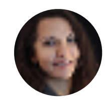
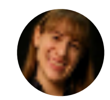
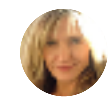
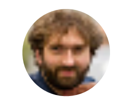
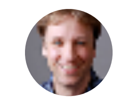
BubbleView: an interface for crowdsourcing image importance maps and tracking visual attention

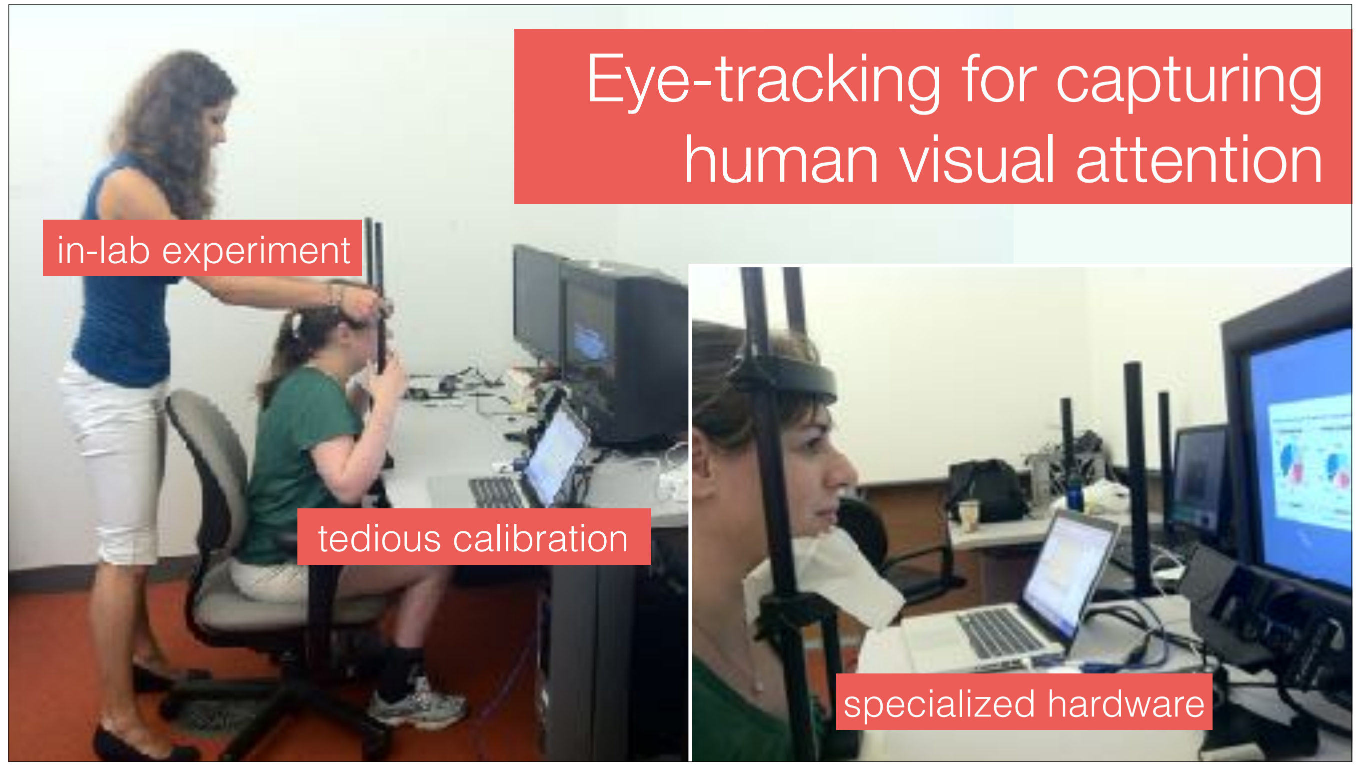
Nam Wook Kim*, **Zoya Bylinskii***, Michelle Borkin, Krzysztof Gajos, Aude Oliva, Fredo Durand, Hanspeter Pfister

CSAIL

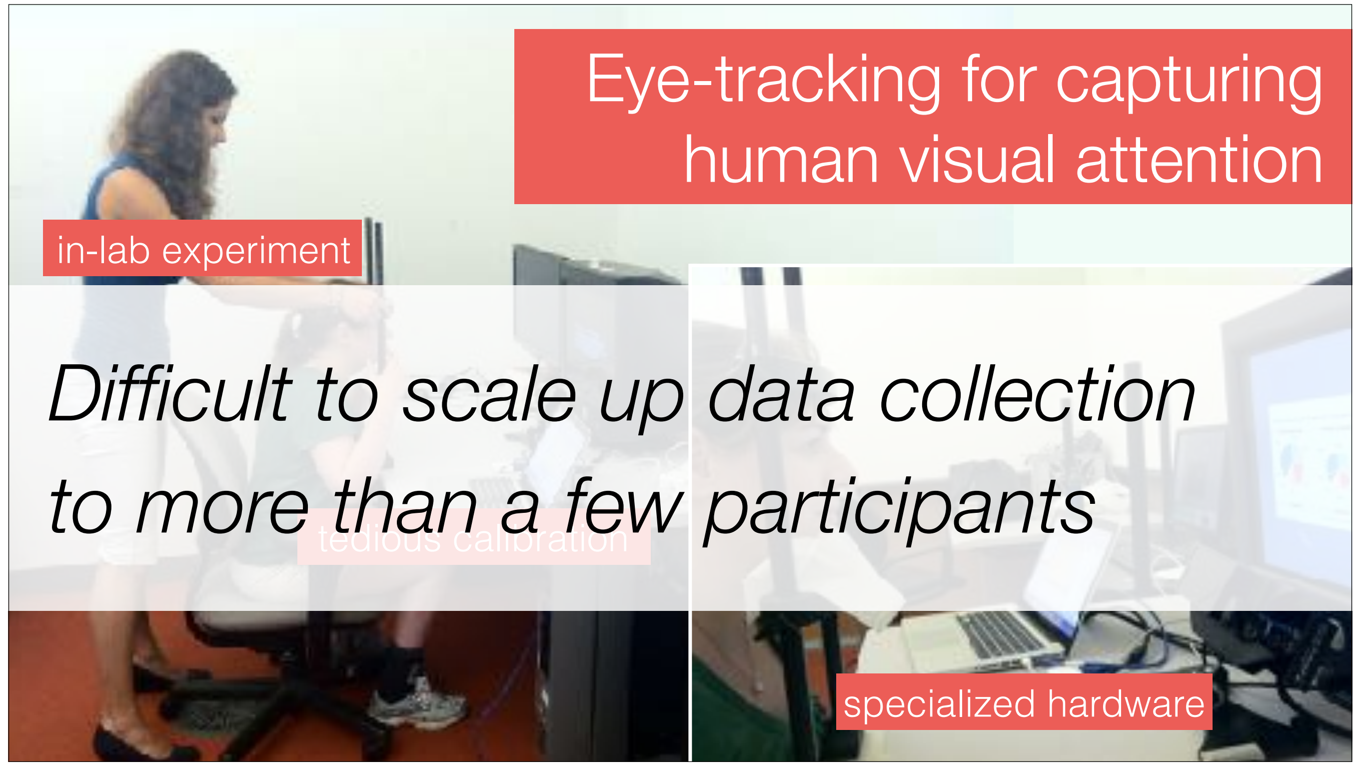
HARVARD John A. Paulson School of Engineering and Applied Sciences

Northeastern University College of Computer and Information Science

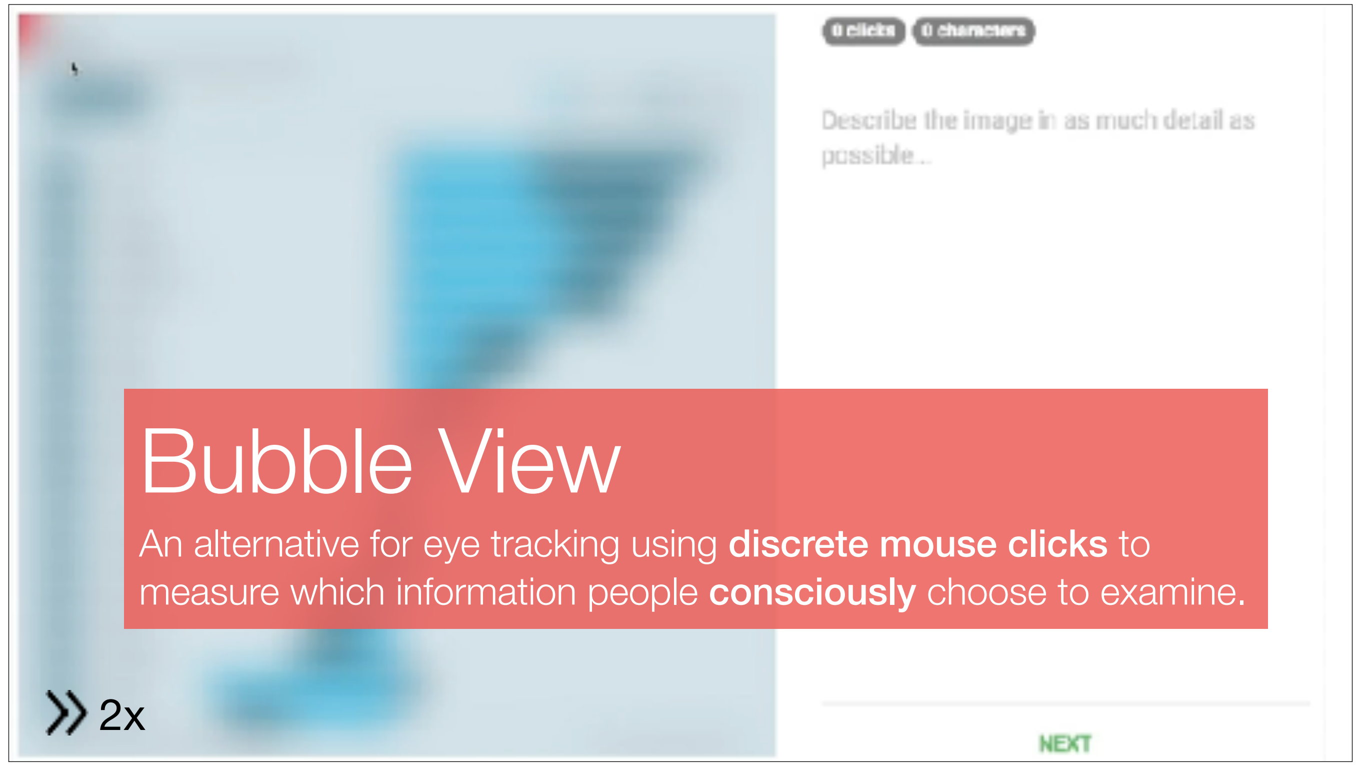
Eye tracking is an extremely valuable tool for capturing human visual attention, to study where people look at and how long they spend.



However, it has a couple of limitations. You can run these types of experiments outside of a lab setting. Eye-trackers usually require tedious calibrations and expensive & specialized hardware like infra-red sensors.

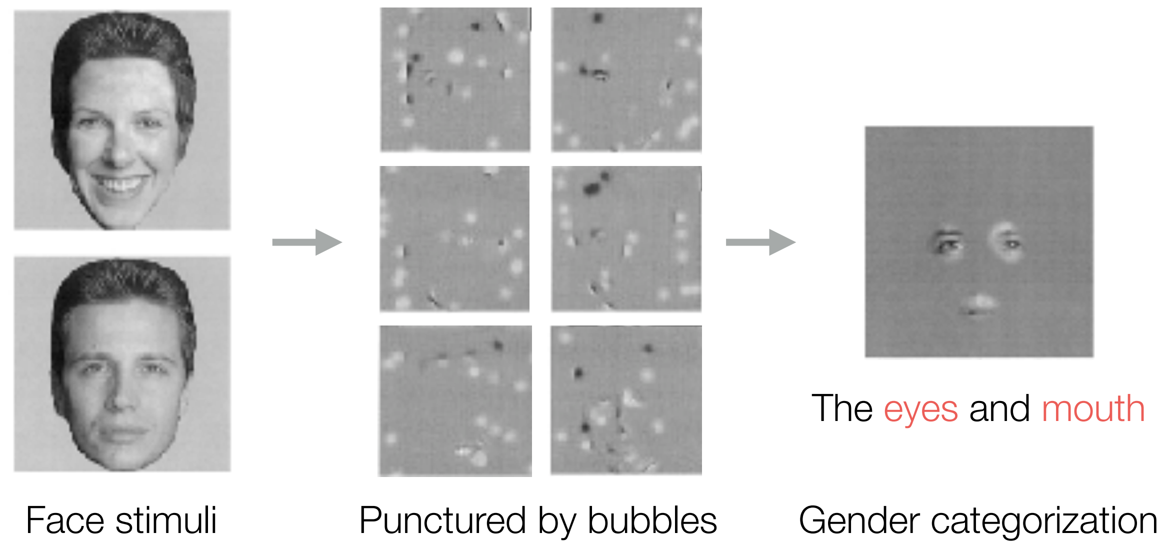


This means it is difficult to scale up data collection to more than a few participants.



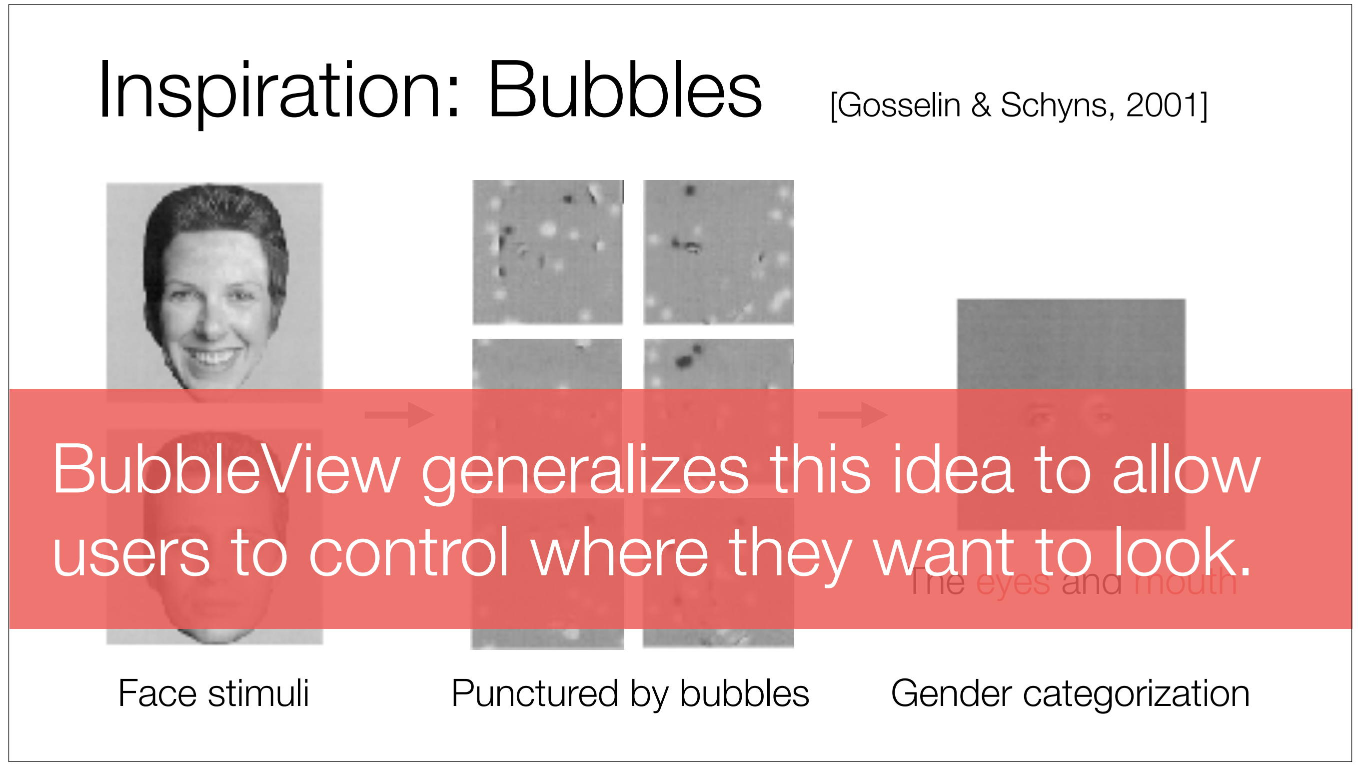
To address this issue, we developed BubbleView, a click-based alternative to eye tracking. The BubbleView interface presents a blurred static image and allows users to click to reveal bubbles, small circular areas of the image at the original resolution, just like having a confined area of focus similar to the human eye fovea.

Inspiration: Bubbles [Gosselin & Schyns, 2001]



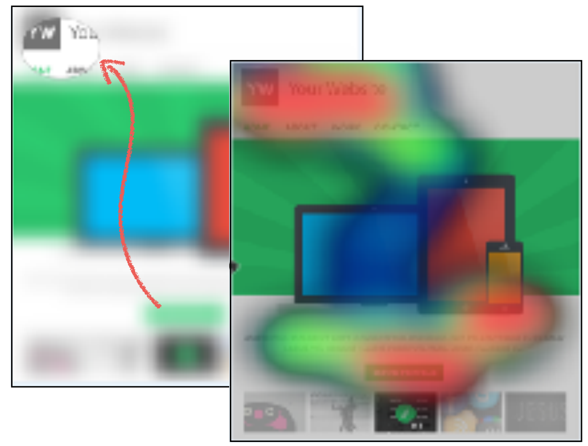
BubbleView was inspired by the work in Psychology by Gosselin & Schyns. In their Bubbles paper, they partially revealed the faces by a grey color mask punctured by randomly generated bubbles which are Gaussian windows. They gave parcipants categorization tasks to identify which image regions are important to the tasks. For example, they found the eyes and mouth are important diagnostic regions for recognizing gender.

“Bubbles: a technique to reveal the use of information in recognition tasks” (2001)

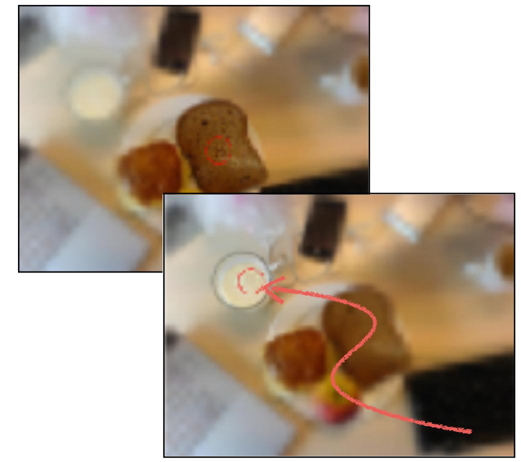


BubbleView generalizes this idea to allow users to control where they want to look.

Cursor-Based Attention Tracking

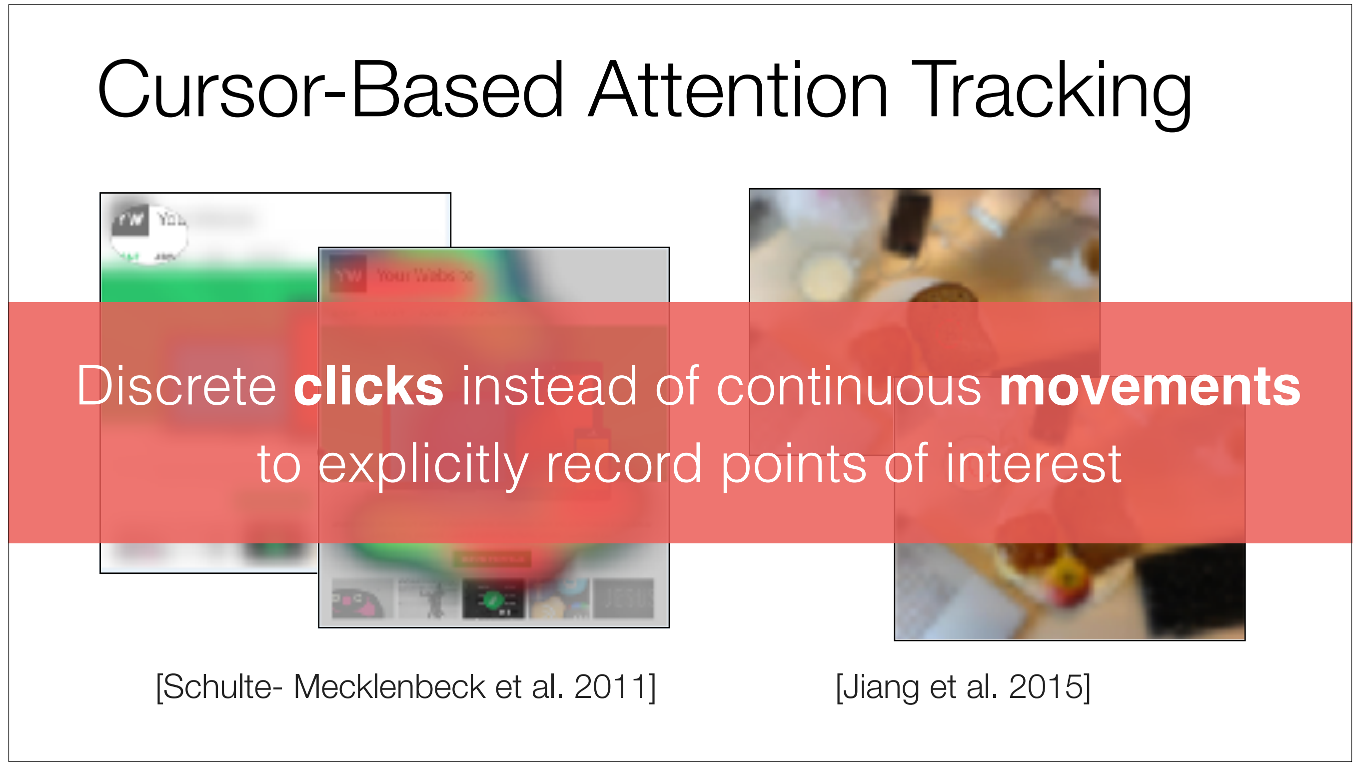


[Schulte- Mecklenbeck et al. 2011]

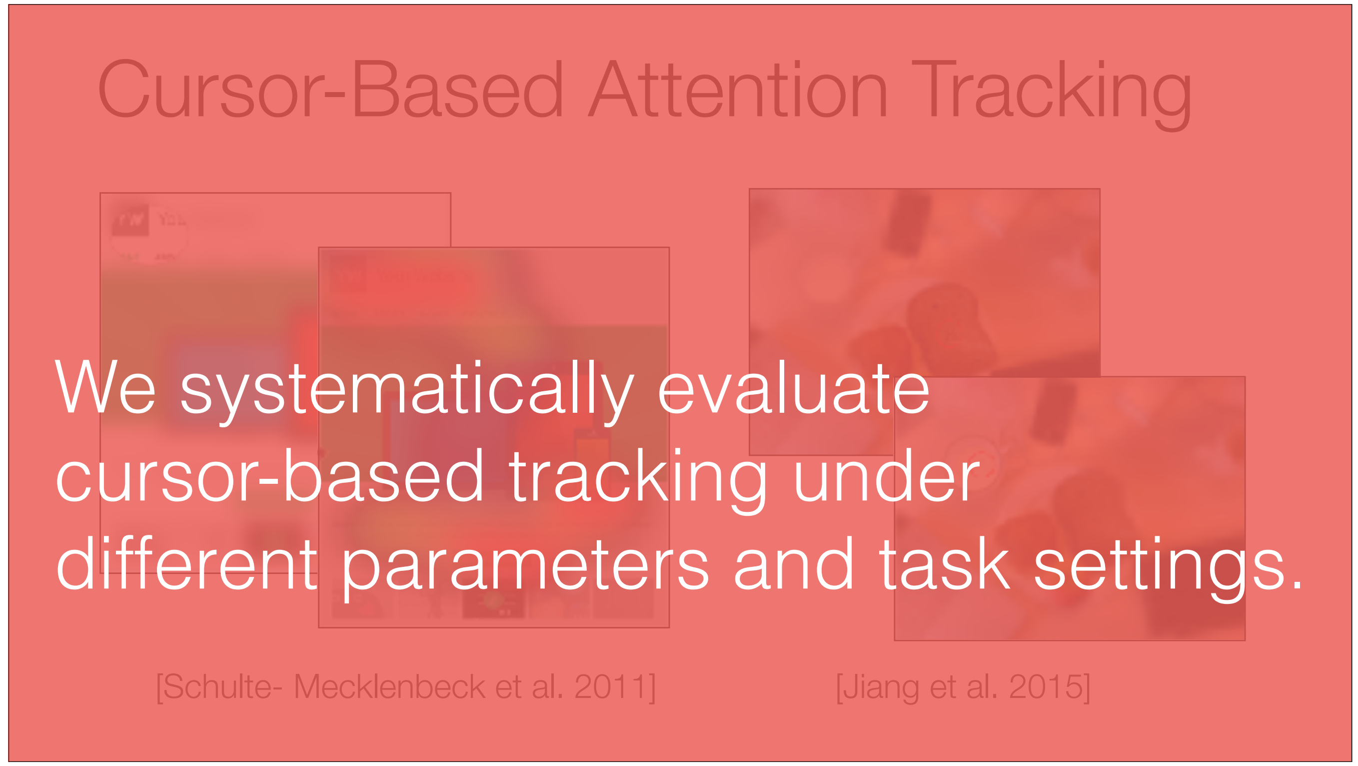


[Jiang et al. 2015]

There have been many cursor-based methods.

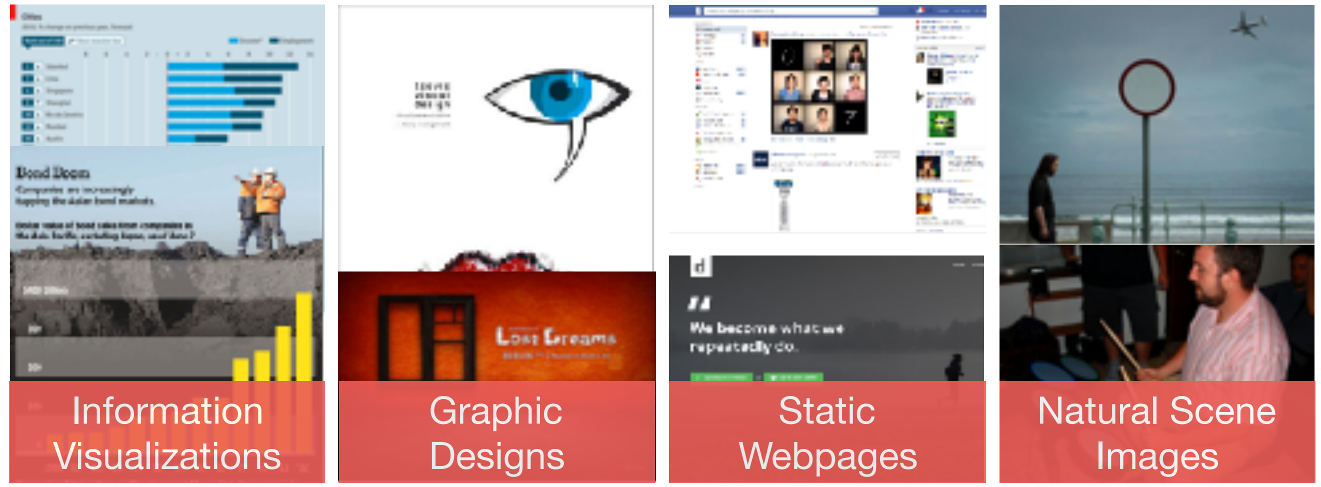


One of the main differences is that BubbleView uses discrete clicks instead of continuous movements to explicitly record points of interest.



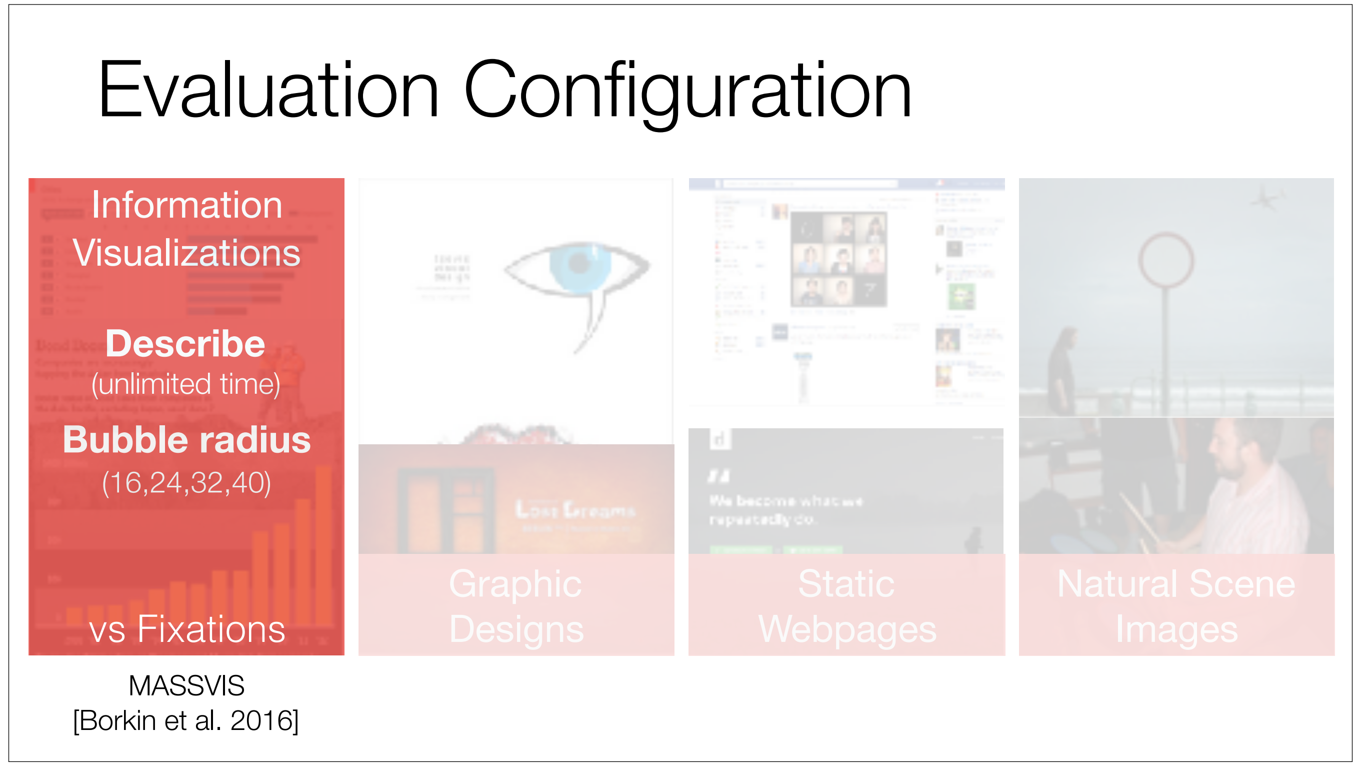
More importantly, unlike previous approaches, we systematically evaluate cursor-based tracking under different parameters and task settings including a comparison of mouse clicks and mouse movements.

Evaluated on Various Image Types

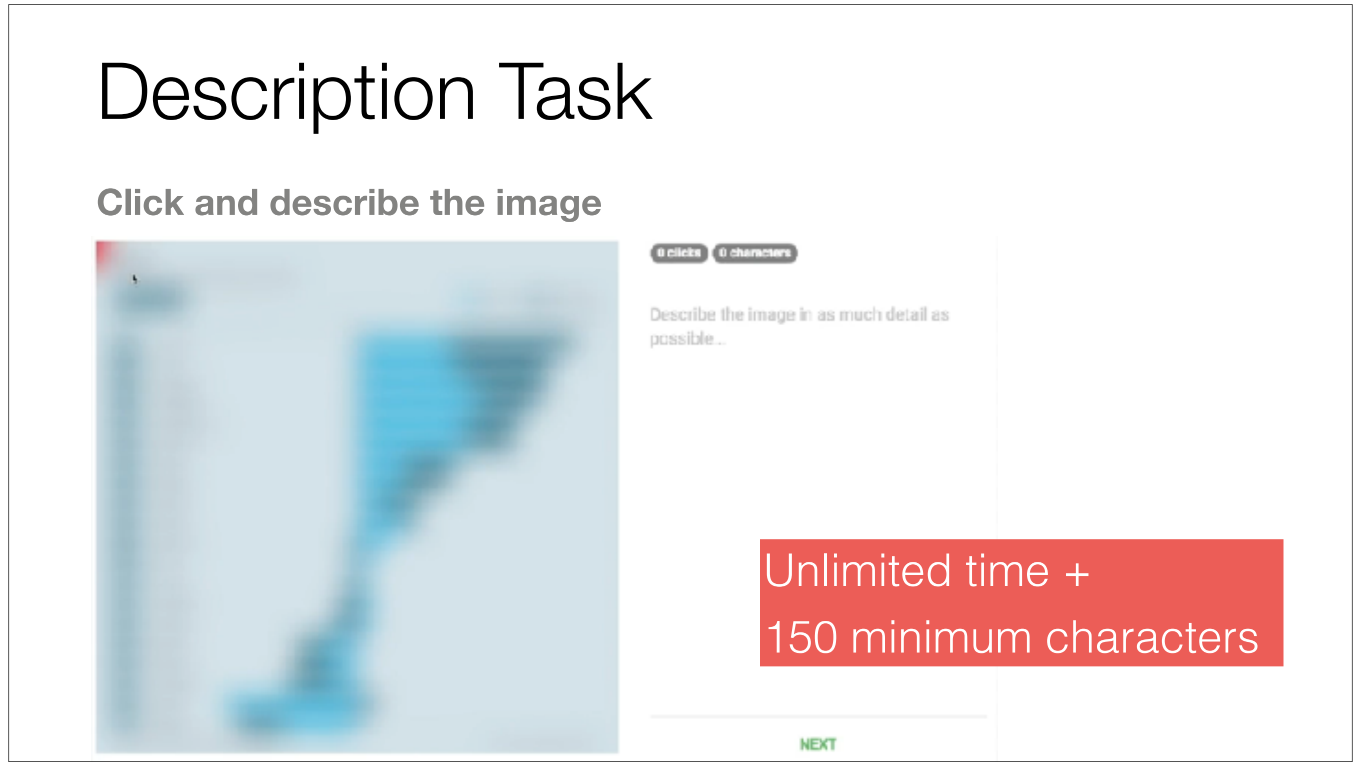


MASSVIS [Borkin et al. 2016] | GDI [O'Donovan et al. 2014] | FIWI [Shen and Zhao 2014] | OSIE [Xu et al. 2014]

We evaluated BubbleView on various image types: Information Visualizations, Graphic Designs, Static Webpages, Natural Scene Images. We collected these images from prior works.



For each image type we had a different parameter configuration. For information visualizations, we had a description task…



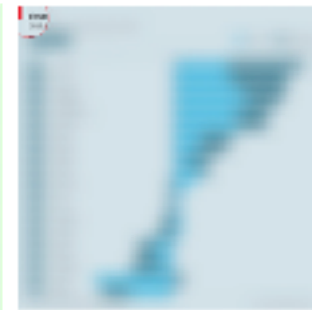
Participants used the BubbleView interface to click to see the image and they were asked to describe the image content. They had unlimited time for viewing the image but had to write a minimum of 150 characters in the description. We wanted to ensure they completed the task with enough thoroughness instead of randomly clicking around.

Varied Bubble Sizes

How does bubble radius size affect performance?

16 pixels

24 pixels

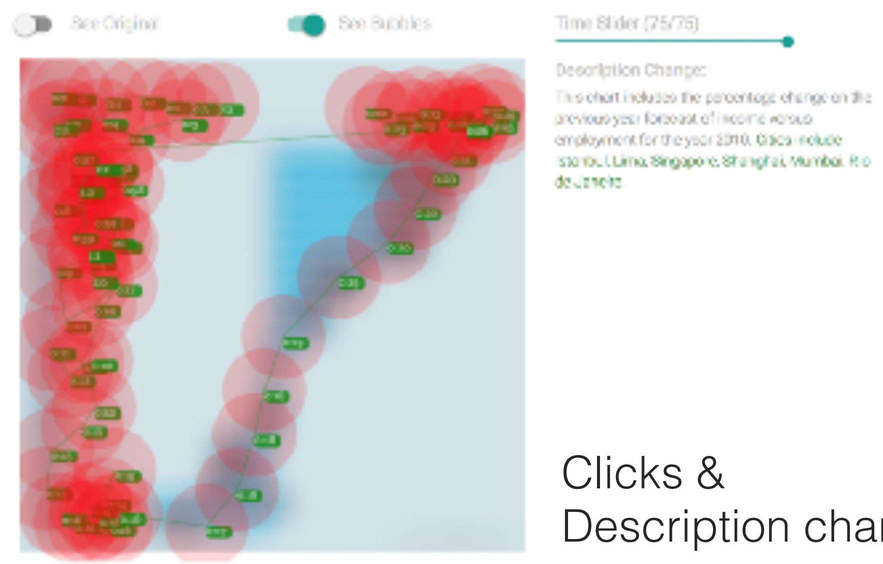


32 pixels

40 pixels

We also experimented with different bubble sizes to see how they affect task performance.

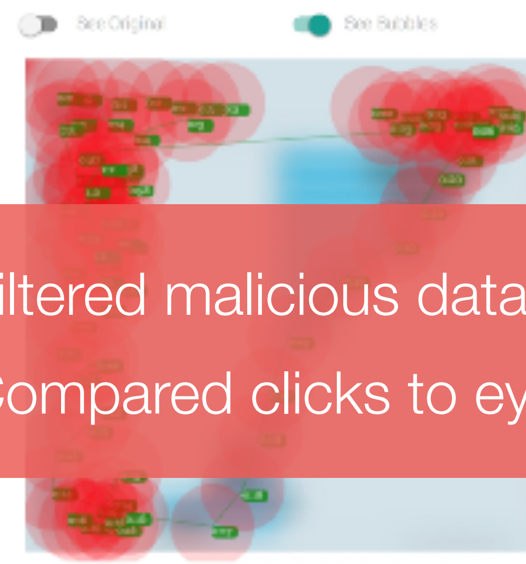
Collected Data



Clicks &
Description changes over time.

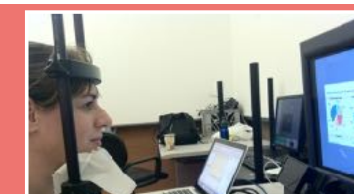
At the end, we collected mouse clicks with different bubble sizes as well as description changes over time.

Collected Data



Filtered malicious data &

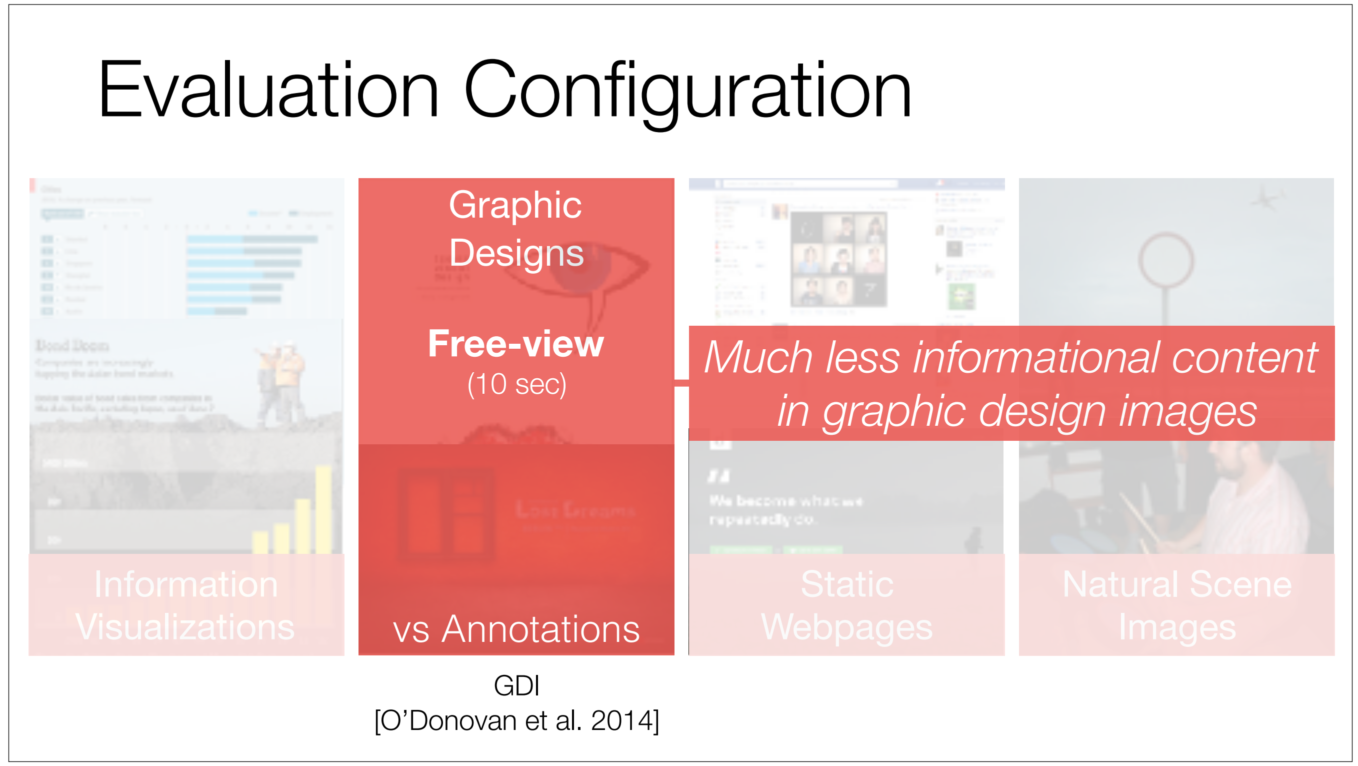
Compared clicks to eye-fixations



MASSVIS
[Borkin et al. 2016]

Clicks &
Description changes over time.

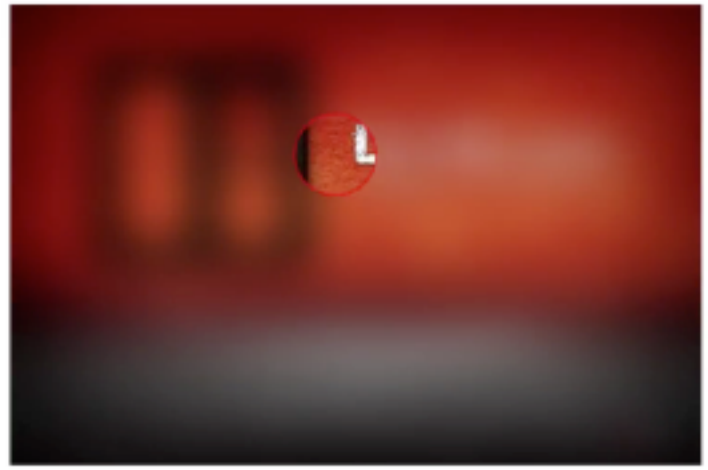
We then filtered malicious data and compared the resulting clicks to eye-fixations collected from an eye-tracker in the previous work.



For graphic designs, we used a free-viewing task because we found they have much less informational content.

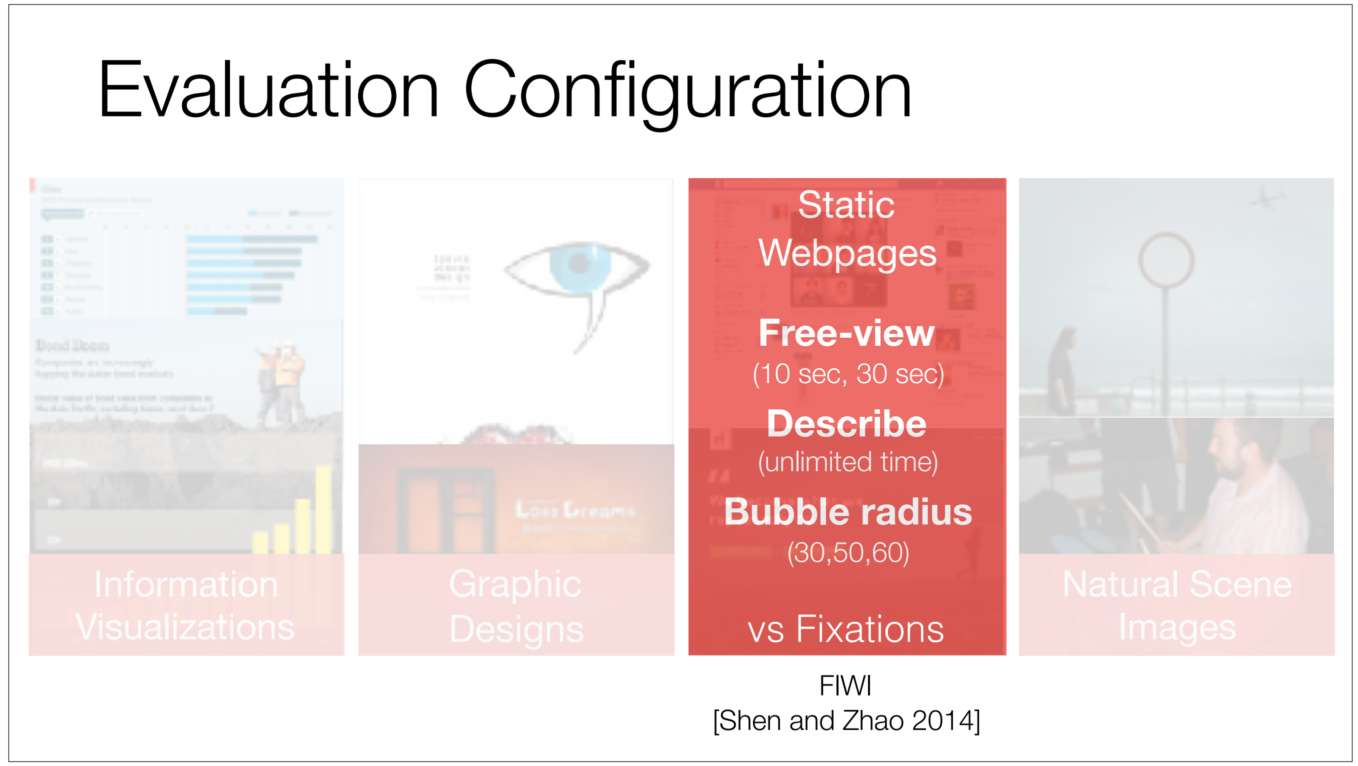
Free-Viewing Task

Click anywhere you want to look.

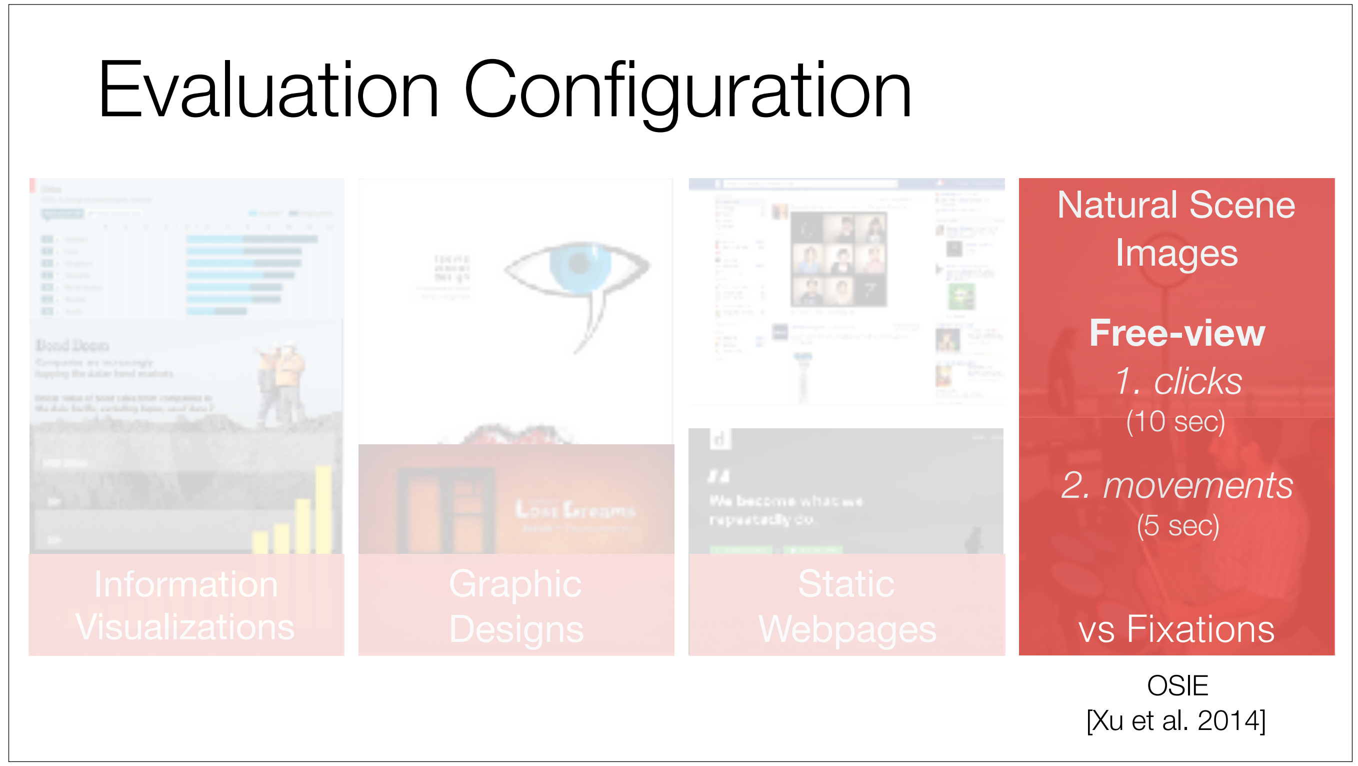


10 seconds of viewing
No description required

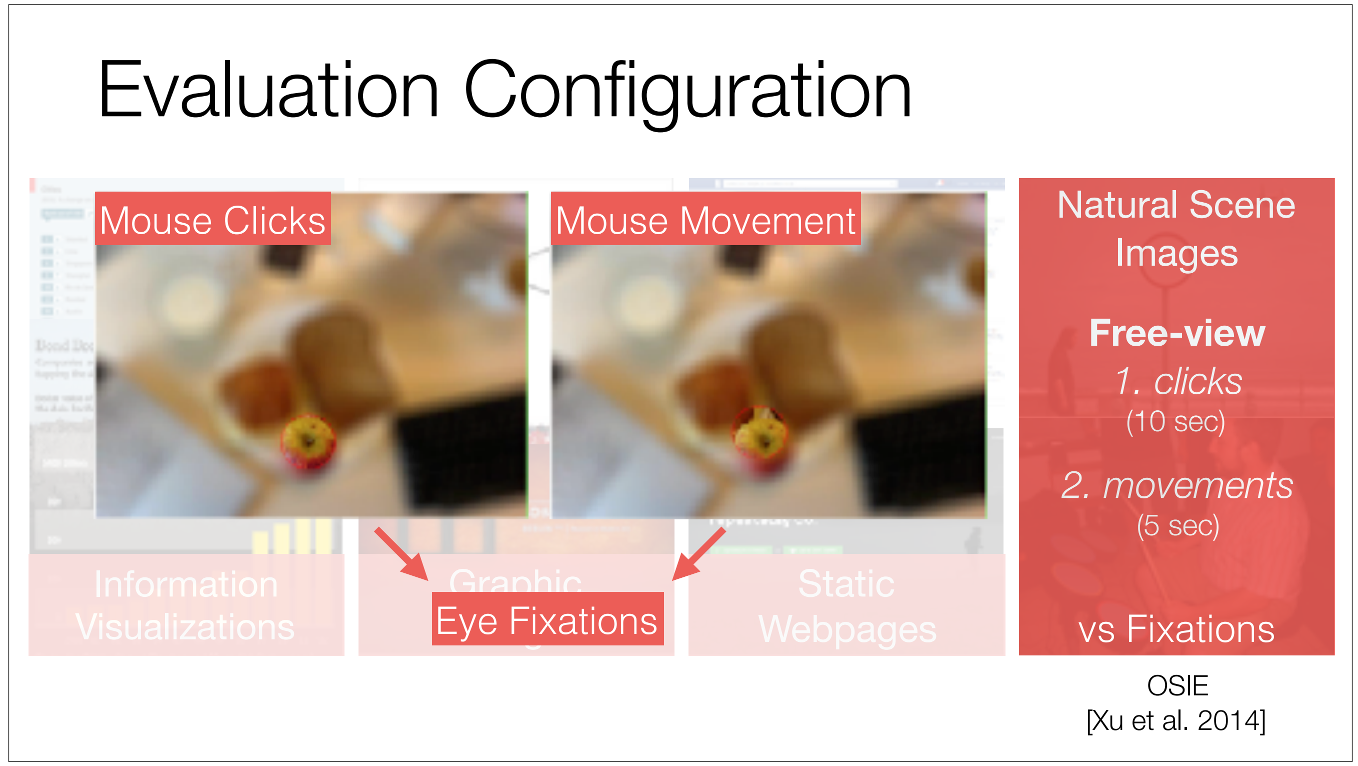
In the free-viewing task, participants can freely explore the image. They don’t have to describe the image. However we restricted the viewing time to 10 seconds



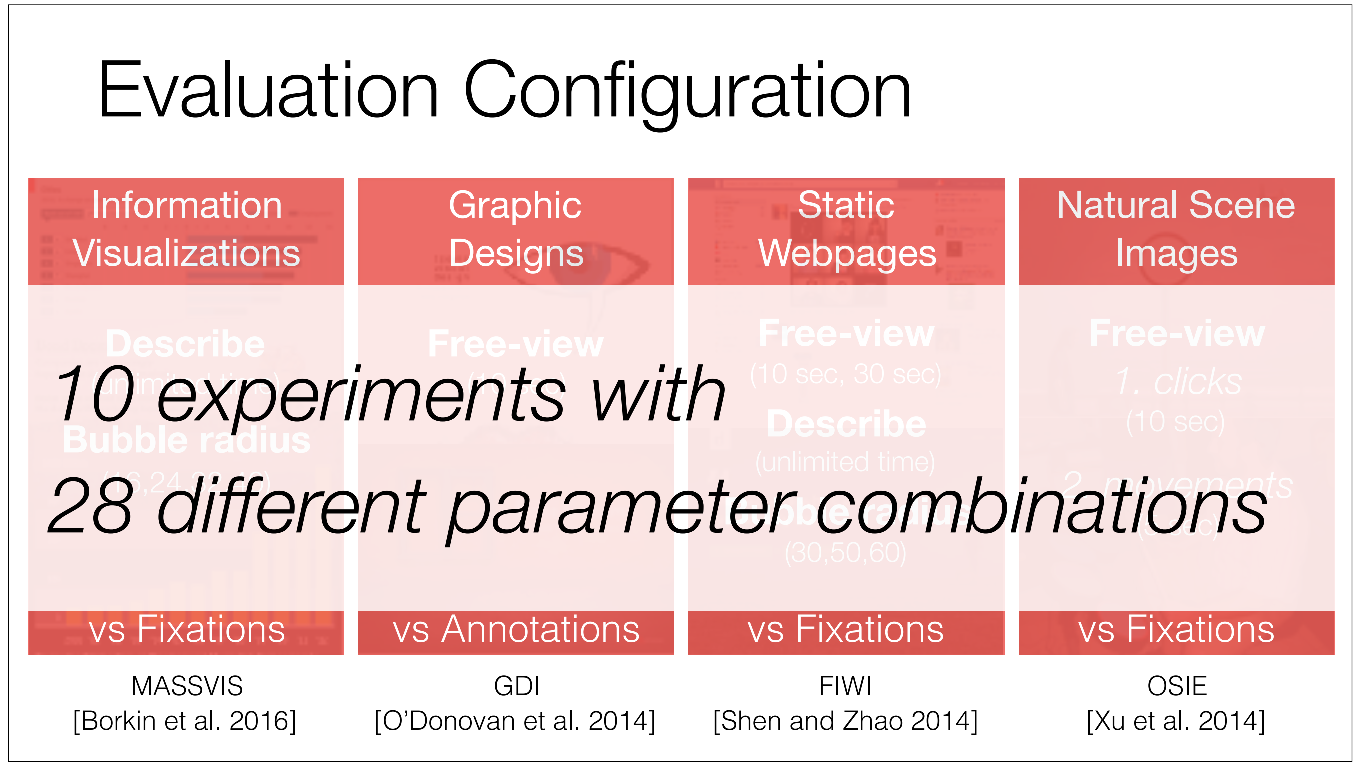
For static webpages, we used both free viewing and description tasks. We wanted to see which one is better in terms of approximating eye fixations.



For natural scene images, we only had a free-viewing task.

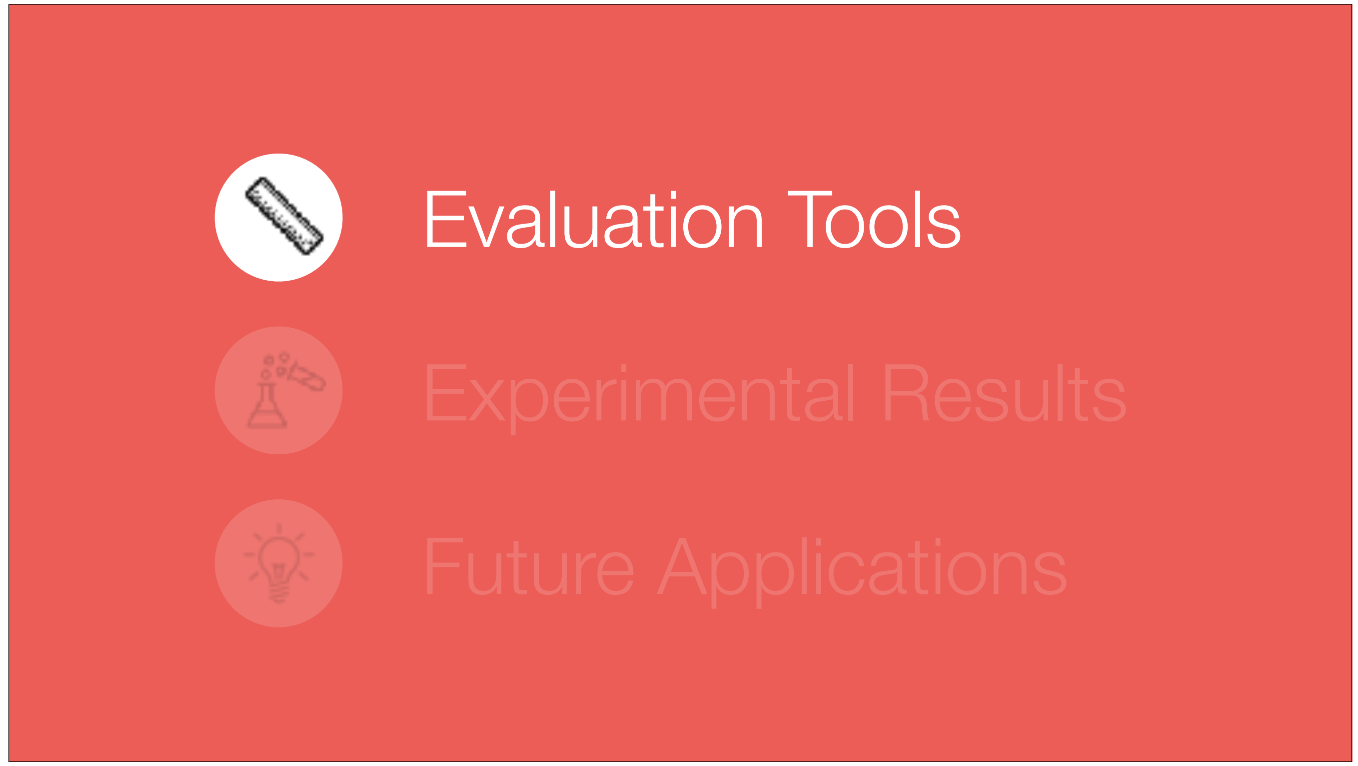


However, we vary the mouse modality, comparing mouse clicks and mouse movements, and analyzed how well each of them approximates eye fixations.

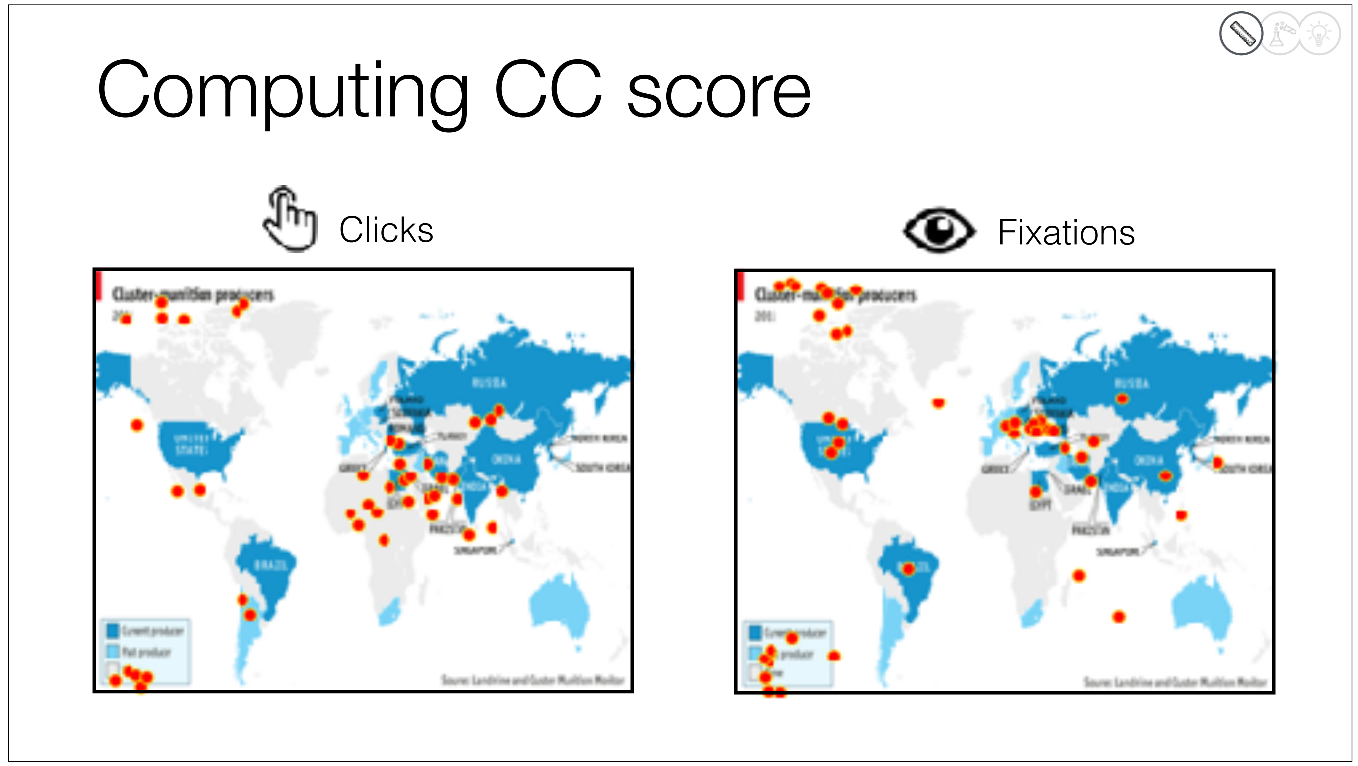


In total, we ran 10 experiments with 28 different parameter combinations in order to rigorously evaluate Bubble in various settings.

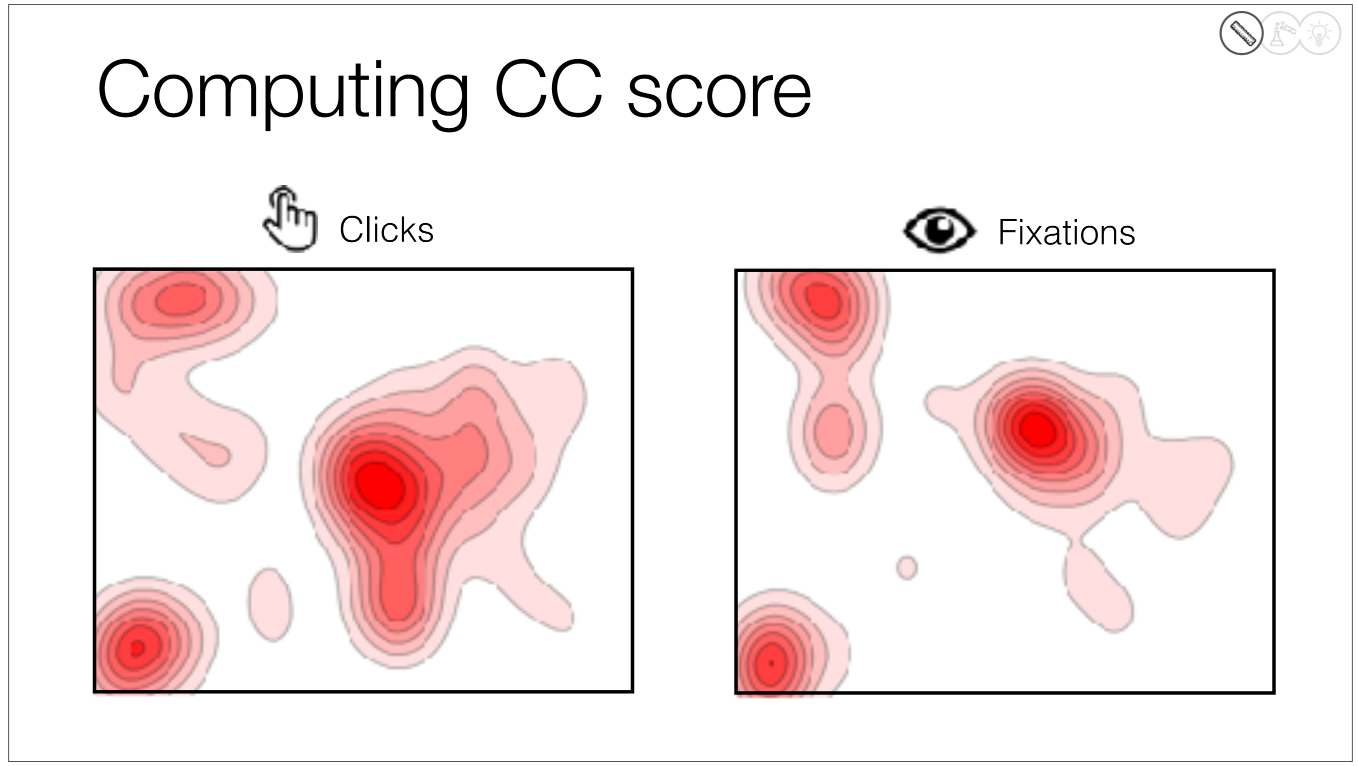
In this next section of the talk, I'll cover our evaluation, the high-level take-aways from all of our experiments, and hand it over to Nam for applications that can be built on top of BubbleView.



First, let's talk about how we compare clicks to fixations to determine if BubbleView can serve as a viable alternative to eye tracking.

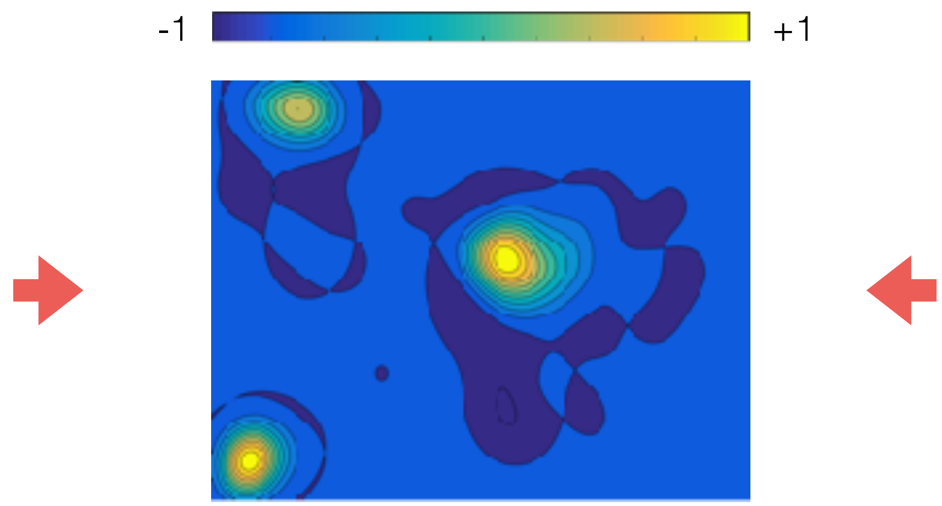


Both clicks and fixations produce a set of attention points on an image, so to robustly measure similarity,



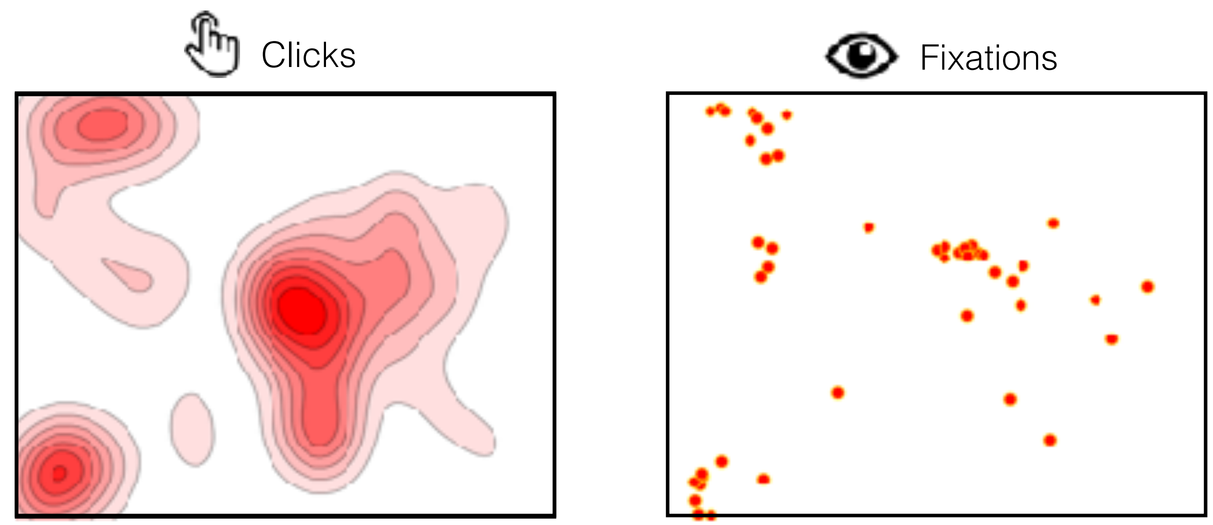
we can first smooth both sets of points into heatmap distributions.

Computing CC score



And compute a normalized cross correlation between the heatmaps, a common metric in saliency evaluation.

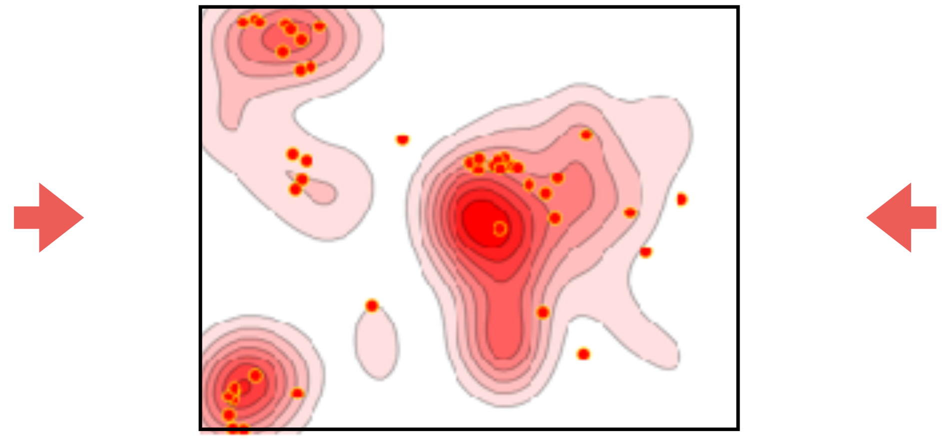
Computing NSS score



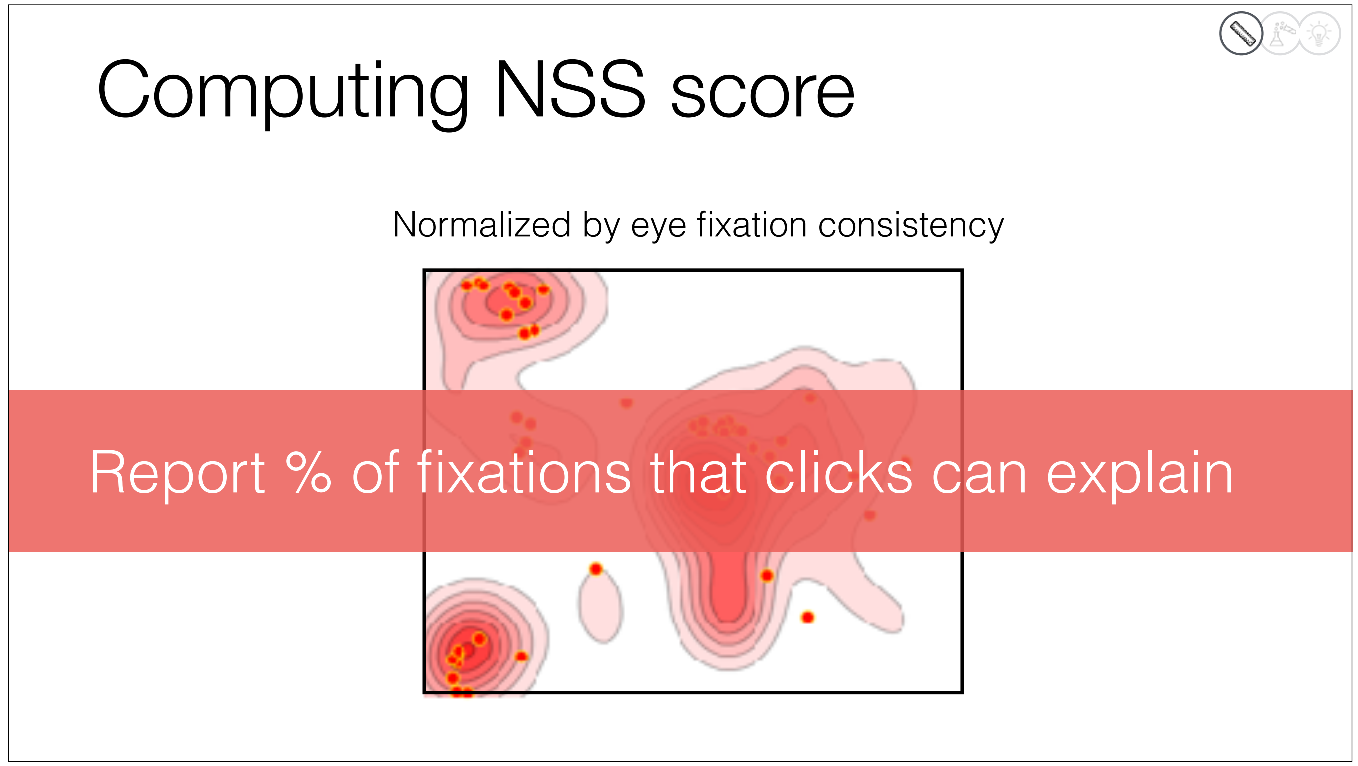
Another similarity score we use is the normalized scanpath saliency, which entails converting only the clicks to a heatmap

Computing NSS score

Normalized by eye fixation consistency



and measuring the average heatmap value at all the fixation points. In our study, [click] we normalize this score by human eye fixation consistency

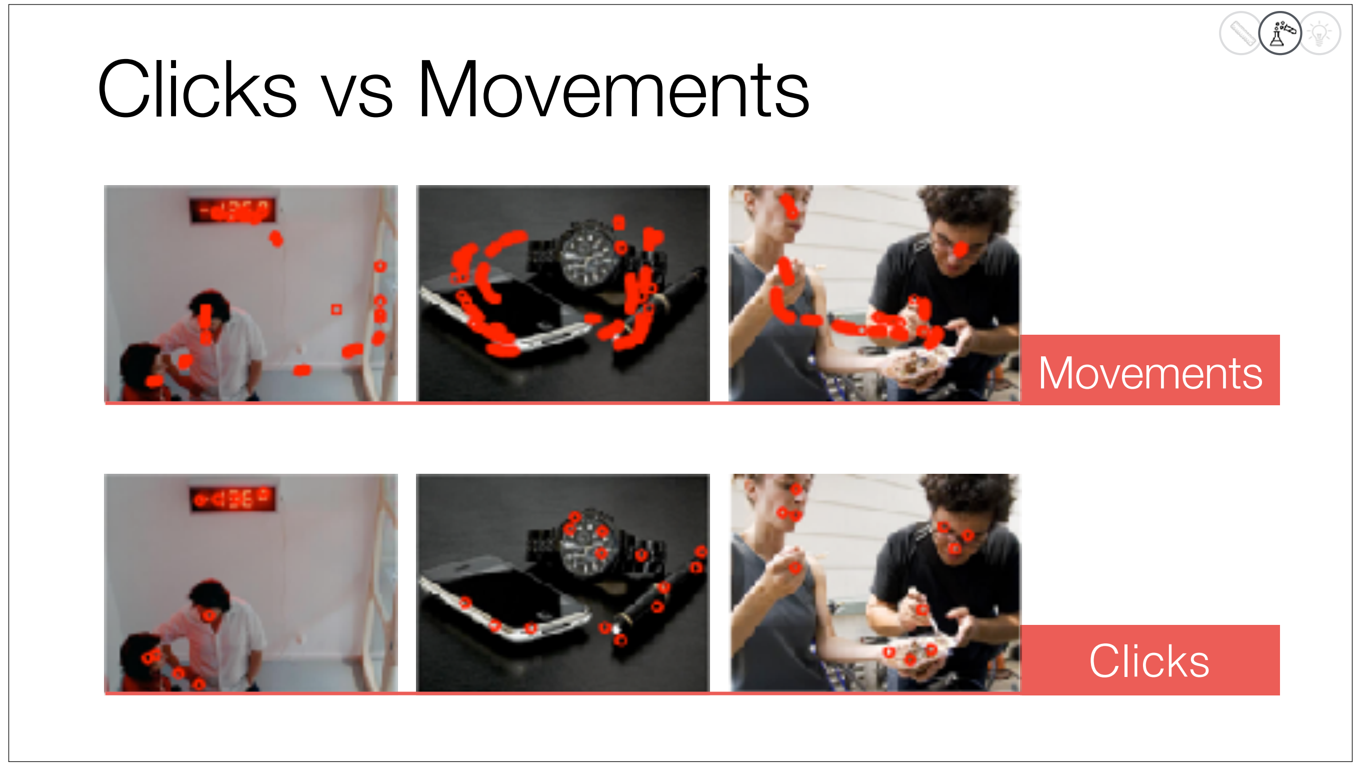


to report the % of fixations that clicks can explain relative to ground truth. This is the value we'll be reporting in this presentation. Detailed scores can be found in our paper.

Next, I’ll tell you about the 4 high-level take-aways across our 10 experiments. We refer you to the paper for thorough per-experiment results.

Take-away #1:

Clicks are more effective than mouse movements for measuring observer behavior.



Here’s an example of the attention data we obtain by capturing mouse movements and mouse clicks on the same images. In the top row, we see motion traces that are the byproducts of allowing a participant to freely move the mouse rather than just click. This would require careful thresholding and cleaning to capture the main points of interest. Instead we can use clicks directly as points of interest without post-processing.

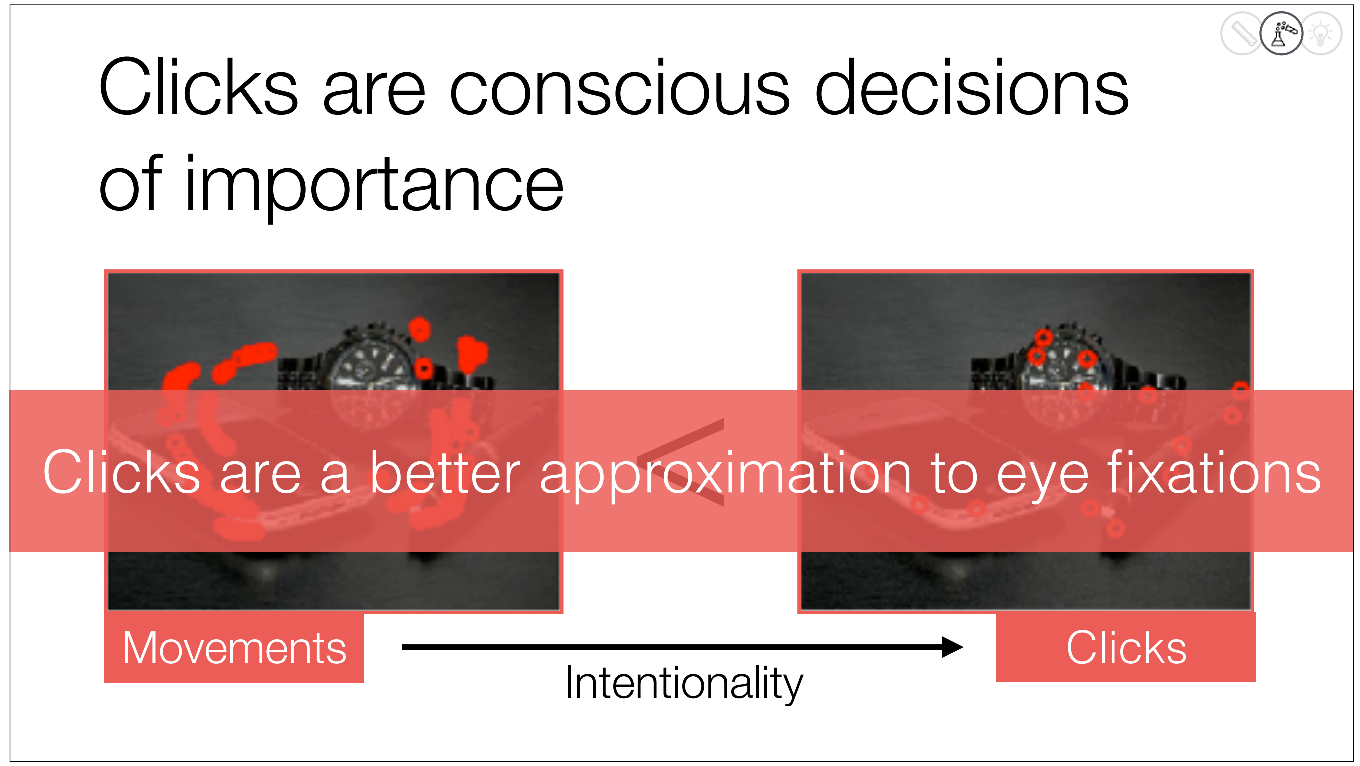
Clicks are conscious decisions of importance

$<$

Movements → Clicks

Intentionality

Clicks correspond to conscious decisions of importance. When a participant clicks somewhere, they’re telling us that they want to look there.

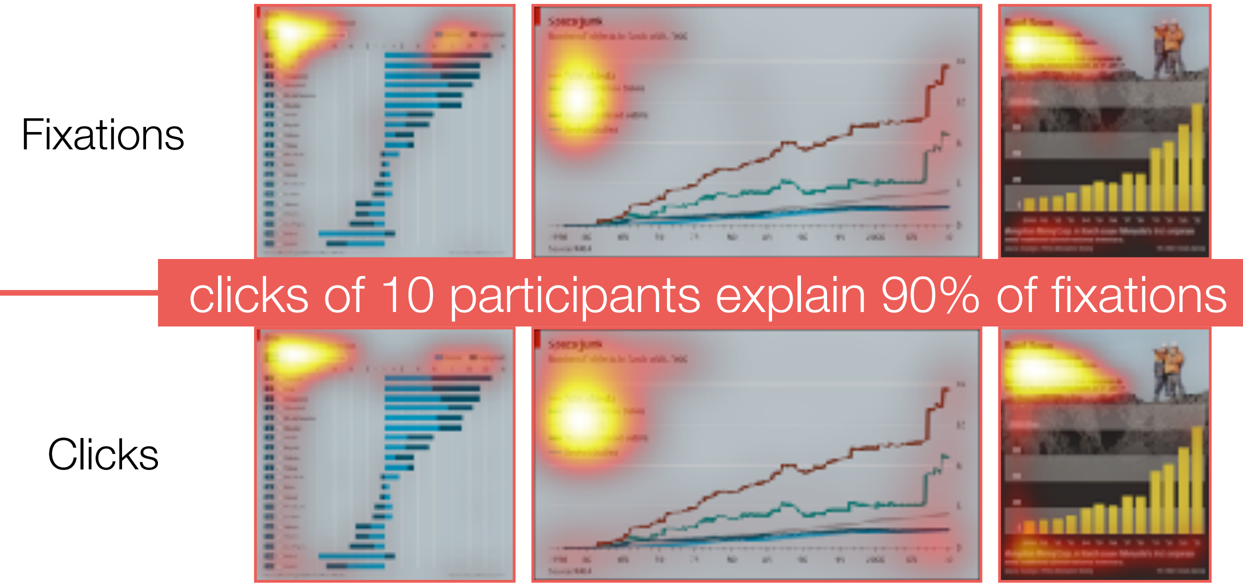


Quantitatively, we found that clicks better approximate eye fixations than mouse movements. This is partially because the mouse traces in mouse movements are not like natural eye movements anyway.

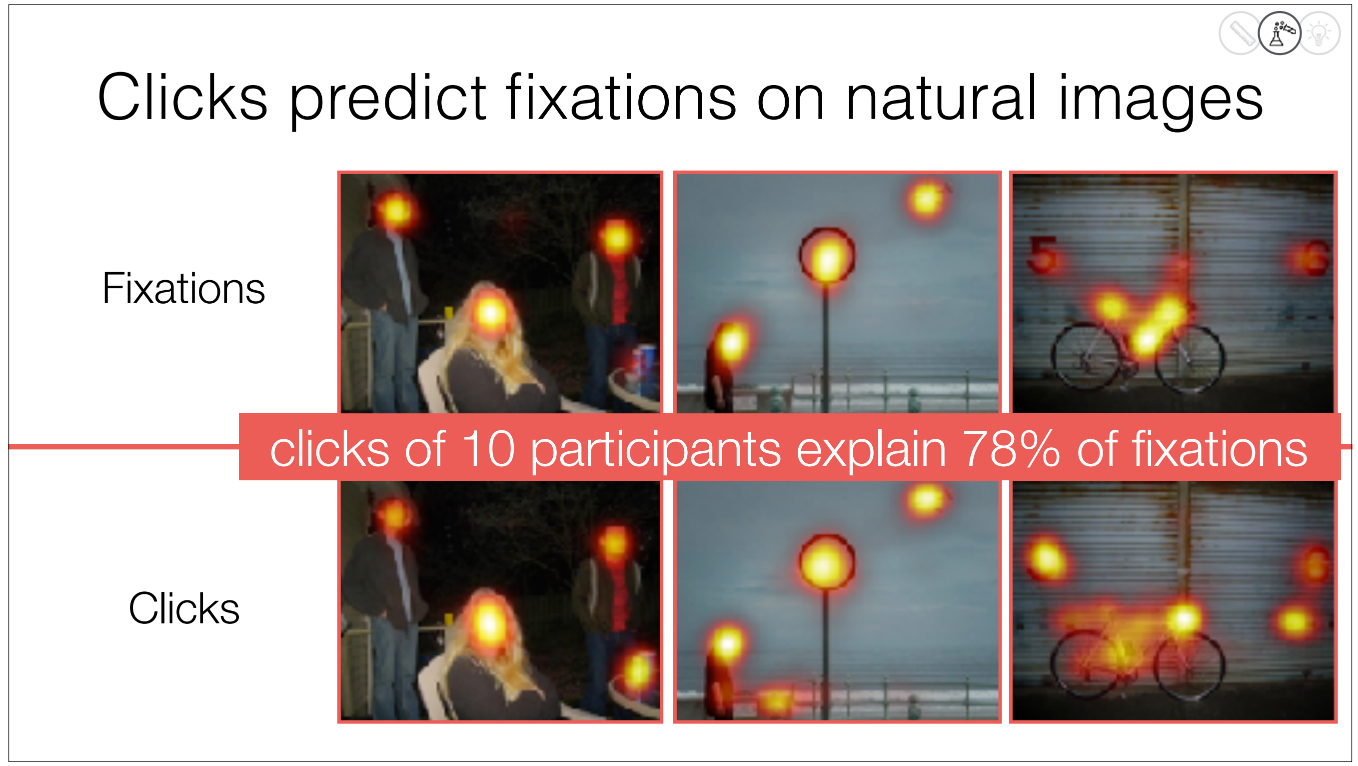
Take-away #2:

Clicks are predictive of eye fixations across a variety of image types and tasks.

Clicks predict fixations on visualizations

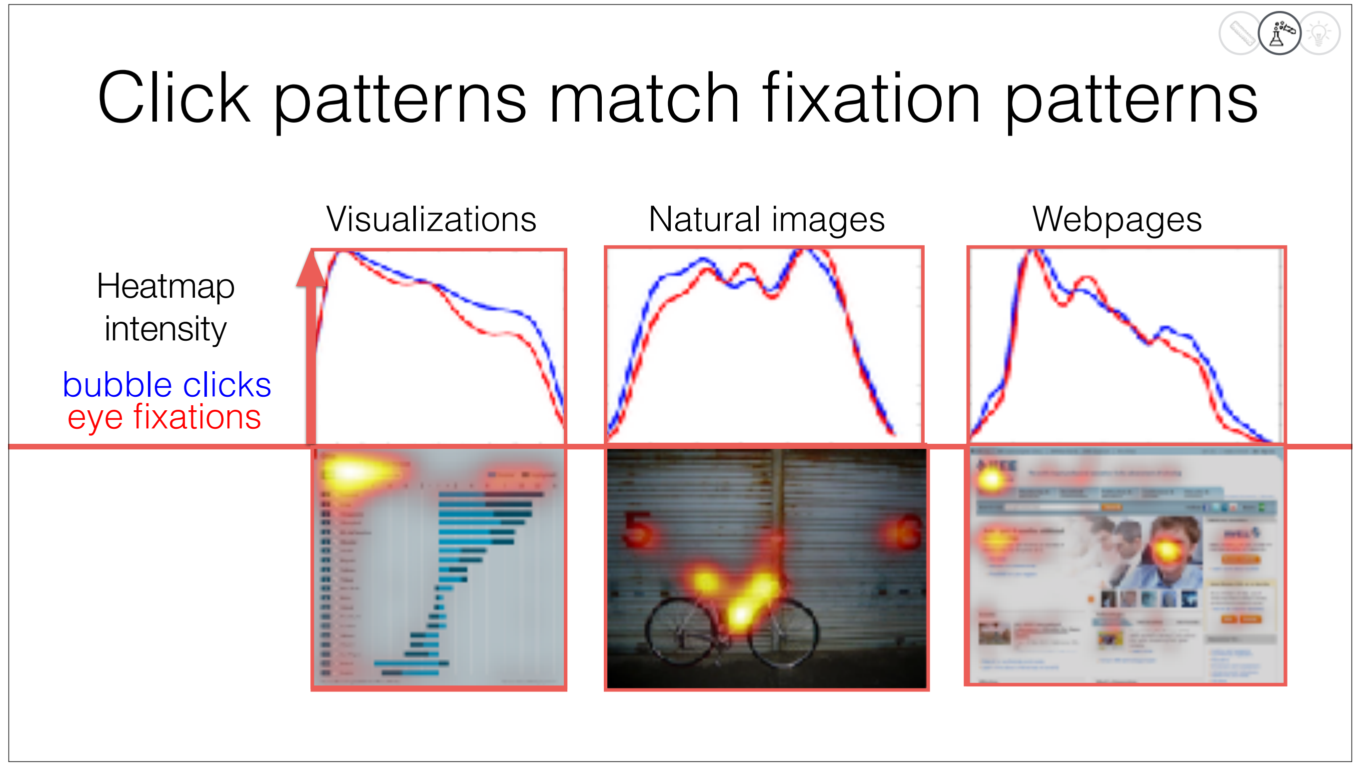


Here we show some examples of our fixation and click maps on visualizations. The heatmaps were computed by averaging and blurring the fixations/clicks of all participants.
With a well-defined description task we found that…

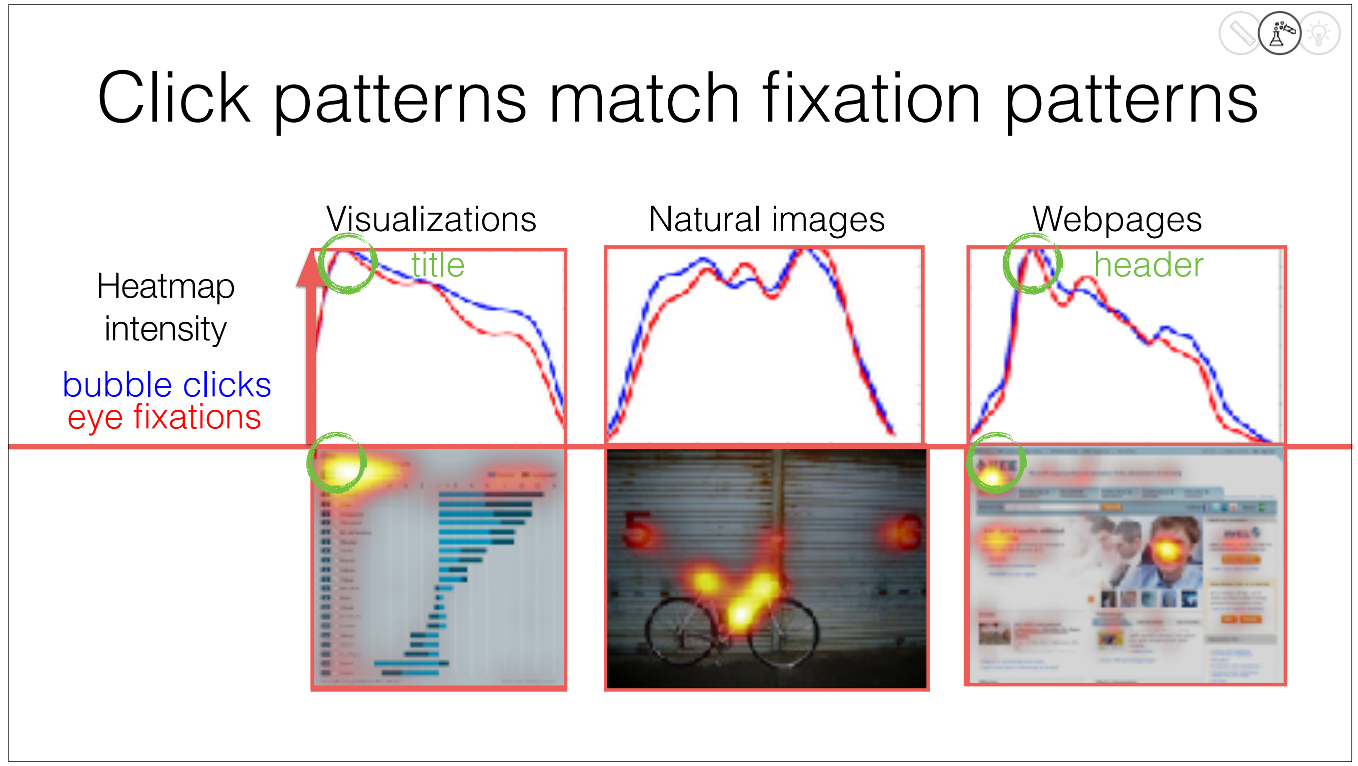


We also compared clicks to fixations on natural images with a free-viewing task. Because free-viewing is a less constrained and less defined task, we begin to see some differences between clicks and fixations. Overall, … , which is high enough to serve as an approximation for eye tracking for many applications.

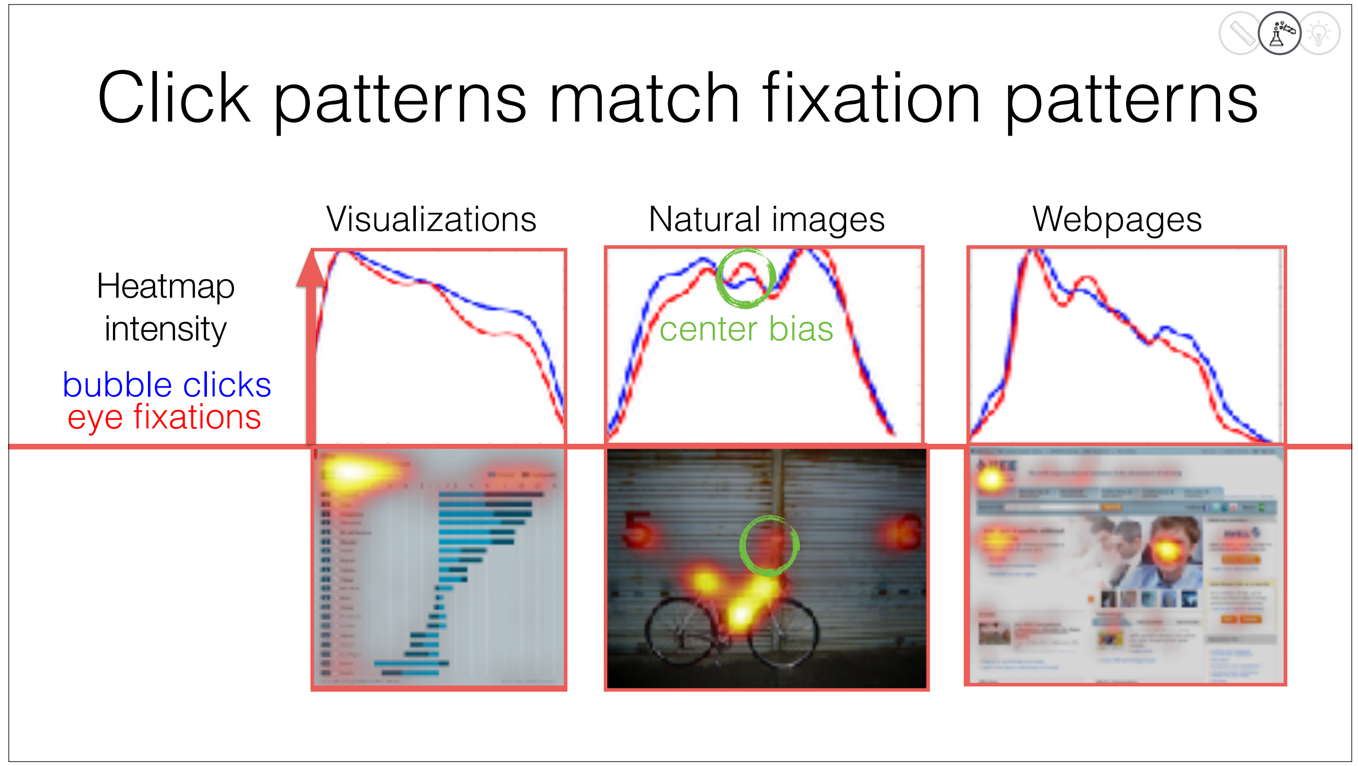
We collected clicks on webpages with both description and free-viewing tasks. Quantitatively, … , with both tasks. This is lower than for visualizations because the webpages tended to contain a lot more elements for people to explore, lowering overall consistency across participants.



Where are some of the similarities and differences? Let's take a look at the intensity of the fixation and click maps by location on an image. Now this is not for a specific image, but rather averaged across all the images in our datasets. The peaks in these plots show the x coordinates with the most fixations or clicks.

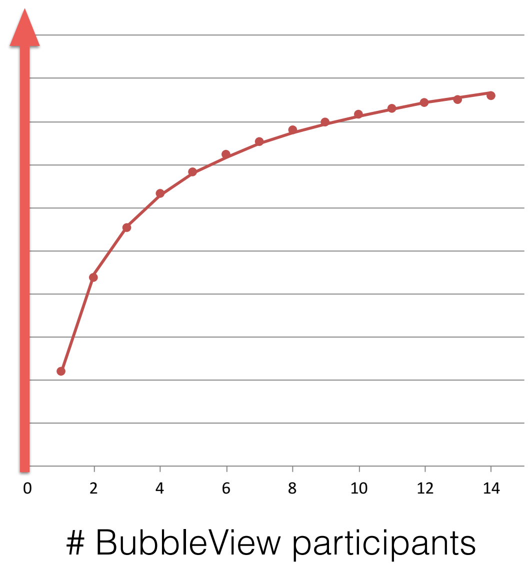


We can see that many fixations and clicks land at the top left of visualizations, which often corresponds to the location of the title; and the top left region of webpages, which corresponds to the location of the webpage header. Overall, the distributions of the click patterns match the the fixation patterns.



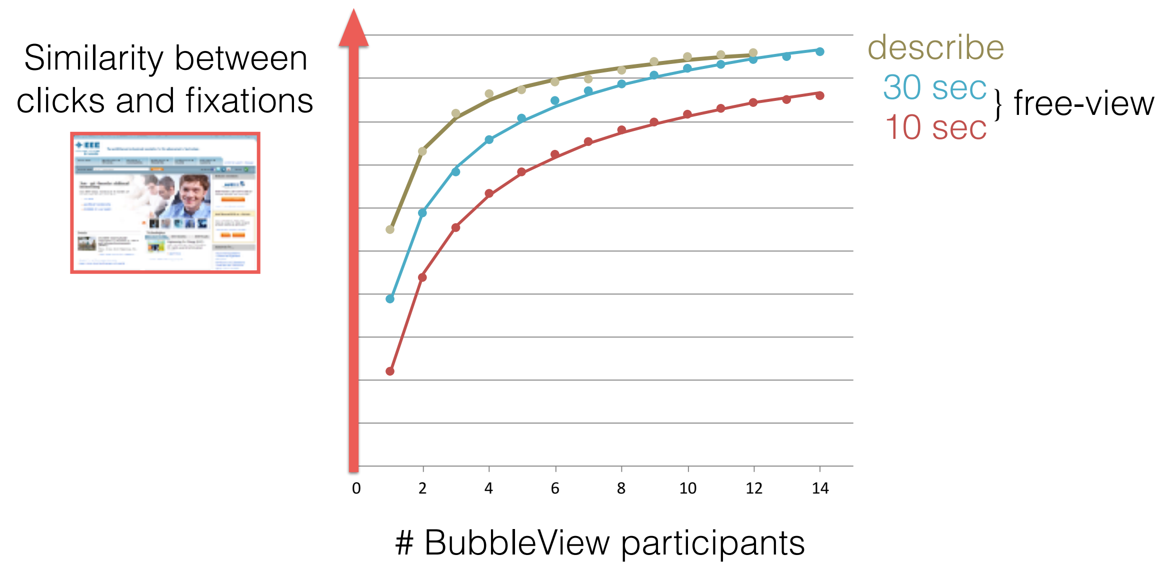
However, eye fixations have a known bias towards the center of the image. This is an aspect of attention behavior that is image-independent and often unconscious. Clicks avoid this bias, as they tend to be more conscious decisions of where to attend next.

More involved tasks lead to better clicks



In this plot, we see how the similarity between clicks and fixations increases as we capture the clicks of more BubbleView participants. In the paper, we show that this begins to saturate after about 10-15 participants.

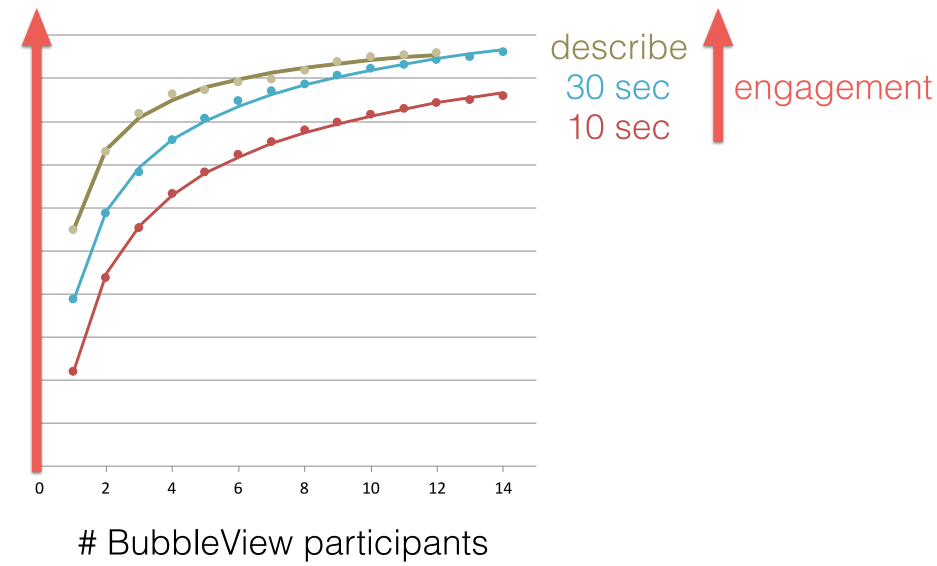
More involved tasks lead to better clicks



We found that clicks approximate fixations better when participants are given a longer viewing time during the free-viewing task, or the more constrained task setting of describing images.

More involved tasks lead to better clicks

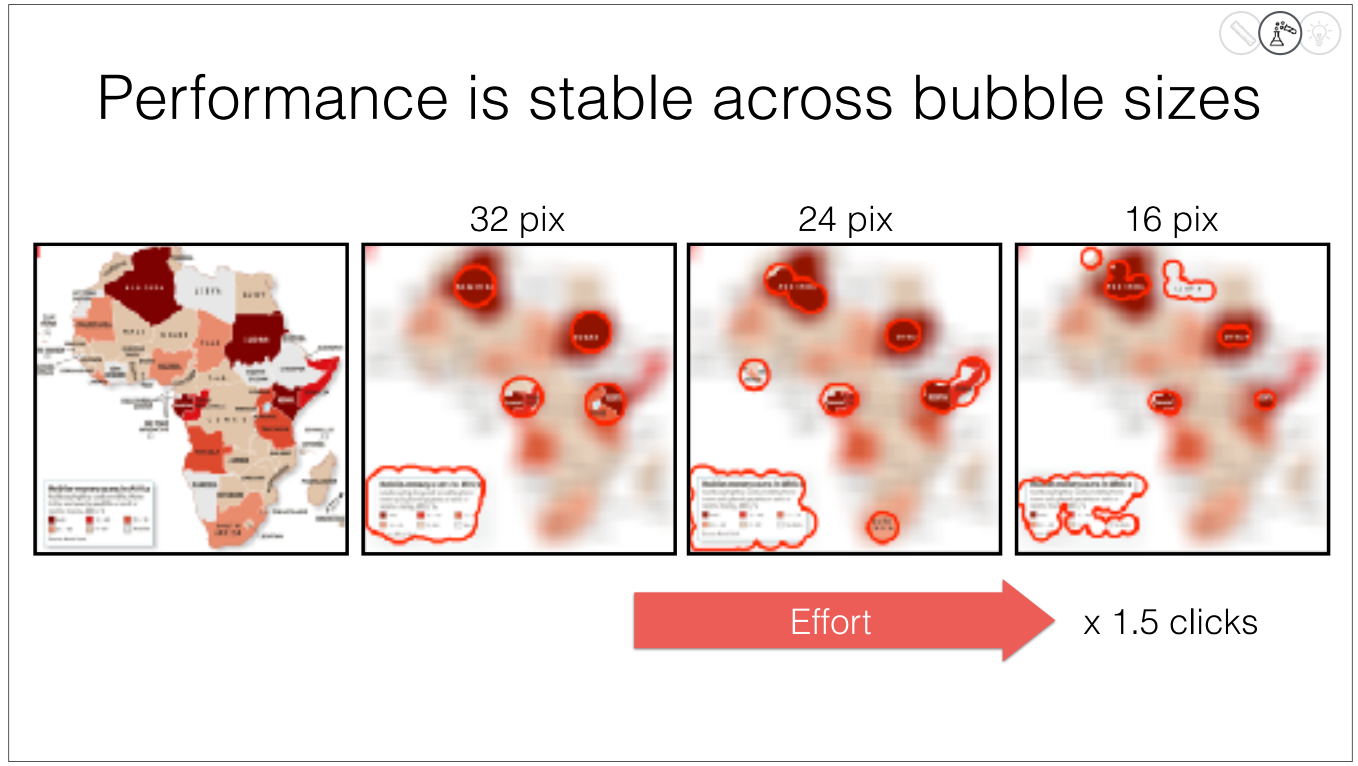
Similarity between clicks and fixations



When click participants are more engaged in the task, we get better quality clicks. When smaller numbers of participants are available, we recommend a longer task duration or more constrained task to get better results.

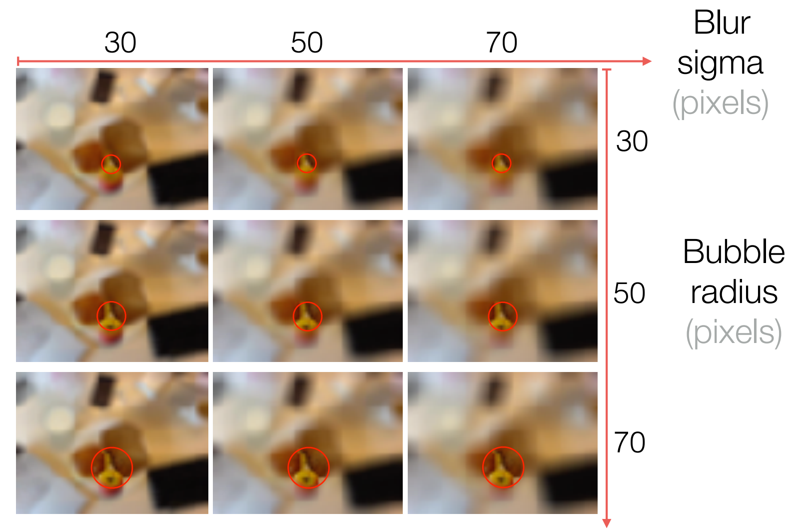
Take-away #3:

Task time and bubble size interact to affect clicks.



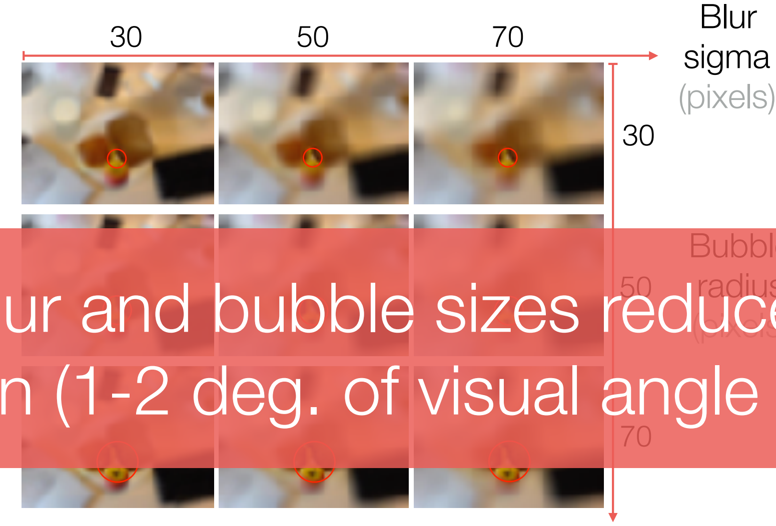
Testing our interface with different bubble sizes on visualizations, we found participants end up clicking on the same portions of the visualization overall with more effort for the smallest bubble size. Because we found no statistically significant differences in the similarity scores across the bubble sizes, we recommend avoiding bubble sizes that are too small to be comfortable within a UI.

Blur affects clicks more than bubble size



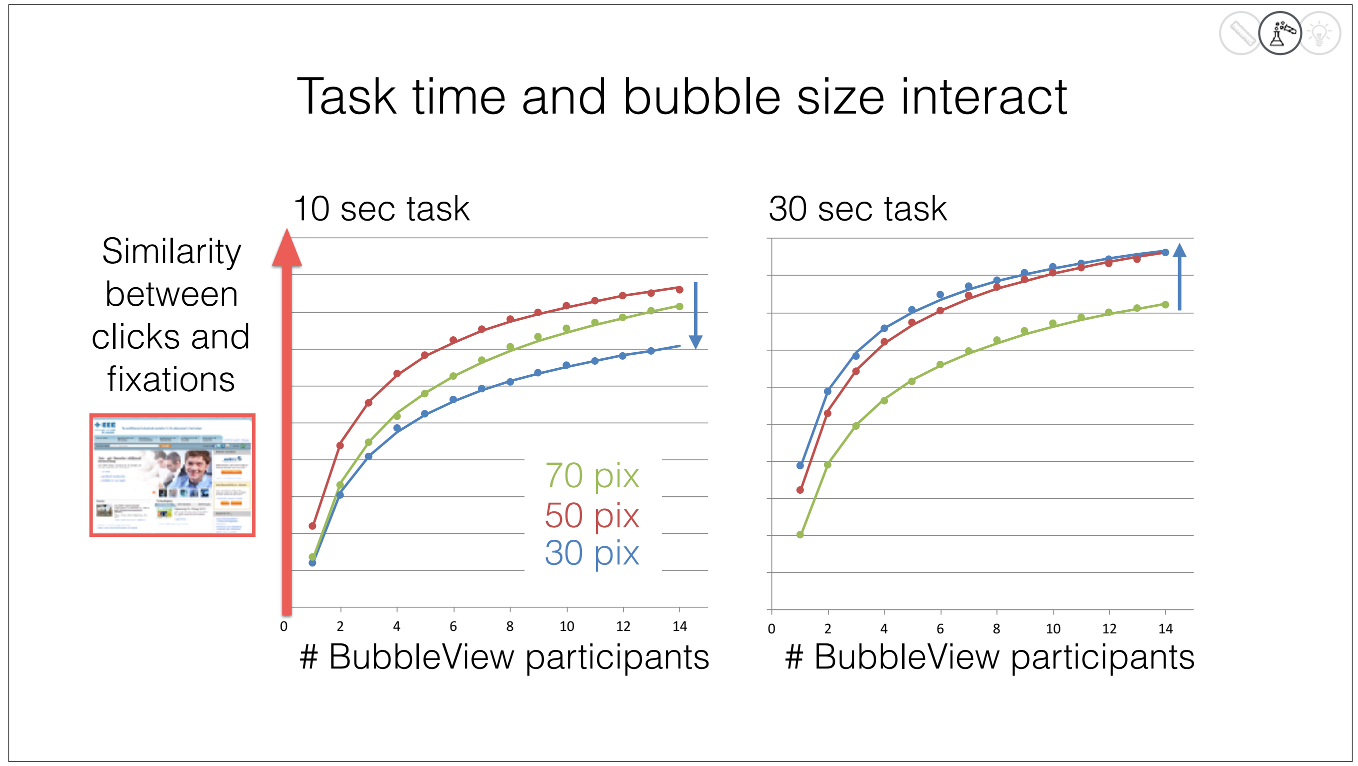
We also jointly varied both bubble size and blur amount on natural images to determine the settings to best approximate natural viewing.

Blur affects clicks more than bubble size



Largest blur and bubble sizes reduce exploration (1-2 deg. of visual angle best).

Again, we found no statistically significant differences across bubble sizes, although largest bubble size started to deviate from fixation patterns. We found that the largest blur amounts reduced exploration and led to significantly worse scores. We recommend keeping both bubble and blur size within 1-2 degrees of visual angle (this was about 30-50 pixels on our images of size 500 pixels). More precise recommendations can be found in our paper.

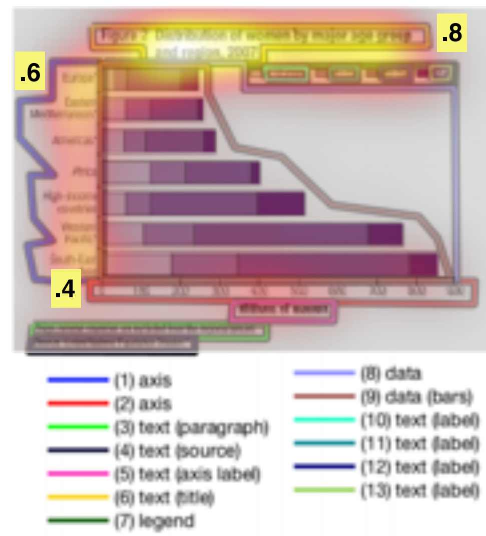


On our experiments with webpages, we found that bubble size interacts with task timing. The smallest bubble size of 30 pixels [click] was worst at approximating fixations with 10 sec of viewing [click], but did best with 30 seconds of viewing [click]. For best overall results, we recommend a longer task duration [click] but paired with a smaller bubble size.

Take-away #4:

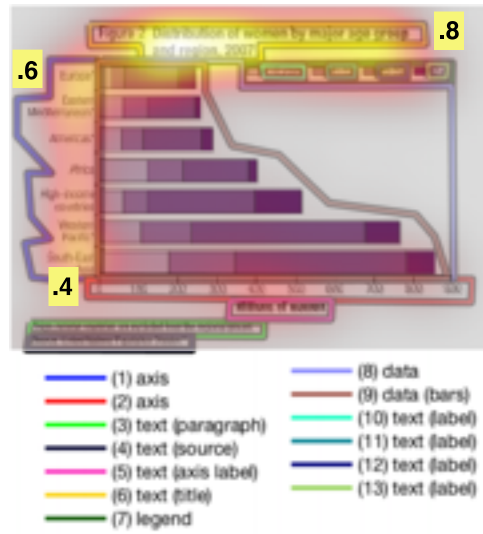
BubbleView can be used to rank image elements by importance.

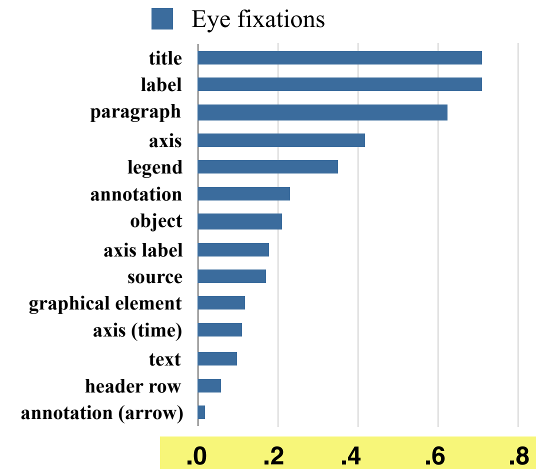
Ranking elements by importance



We analyzed the similarity of clicks and fixations across visualization elements. For this, we used element annotations for title, axis, etc. on the visualizations, which we overlapped with the fixation maps [click] and aggregated the maximum heatmap value per element to get an importance score for that element (from 0 to 1).

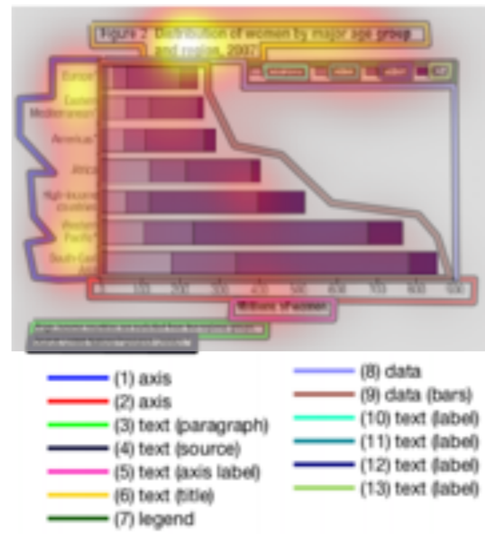
Ranking elements by importance

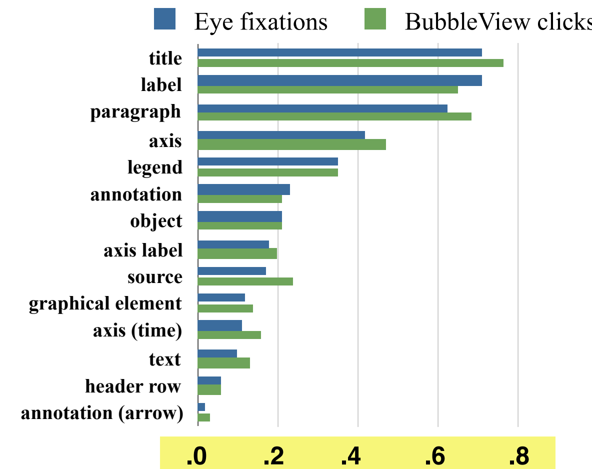




When we average these scores across all the visualizations in our dataset, we get a ranked list of visualization elements, sorted by “importance”.

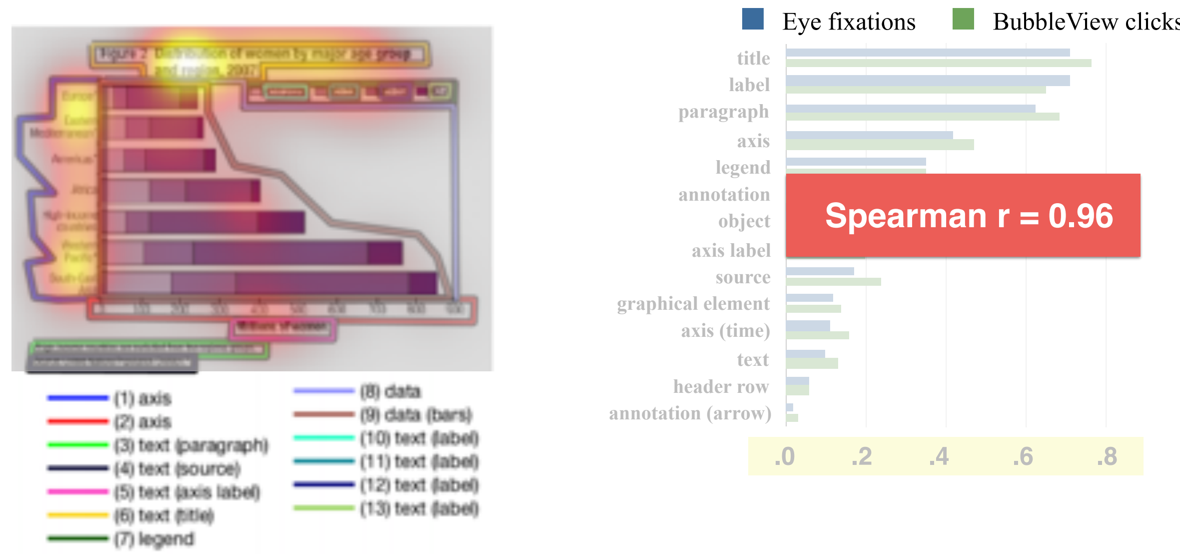
Ranking elements by importance



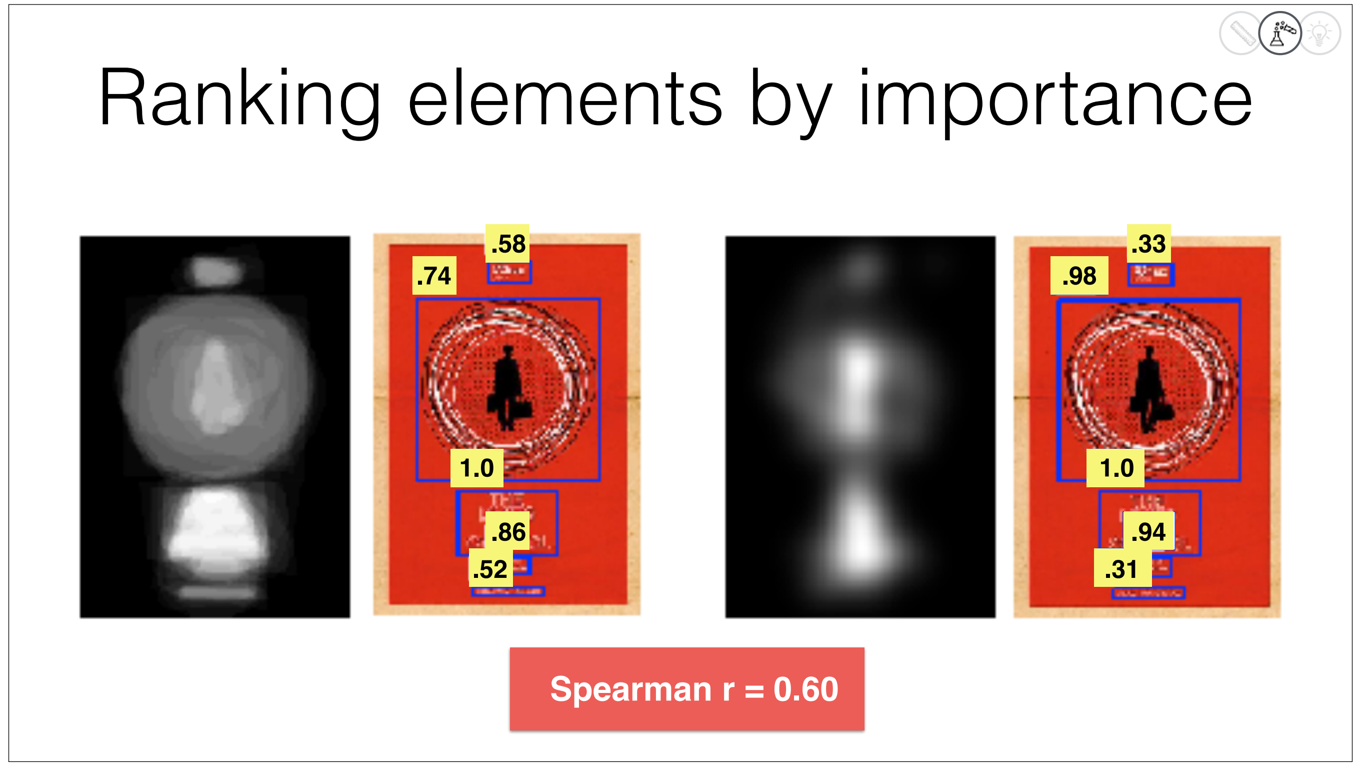


Repeating the same analysis for BubbleView click maps, we see the same pattern emerge: the same visualization elements receive the highest scores.

Ranking elements by importance



Overall, the rank correlation between the eye fixations and bubbleview clicks in assigning importance values to visualization elements is 0.96.

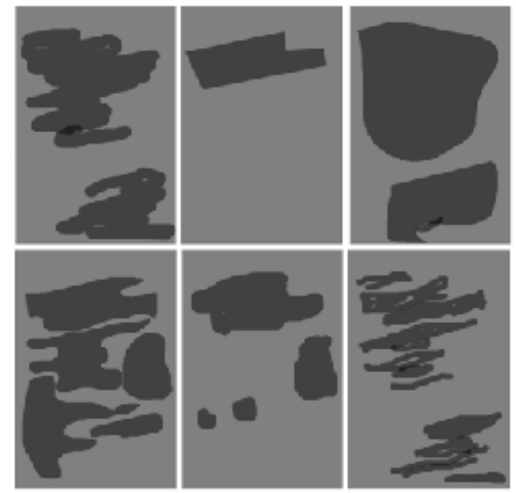


We did a similar analysis for the graphic designs, segmenting graphic designs into individual elements and aggregating heatmap values per element. On an per-design basis we compared the ground truth and bubbleview importance scores. The average spearman correlation is 0.6, which is lower than for visualizations, but sufficient for design applications, which we'll demo in a few slides.

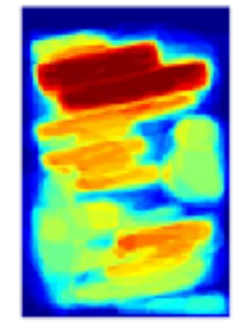
Another measurement of importance

Graphic design

Crowd annotations

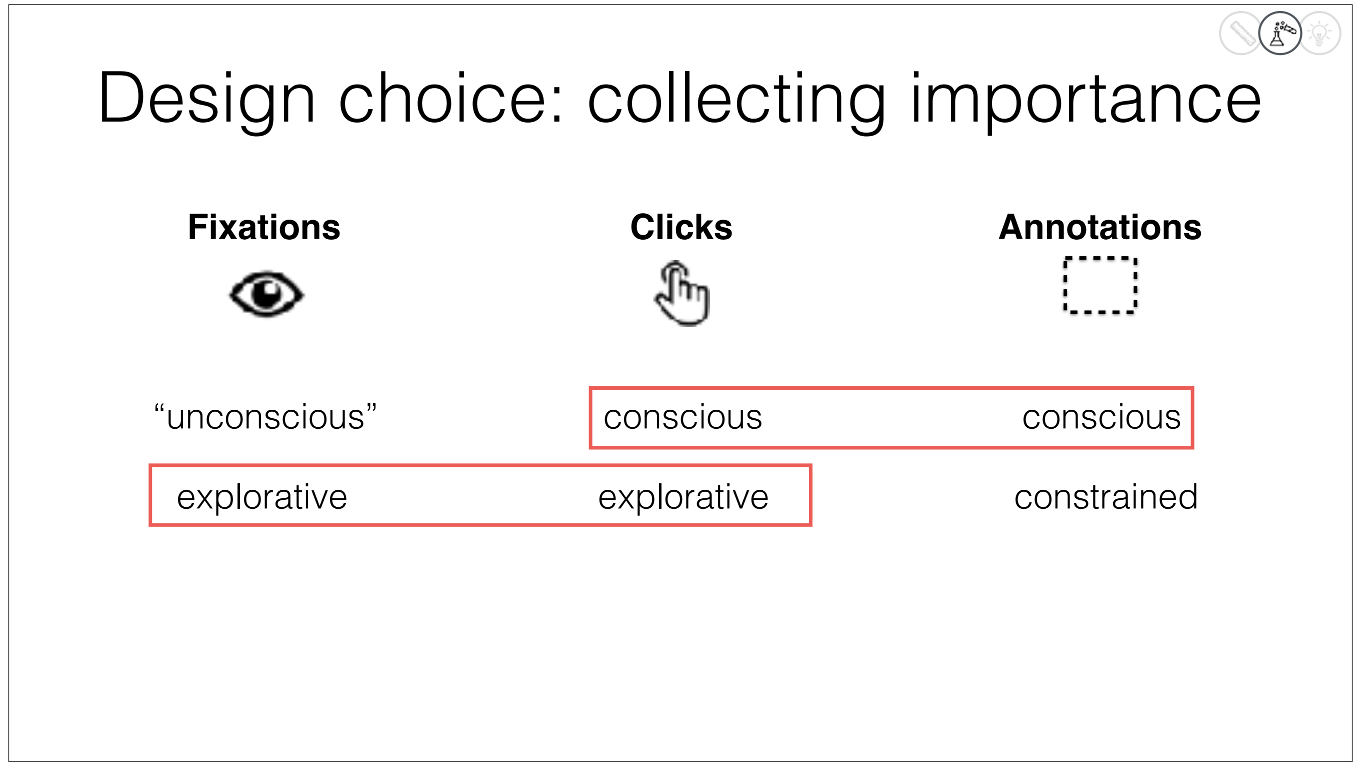


Avg. annotation

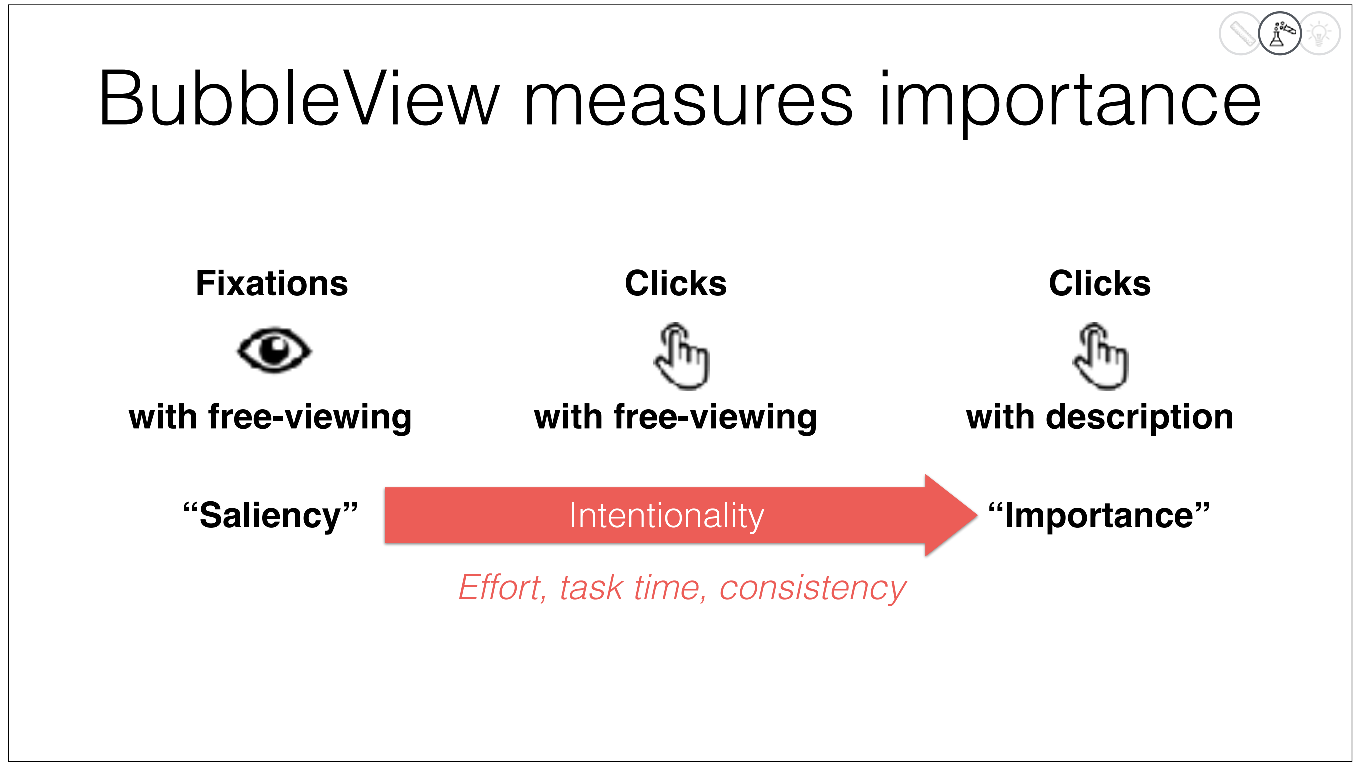


O'Donovan et al. [TVCG'14]

We note that for graphic designs, the ground truth importance did not come from eye fixations, but rather explicit annotations of importance. Participants were asked to draw binary masks over design regions they found important. The overall importance heatmap is obtained by averaging these annotations.



This leads to a data design choice: how to collect importance on images? This can be done with eye movements, BubbleView clicks, or explicit annotations. Clicks have a nice trade-off. Unlike fixations, both clicks and annotations invoke slower cognitive processes and provide more conscious decisions of importance. However, fixations and clicks are less constrained than annotations and capture the regions of images people choose to explore in a more natural setting.



Overall, as we move from fixations with free-viewing to clicks with description, the intentionality of participant behavior increases. This comes with increasing task effort and time, but as a result higher consistency in behavior. Whereas fixations with free-viewing have traditionally been used as measures of saliency, we associate clicks with constrained tasks like description, as a measure of importance. And it is this importance that we can build applications on top of.

Evaluation Tools

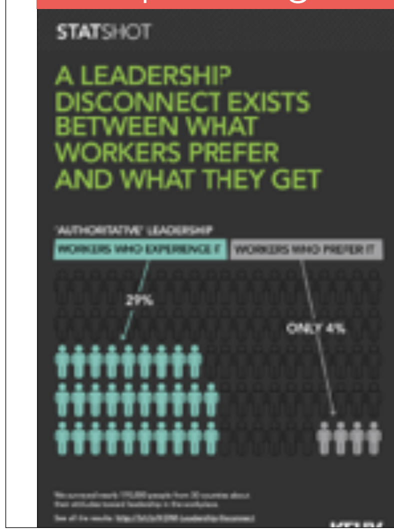
Experimental Results

Future Applications

Nam: I will now briefly preview some applications that can be built on top of BubbleView clicks and the notion of importance.

Retargeting & Thumbnailing

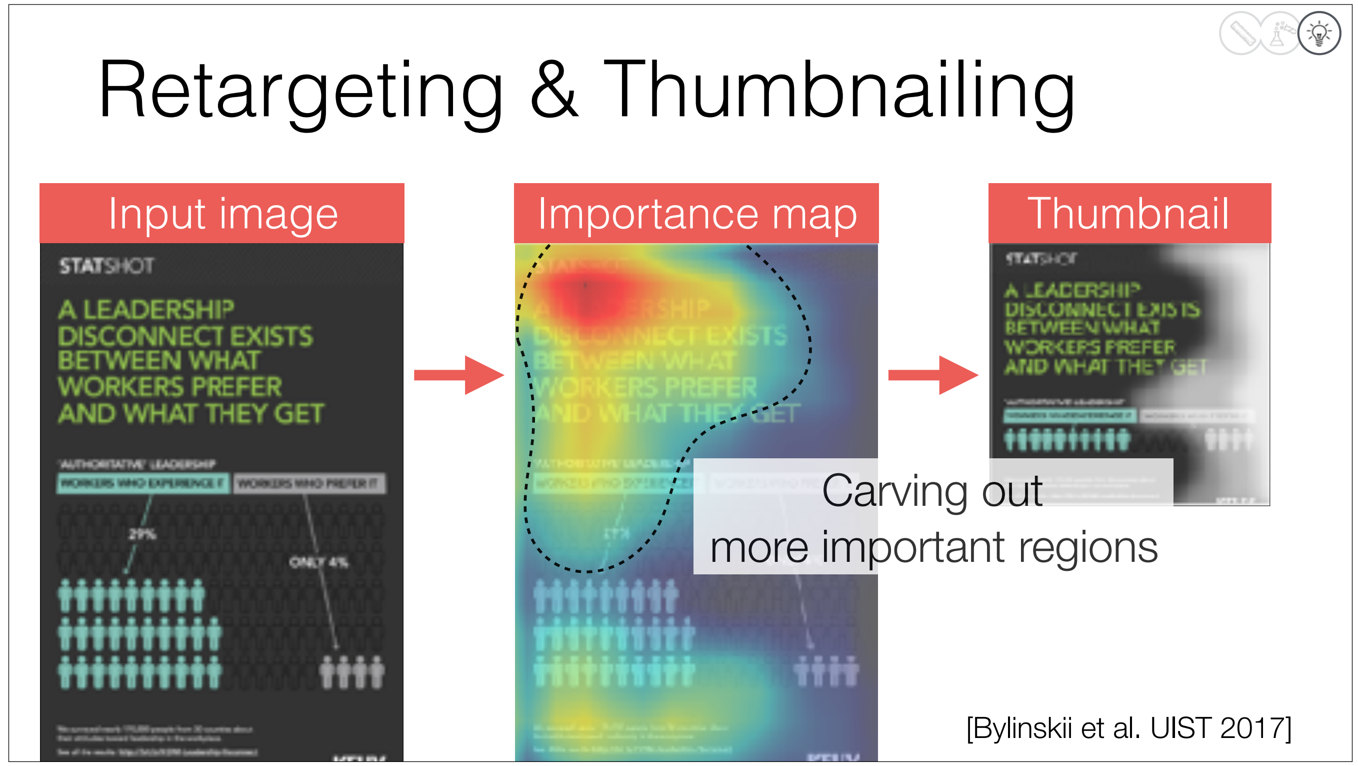
Input image



Importance map

[Bylinskii et al. UIST 2017]

One application is image retargeting and thumbnailing. For example, given an infographic image, we can collect attentional data using BubbleView and create an importance map.



By carving out more important regions of the image, we can generate a thumbnail of the image.

Prediction of Visual Importance

Providing real time feedback based on importance predictions

[visimportance.csail.mit.edu]

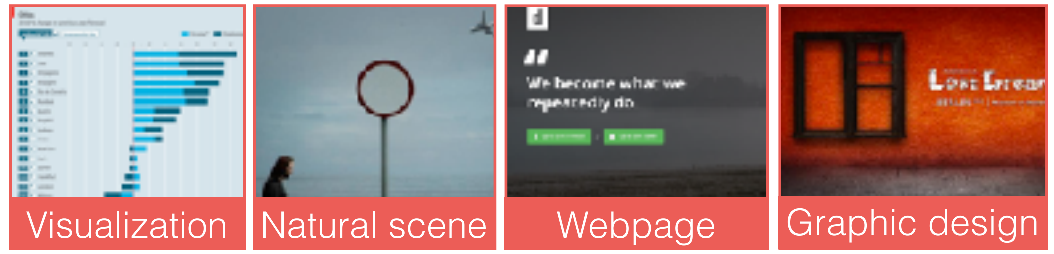
Using BubbleView, you can collect large scale human attentional data and build computations models to predict where people would look. This is a proof-of-concept design tool providing real time feedback based on importance predictions.

Task-specific attentional data

Which city is ranked first (find an extremum)?

You can use BubbleView to collect task-specific attentional data by asking more guided questions such as finding an extremum instead of asking them to freely explore or describe the image.

Other image domains & tasks

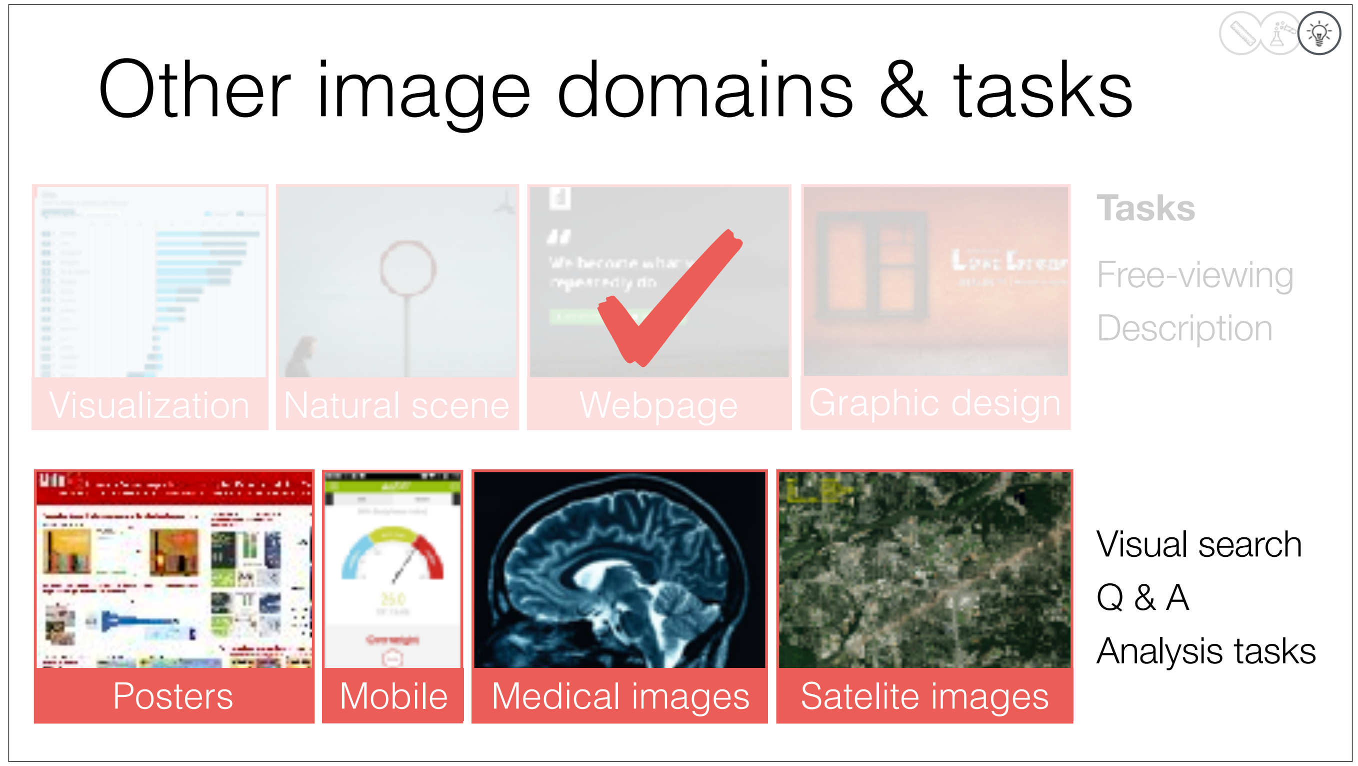


Tasks

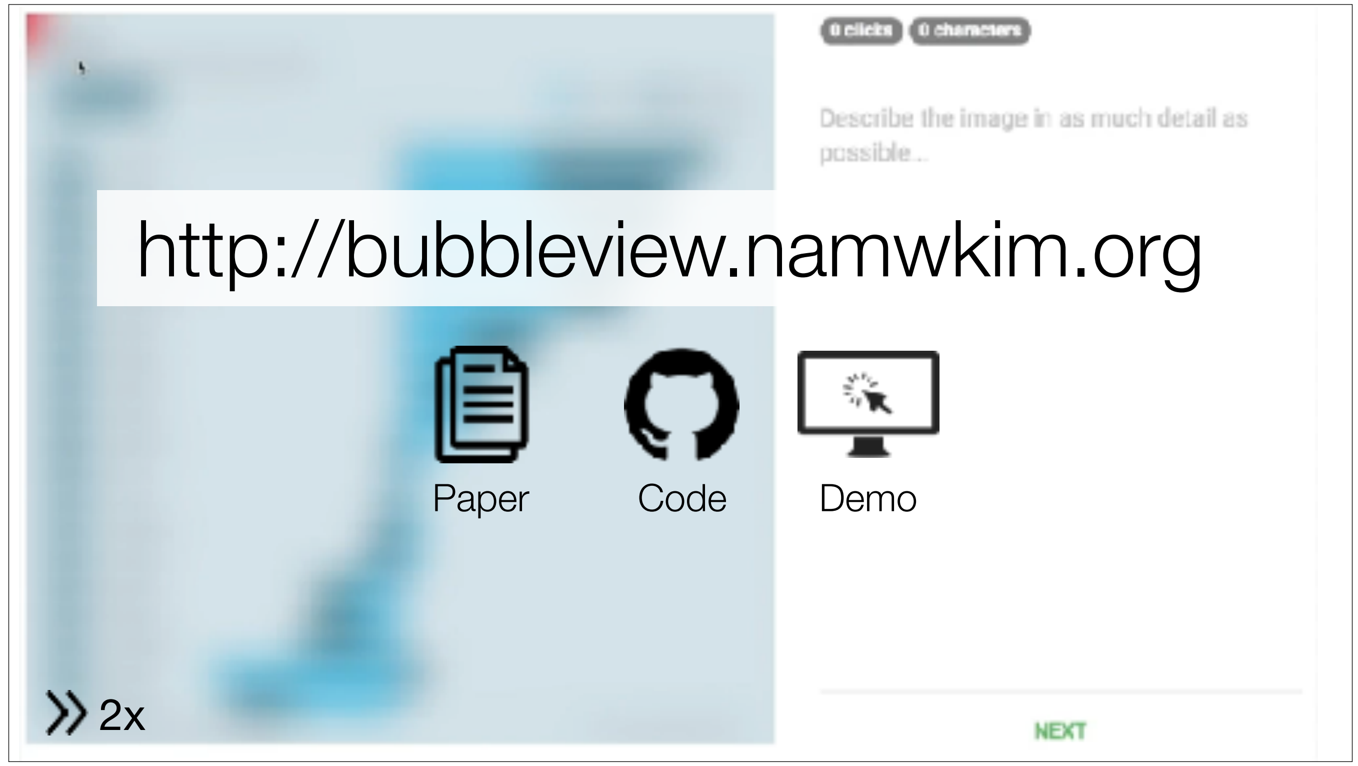
Free-viewing

Description

In this work, we showed that BubbleView generalizes to different types of images, as well as free-viewing and description tasks.



This can be expanded to new image types, for instance for studying medical images, geographical maps, user interfaces, slides and posters.



For future explorations, you can find more information about BubbleView on this website. Thank you


THE RISE
OF
THE BRICKS