
Different Color perceived as same





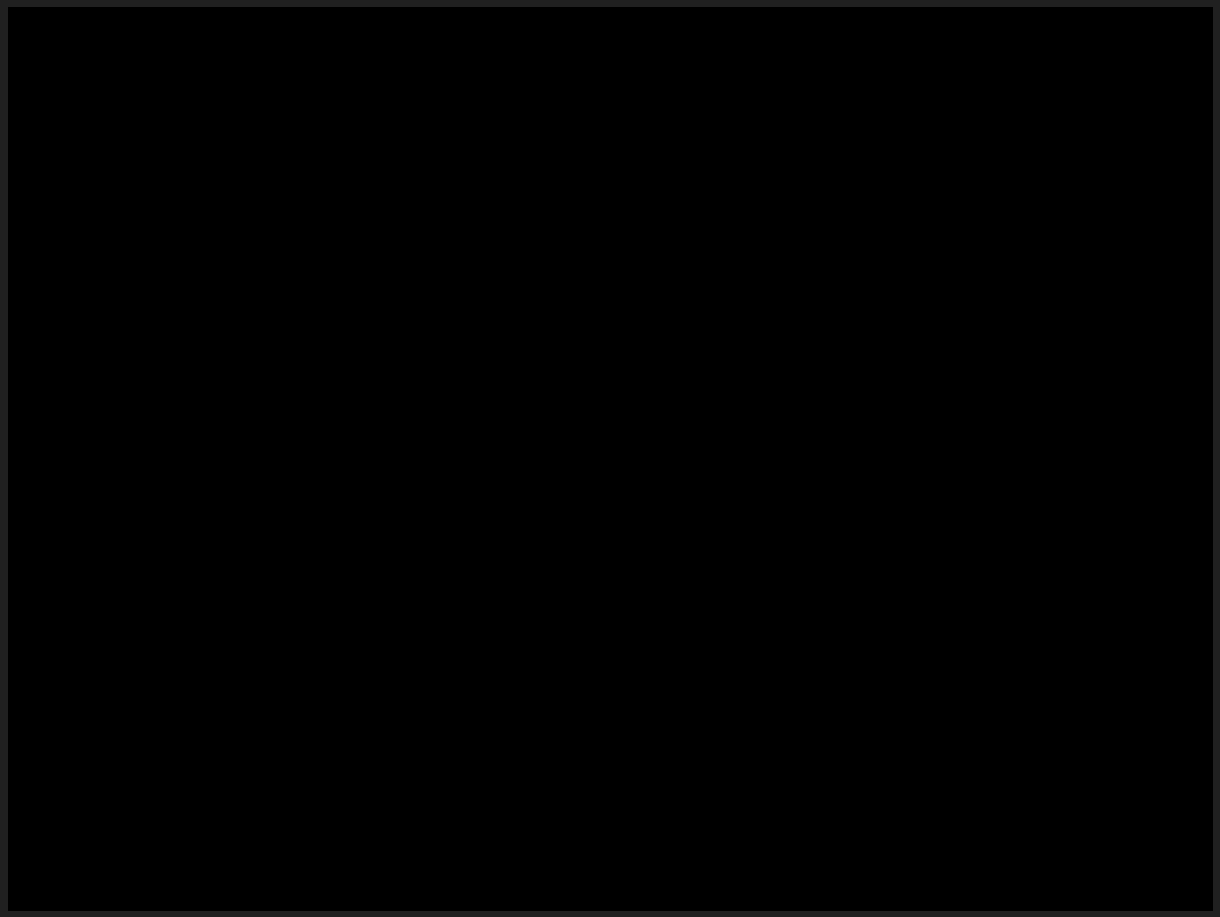
Motion Induced Blindness

- Fixate on the center – The yellow spots disappear once in a while (sometimes 1 and sometimes all 3)



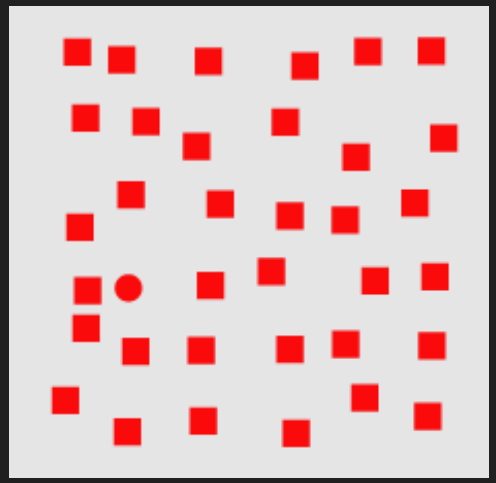
http://www.michaelbach.de/ot/mot_mib/index.html

Inattentional Blindness

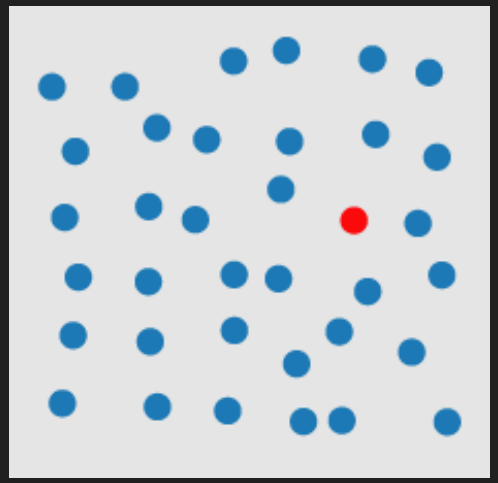


http://www.youtube.com/watch?v=Ahg6qcgoay4

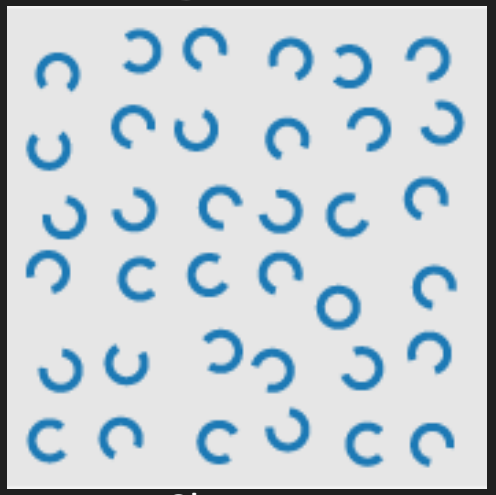
Preattentive Processing



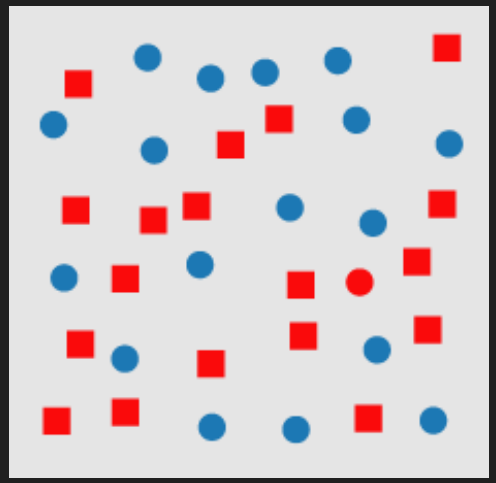
Shape



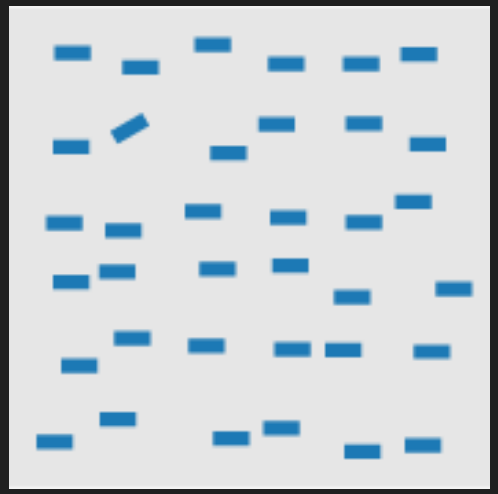
Color



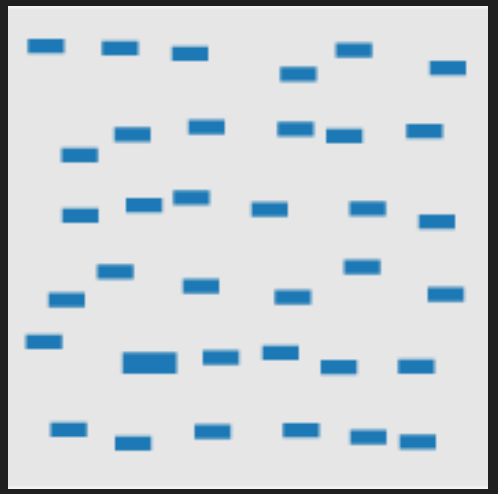
Closure



Shape and Color



Orientation



Size

Credits: http://www.csc.ncsu.edu/faculty/healey/PP/

Need for Evaluation

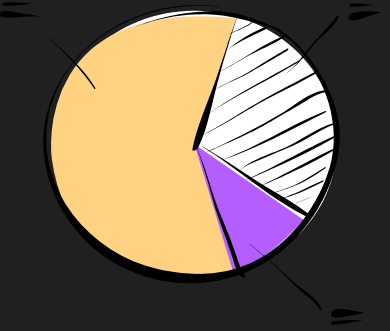
- Visual system is extremely complex
- Misrepresentation of data can have catastrophic results in some cases
  - Challenger disaster could have been averted - Tufte
- Need to ensure consistent interpretation of data using a visualization technique in various scenarios
- “User studies offer a scientifically sound method to measure a visualization technique’s performance”
  - Kosara et al.

Need for Evaluation

- Novel visualization techniques provide a unique representation of data
- How do you measure the effectiveness of the technique?
- How can we be sure that the new technique is not misrepresenting/misleading the user?
- How can we confidently say that your visualization technique is ‘better’ than existing techniques?

User Evaluations

- Conduct a user study to evaluate which visualization technique is ‘better’
- A user study involves users performing tasks using a variety of visualization techniques
- The idea is to evaluate the visualization techniques based on the performance of the users – For example, Yost and North ’06 (The Perceptual Scalability of Visualization)
- How do you measure the performance of a user?



User Study

- Compare Pie Charts and Bar Graphs
- Learn about the components of a user study
- Analyze the results

Dependent Variables

- How do you measure the performance of a user?
- Performance metrics
  - User Accuracy at performing the tasks
  - Time required to perform the tasks
  - Emotional response
    - Confidence that the users' have in their answers (which indicates the effectiveness of a technique)
    - Stress caused by a particular technique
  - Qualitative feedback in the form of a questionnaire that lets the user provide additional feedback

Conducting a User Study

- What are you testing?
  - Efficacy of a novel visualization technique
  - Evaluating pre-existing techniques to find the best one for a certain task
- Identify your **Null Hypothesis**
  - For example, the Null Hypothesis would be that that the novel visualization technique is not more effective than currently used techniques

Independent Variables

- What are the variables in the study that are independent?
- Variables that we manipulate in the study
- The visualization techniques that you show the user are your independent variables
  - E.g. You have three independent variables if you are comparing Treemaps vs Bar Charts vs Pie Charts

Datasets

- The datasets chosen for the study should be standard and anonymous
- Mock/Test datasets are acceptable as long as they do not favor a certain technique
- Ideally, the study should evaluate the visualization technique on a couple of real-world datasets
- Provides legitimacy to the results

Before you start the study

- Identify the tasks that users/experts would do on a regular basis
  - Try to keep the tasks general and reasonably short
- Create *fair* visualizations of the data for all the visualization techniques involved
- Identify/Solicit unbiased participants for the study
  - Attempt to have a balanced gender, age pool for the study
  - For expert studies, identify appropriate participants

Before you start the study

- In addition to verbally describing the study and its purpose, have a *written description* of the ready for the participants
  - Provides perspective to the participants
  - This allows participants to leave if they feel uncomfortable with the study
- Make sure you have the questionnaire ready for users to fill out after the study is completed
- Ensure that you have ways to record data about the participants before the study commences

Ethical Considerations

- Sometimes tests can be distressing and tiring
- Ensure that you are clearly communicating to the participants that
  - They can leave at any time
  - They are not pressured to participate
  - The data collected will be anonymous
- Every user study has to be approved by an *Institutional Review Board*

Capturing User Data

- Automatically recording performance metrics
- Requesting participants to ‘think aloud’
  - Tell us what they are trying to do
  - Tell us what questions arise as they work
  - Tell us what problems are arising as they work
- Video recording the participants study
- Recording on-screen activity

Pilot Evaluation

- Before you start the study, it is always a good idea to
  - Take the study yourself to identify potential problems with the study
  - Perform a pilot evaluation with 2/3 participants to make sure the tasks, visualizations and description is not ambiguous
- Invariably end up finding problems and allows you to fix them before you start the study

Conducting the user study

1. Welcome and introduce yourself to the participant
2. Explain the entire process to the participant
3. Clearly indicate to the participant that s/he is allowed to leave the study at any time
4. Allow the participant to read the written description and obtain consent from the participant
5. Introduce the participant to the basics of the study

Conducting the user study

6. Inform the participant that you will be measuring metrics (time, accuracy etc.)
7. Patiently answer all the questions that they may have
8. Let the participant start the study
9. Once the study has been completed, request the participants to fill out the questionnaire
10. Thank the participant for completing the study

Ordering Effects

- Ordering of conditions is a variable that can confound the results
- Randomization – To ensure that there is an ordering effect does not affect the outcome of the study
- Control – To ensure that a variable is held constant for all cases
  - Same Datasets for multiple visualization techniques

Ordering Subjects

- Two types of ordering the participants
- Within-Subjects design
  - All participants see all visualization techniques
  - Ordering and fatigue effects
- Between-Subjects design
  - One set of participants sees one technique and performs tasks pertaining to that techniques
  - Cannot isolate effects due to individual differences
  - Requires more participants

Measuring Effectiveness

- Performance metrics can provide insight
  - Consistently accurate results using a technique
  - Significantly faster using a technique
  - Higher confidence when using a technique
- Do not combine time with ‘think aloud’
  - Talking will affect speed
- Caveats
  - Faster is not always better
  - Higher confidence/speed but less accurate?

Pie Charts vs Bar Graphs

- Complete the study based on the instructions given – 10 mins
- Turn in your study when you are done

Quantitative analysis

- Identify statistical significance for performance metrics
- *Student t-test* computes statistical significance if two quantities are being compared
- *ANOVA (Analysis of Variance)* computes statistical significance if more than two quantities are being compared
- Output in the form of *p-value*
  - $P < 0.05$ = statistically significant
  - $P < 0.01$ = highly statistically significant

Qualitative Analysis

- ‘Think aloud’ feedback useful
- Answers to the Questionnaire provide crucial feedback about preferences, difficulties with a certain technique, ideas for improvement etc.
- Use Feedback with Quantitative results
  - To determine which technique is ‘better’
  - Make a list of positive and negative events
  - Refine the technique/product for a second round of studies

Report your Results

- Report your results in full detail
  - Research paper
  - Marketing brochure
- Provide sufficient detail to allow another team to conduct the same study
- Results provide a legitimacy to your claims of a more effective visualization technique/system
  - Example: Yost and North, Laidlaw et al. and so on

Issues of Perception and Cognition

- Advent of huge displays and computing power overwhelming our perceptual senses
- Visualizing huge multimodal datasets on these displays may cause perceptual and cognitive overload
- Need to evaluate and design visualizations that are effective and easy to understand

Now for something completely different!

- Pixar Short – Lifted
  - http://www.youtube.com/watch?v=pY1_HrhwaXU

User Studies

- Robert Kosara, Christopher G. Healey, Victoria Interrante, David H. Laidlaw, and Colin Ware. User Studies: Why, How, and When?. IEEE Comput. Graph. Appl. 23, 4 (July 2003), 20-25.
- Increase the awareness of the need for user evaluation in visualization
- Discuss scenarios where user studies may be appropriate and situations where other techniques may be more appropriate (videotaping users using the system, etc.)

Other Evaluation Methods

- Expert evaluation
  - Evaluate with the help of a small group of domain experts
- Interview-based evaluation (ManyEyes paper)
- Multi-dimensional In-depth Long-term Case studies (MILC)

Multi-dimensional In-depth Long-term Case studies (MILC)

- Ben Shneiderman, Catherine Plaisant, Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies, BELIV '06.
- Assessing insight, discovery is hard
- Efficacy can be assessed by documenting
  1. usage (observations, interviews, surveys, logging etc.)
  2. expert users' success in achieving their goals

Challenge of Information Visualization Evaluation

- Plaisant, C. The Challenge of Information Visualization Evaluation. Advanced Visual interfaces, Italy, 2004, ACM Press.
- Discuss other evaluation approaches that take into account the long exploratory nature of users tasks
- Address the problems associated with evaluating benefits

Information Visualization Evaluation

- "Discovery is seldom an instantaneous event"
  - Observe users over a larger period instead of a short user study with a single tool
- "Way to answer questions you didn't know you had"
  - Allow free exploration of data and report on what they learned and understood (less effective with non-experts)
  - Letting subjects use their own data can increase engagement and interest

Information Visualization Evaluation

- "Factoring in the chances of discovery and the benefits of awareness"
  - Potential adopters have to consider risks associated with errors caused due to visualization
  - Streamlining repetitive tasks is also a benefit that adopters like in place of novel discovery
- "Success cannot be tracked back to visualization"
  - Freire and Silva propose the use of provenance to track user tasks and "quantify" insight

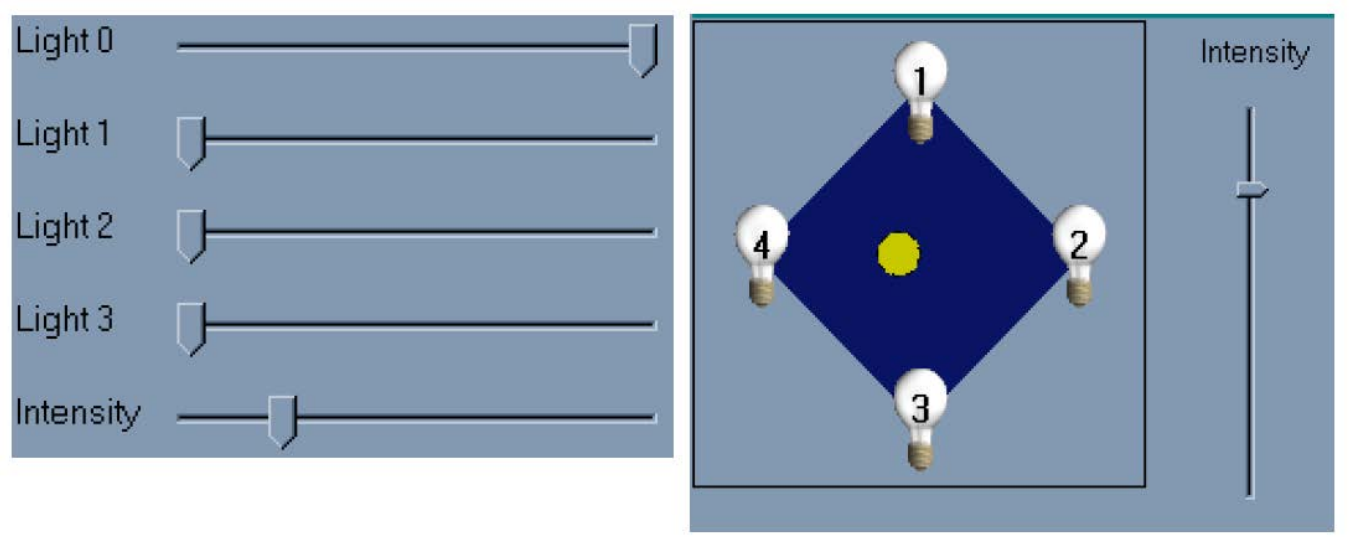
Examples of Tech Transfer

From treemap to the Map of the Market and other treemap tools

Expert Reviews

- Tory, M., Moeller T. Evaluating Visualizations: Do Expert Reviews Work? IEEE Computer Graphics and Applications, 25(5), 2005, 8-11.
- Asking a few friends for their opinion isn't sufficient and may miss valuable information
- Expert reviews identify usability problems and are efficient (5 experts find 75% problems)
  - Compared to a study of 50 participants

Comparing light control widgets



- Heuristic evaluation that focuses on GUI issues, generic visualization tasks and tasks specific to their tool
- Light dial (right image) facilitated faster exploration
- Sliders better for understanding light contributions

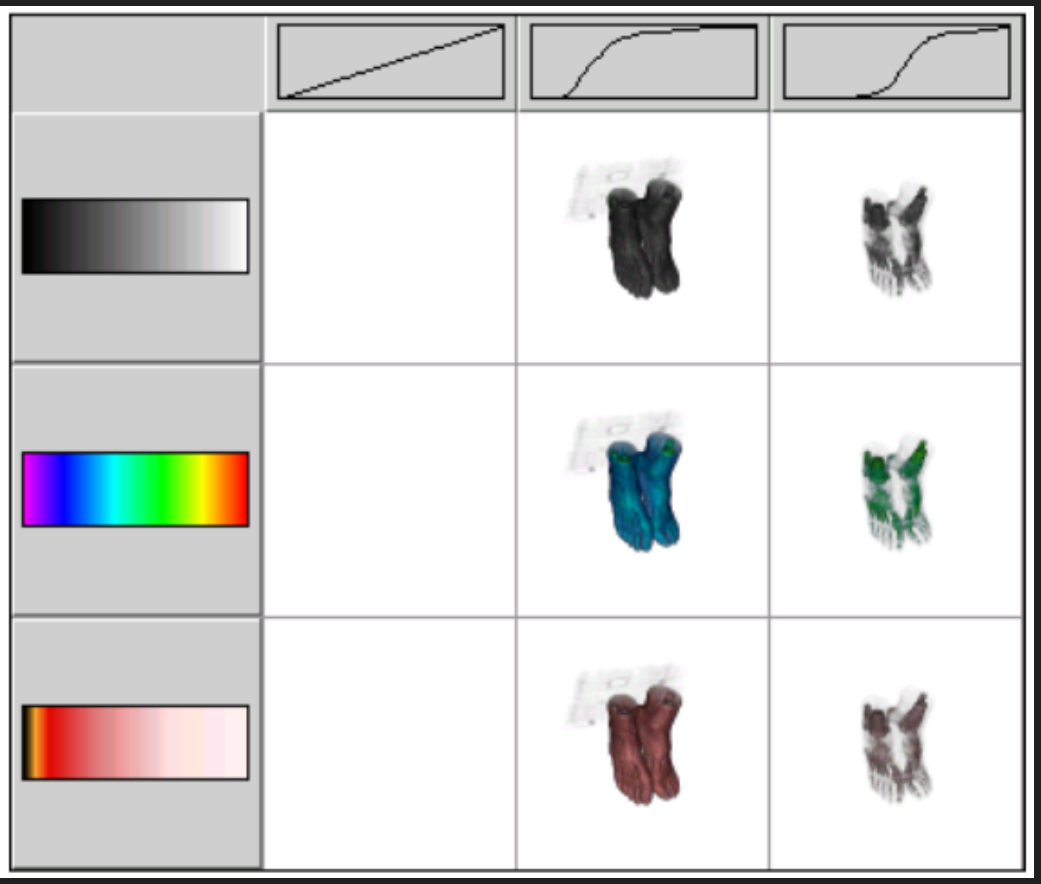
Comparing light control widgets

- Usability experts had far more insight
- Of the experts, only two focused on data analysis tasks
  - Due to their training in medical imaging
- Involving usability experts with domain knowledge is more important than just usability experts

Comparing volume rendering interfaces

- Expert reviews on two volume rendering interfaces
- Table interface vs parallel-coordinates style interface
- Used usability experts with domain knowledge for this study

Table Interface



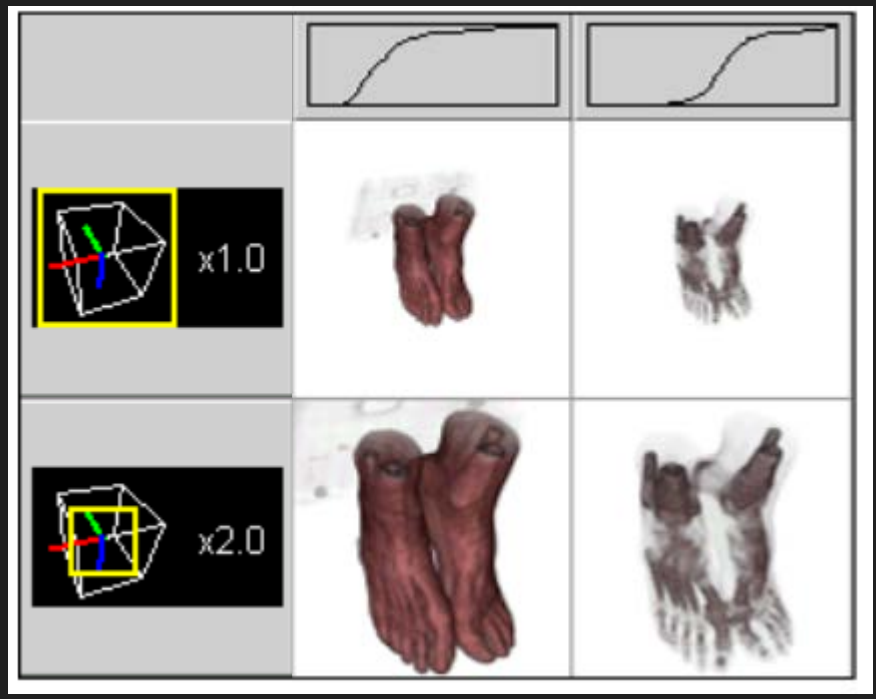


Image credits: TJ Jankun Kelly and Kwan-Liu Ma

Parallel-Coordinates Style Interface

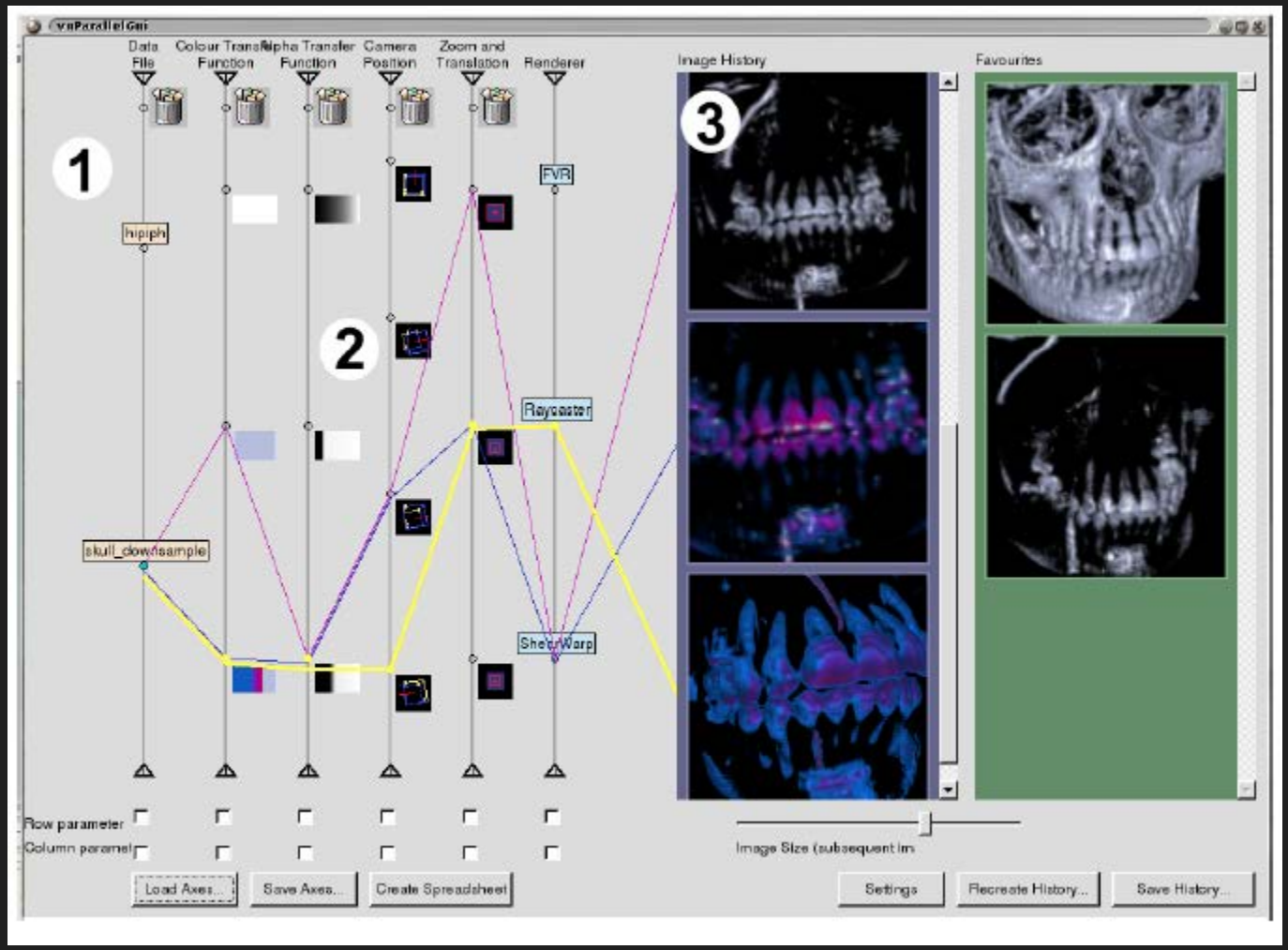


Image credits: M. Tory and T. Moeller

Two tasks

1. Explore several data sets
2. Search for an identifiable object using the interfaces

- Experts provided written feedback along with ratings
- Observed participants and recorded their opinion throughout the study
  - Led to identification of misconceptions

Comparing volume rendering interfaces

- Table interface – Useful for quick exploration of settings
- Parallel coordinates interface – Identifying available display options and manipulating display settings
- Small images were seen as a problem

Summary

- Recording and analyzing the observations is hard
  - Explore video recording
- Highly valuable and applicable feedback
- Experts provide quick and valuable insight
- Not a substitute for user studies
- “Problems” found by experts may not affect end users
- Experts with usability experience can provide applicable feedback

Seven Scenarios - InfoVis Evaluation

- Heidi Lam, Enrico Bertini, Petra Isenberg, Catherine Plaisant, Sheelagh Carpendale, Empirical Studies in Information Visualization: Seven Scenarios, TVCG, December 2011.
- Provide a descriptive view of the various evaluation scenarios in Information Visualization
- Surveyed 850 papers to categorize them into seven distinct classes (scenarios)

Scenarios

- Identify your user evaluation goals
- Provide an insight into the kinds of questions that can be asked for specific scenarios
- Each scenario contains a description of the situation in which you could use specified questions and methods to evaluate results

Scenarios

1. Understanding Environments and Work Practices (UWP)
   - MILC falls into this category
   - Spend time at the site of the client trying to understand the kinds of tasks, visualizations and barriers being used
2. Evaluating Visual Data Analysis and Reasoning (VDAR)
   - Study the data analysis oriented tasks (hypothesis generation, knowledge discovery, decision making)

Scenarios

3. Evaluating Communication through Visualization (CTV)
   - Conveying information to a large group of people
   - Aim to evaluate efficacy, usefulness and interaction patterns
4. Evaluating Collaborative Data Analysis (CDA)
   - Evaluate the ability to facilitate seamless collaboration
5. Evaluating User Performance (UP)
   - Identify limits of human perception, compared visualization techniques by examining performance

Scenarios

6. Evaluating User Experience (UE)
   - Usability testing in terms of useful features, missing features, improvements, learning curve
7. Evaluating Visualization Algorithms (VA)
   - Evaluate the efficacy of algorithms in being able to highlight patterns, produce least cluttered view, ability to scale (with large displays, huge datasets)
   - Examine the performance in terms of efficiency of an algorithm

Many-to-many mapping

- Scenarios don't make directly to a single evaluation method
- Situations where your evaluation goals may be broad may include multiple scenarios
  - Exploring data analysts process of knowledge discovery may include UWP, VDAR, UE, VA and so on

User study: 2D Vector Visualization

- D. H. Laidlaw, M. Kirby, C. Jackson, J. S. Davidson, T. Miller, M. DaSilva, W. Warren, and M. Tarr (2005). Comparing 2D vector field visualization methods: A user study. TVCG, 11(1):59-70, 2005.
- Conducted a thorough evaluation of expert and non-expert users for visualization techniques
- Compared six visualization methods for 2D vector visualization