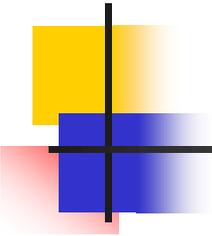


Did you see that thing?

An eye tracking study on the reliability of self-reported awareness measures

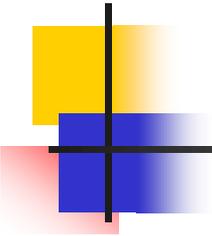
Bill Albert, Design and Usability Center, Bentley University

Donna Tedesco, Fidelity Investments



Research Questions

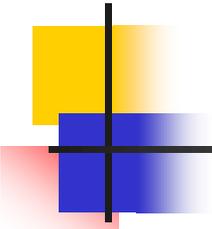
- Usability professionals often asked participants if they noticed a particular object
 - Is this feedback reliable?
 - If so, are there certain situations in which it is more/less reliable?
- Implications for how we ask participants questions, and the need for eye tracking technology as part of basic usability testing



Background Research

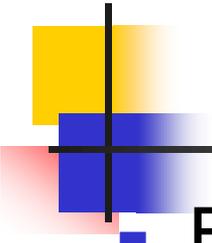
- Guan et al (2006)
 - RTA was validated with eye movement data
 - Omissions occurred 47% of the time

- Johansen & Hansen (2006)
 - Participants recollection of elements were valid about 70% of the time
 - Images, text, and navigation was recollected more than 70% of the time, logos only 30% of the time



Methods

- Half of the participants (n=40) were calibrated with eye tracker, the other half (n=40) were not
 - Is there an impact of the technology on what people report?
- Shown a series of popular website home pages for 7 seconds (study page), followed by a test page
- Test page includes two objects highlighted
 - Objects are images, navigation, or functional based



Experiments

■ Experiment 1

- 40 participants asked “Did you notice this object”?
- 3-point scale (1 = definitely did not notice; 2 = not sure; 3=definitely noticed)
- Half were eye tracked (n=20); half not eye tracked (n=20)

■ Experiment 2

- 40 participants asked “How much time did you spend looking at this object?”
- 5-point scale (1 = no time looking at object ... 5 = a long time looking at object)
- Half were eye tracked (n=20); half not eye tracked (n=20)

Eye Tracking Condition	Expt. 1 (“Did you notice”)	Expt. 2 (“How much time”)
Yes (ET)	20 participants	20 participants
No (NET)	20 participants	20 participants

Example of Study/Test Page

Welcome. [Local weather in 1-click](#) | [Put weather on my desktop](#) | [Customize weather.com](#) | [Sign In](#)

The Weather Channel weather.com

Localweather Enter zip or US/intl city GO

Maps | Video | News | TV | Mobile | Alerts

Home In Season Plan Ahead My Neighborhood Travel Smart Stay Healthy Around the Home Get Out & Play

Hurricane Season
It's hurricane season: [Are you ready?](#)
[Make a preparedness plan](#)
Video: [Know how storms get their names?](#)
Get your forecast:
Enter city or US zip GO

Weather Center August 14, 2007

Hawaii Satellite

Hurricane Central
Flossie nears Hawaii; TS Dean develops
8/14/07 2:50 p.m. ET

Category 2 Hurricane Flossie is nearing the Big Island of Hawaii. Tropical Storm Dean is christened in the Atlantic.

- Tracker | [Interactive map](#)
- Video: [Flossie, Dean updates](#)
- [TS Dean named in Atlantic](#)

[Nat'l Forecast: Storms for Midwest; hot South](#)

Map Room [Click to Enlarge](#)
Select a map

GET TIPS FOR DRIVING IN ANY WEATHER SITUATION

VISIT THE AUTO CENTER

LIKE A GOOD NEIGHBOR STATE FARM IS THERE!

Get Your Weather Your Way!

Desktop

- Weather toolbar [download here free](#)
- weather.com on [your mobile phone](#)
- Get faster maps & [no ads](#)
- Free weather for [your site or blog](#)

Get instant access to The Weather Channel forecasts right on your desktop [Get it now](#)

Memory Test

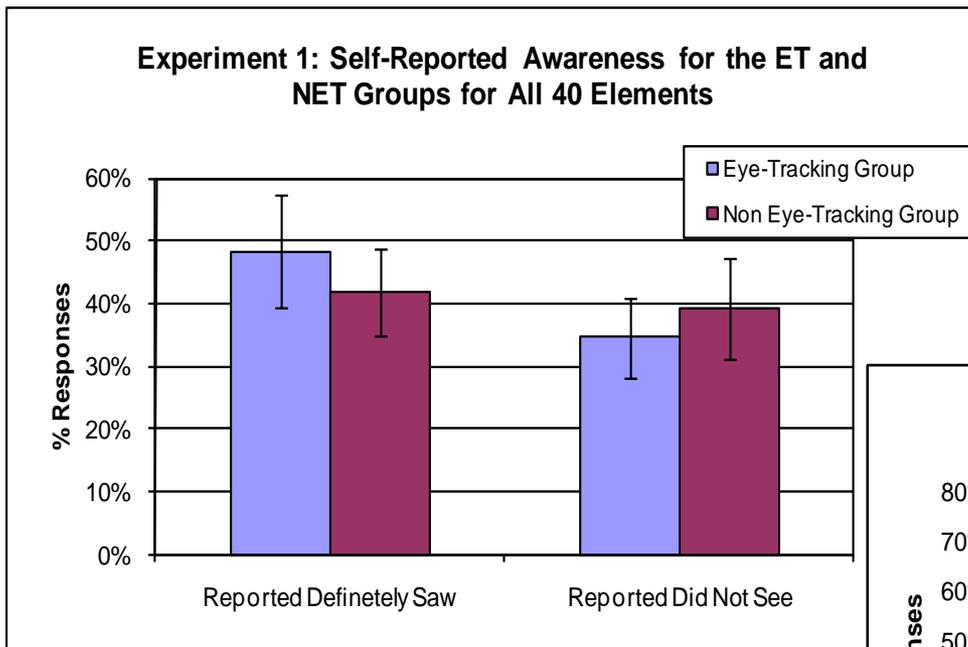
The screenshot shows the eBay homepage with several red boxes highlighting specific elements:

- Top Navigation:** A box highlights the navigation links: Buy Sell My eBay Community Help.
- Search Bar:** A box highlights the search input field, the "All Categories" dropdown menu, and the "Search" button.
- Category Navigation:** A box highlights the "Categories" dropdown menu, the "Motors" link, the "Express" link, and the "Stores" link.
- Specialty Sites:** A box highlights the "SOLD OUT TICKETS" banner.

The page content includes:

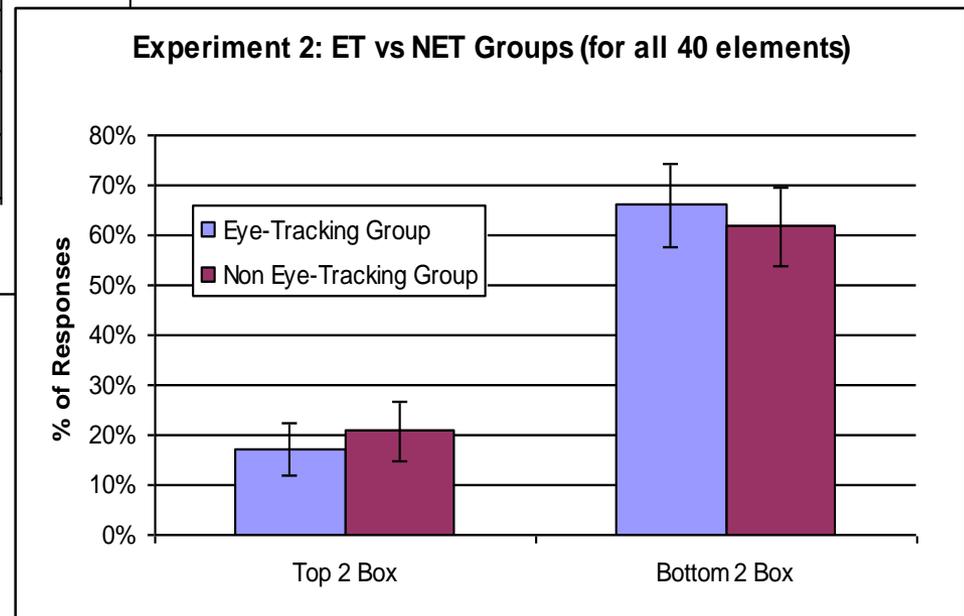
- Header:** eBay logo, "Hello! Sign in or register.", and "Site Map".
- Search:** Search bar with "All Categories" dropdown and "Search" button.
- Navigation:** "Categories" dropdown, "Motors", "Express", and "Stores" links.
- Logos:** "Java™ TECHNOLOGY" and "POWERED BY Sun".
- Live help:** "Live help" link.
- Main Banner:** "Whatever it is...you can get it on eBay".
- My Favorite Searches:** "My Favorite Searches: Sign in to view your Favorite Searches."
- Specialty Sites:** Links for eBay Express, eBay Motors, eBay Stores, eBay Business, Half.com, Apartments on Rent.com, and StubHub Tickets.
- Categories:** A list of categories including Antiques, Art, Baby, Books, Business & Industrial, Cameras & Photo, Cars, Boats, Vehicles & Parts, Cell Phones & PDAs, Clothing, Shoes & Accessories, Coins & Paper Money, Collectibles, and Computers & Networking.
- Advertisements:** "What's your BAG?" (Class, Adventure, Glam, Wild, Eco Friendly), "SOLD OUT TICKETS", "My eBay at a Glance" (Sign in for a snapshot of your personalized information on the homepage), "Not Just for Kids Dolls & Bears", "Get Ready to Rock Musical Instruments", and "Collect them all Coins".

Impact of Eye-Tracking

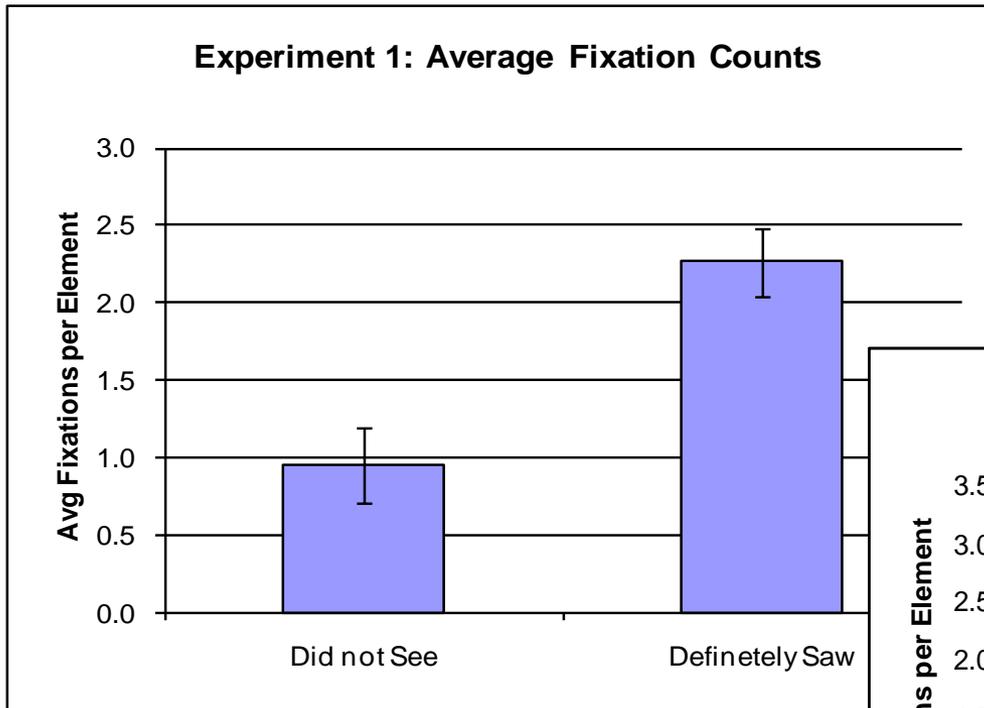


There was no impact of using eye tracking technology on self-reported measures

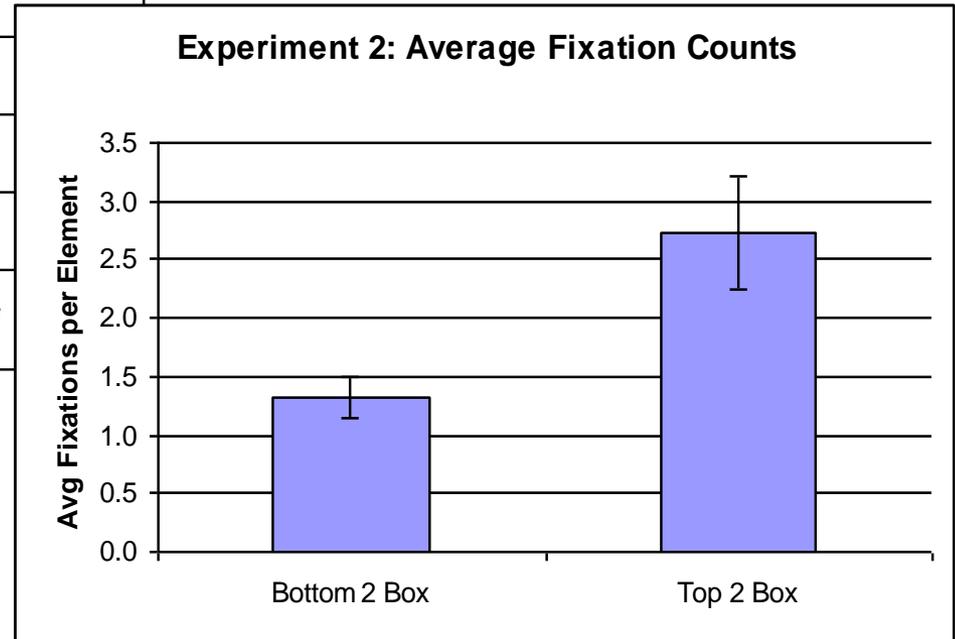
No statistical difference between the ET (n=40) and NET groups (n=40)



Fixation Counts

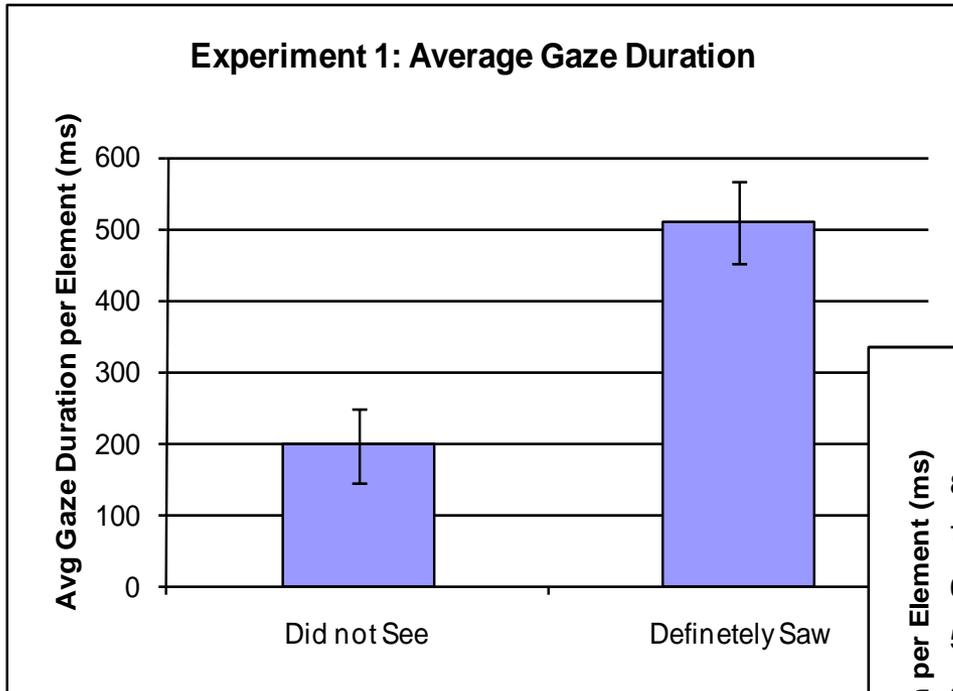


There were significantly more fixations for those that reported "definitely saw" top-2 box

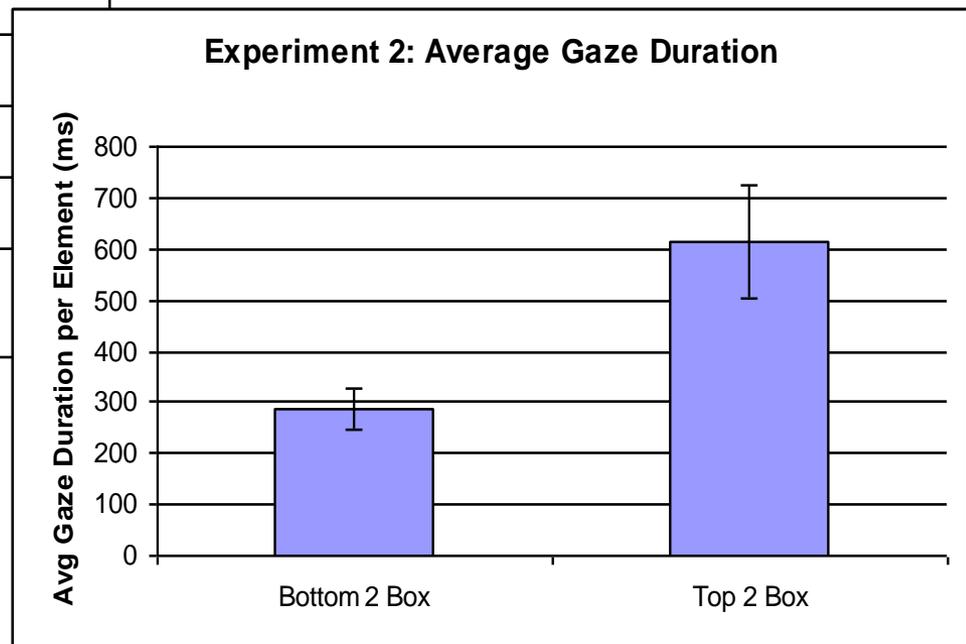


Results indicate that what participants report is initially supported by eye movement data

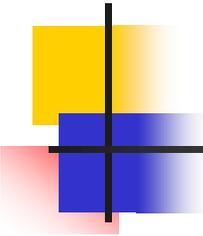
Gaze Duration



Significantly longer time spent looking for those that reported “definetely saw” or were top-2 box



Consistent with fixation counts, self-reported awareness has some basis in the eye movement data



Response Outcomes

Responses	Errors	Success
Definitely saw (Expt. 1) or top 2 box (Expt. 2)	False alarm (gaze duration = 0ms)	Hit (gaze duration > 250 ms)
Definitely did not see (Expt. 1) or bottom 2 box (Expt. 2)	Miss (gaze duration > 500 ms)	Correct rejection (gaze duration < 250 ms)

Experiment 1

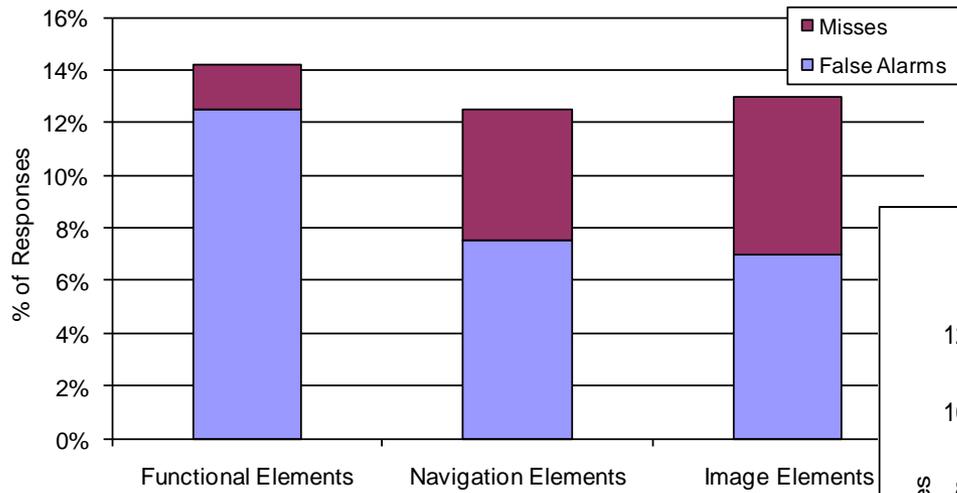
Responses	Errors	Success
Definitely saw	10.2% (false alarm)	28.2% (hit)
Definitely did not see	4.8% (miss)	27.0% (correct rejection)

Experiment 2

Responses	Errors	Success
Top 2 box	4.8% (false alarm)	11.7% (hit)
Bottom 2 box	12.6% (miss)	22.1% (correct rejection)

Object Types

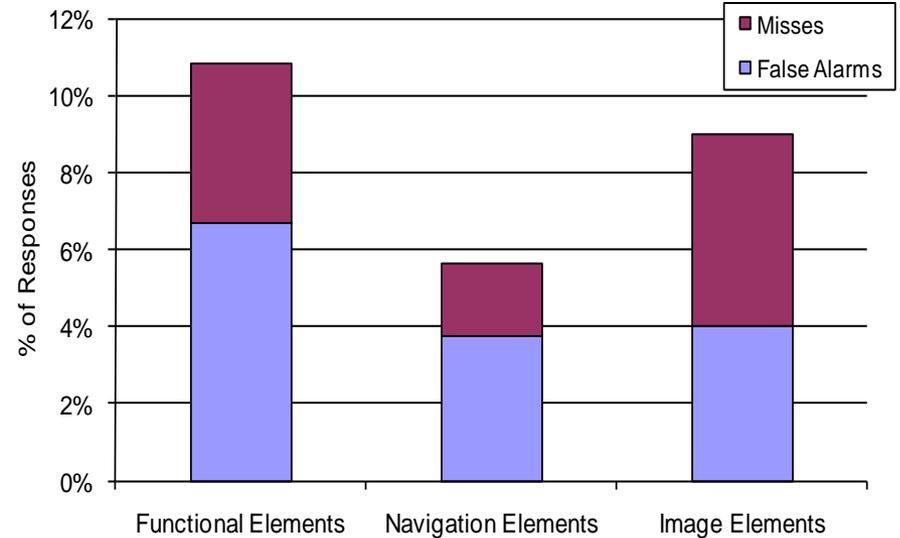
Experiment 1: Error Rates by Element Type

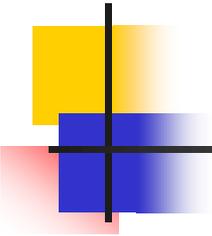


No statistical differences between the three types of elements (Expt 1)

Functional-based elements have higher false alarm rates than other element types

Experiment 2: Error Rates by Element Type



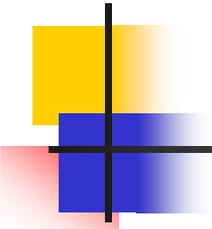


Results of Memory Test

Surprising how many participant had a false recollection

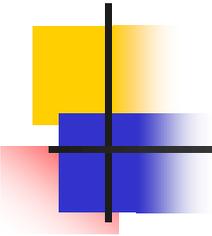
But, a more continuous question gives participants more leeway in how they respond

Response	Percent
Definitely saw (Experiment 1)	26.8%
Top 2 box (Experiment 2)	8.9%



False Alarms Make Sense

- A *false alarm* scenario
 - The design team wants to test if a particular object is noticed
 - During a usability evaluation they ask participants whether or not they noticed a particular object
 - Some participants may say they noticed the object, but did not
 - The design team incorrectly concludes that the object is visually prominent enough, and no steps are required to increase its visual prominence
- False alarm scenarios happen and should be avoided

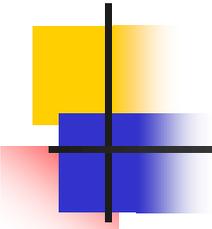


Misses Don't Make Sense

- *Amis* scenario

- The design team wants to make sure an object is NOT noticed
- They run a usability evaluation, and ask the participants if they noticed a particular object
- Some of the participants report not seeing the object, whereas they actually did notice it
- The design team incorrectly concludes that the object is well hidden, and they don't need to make it less prominent

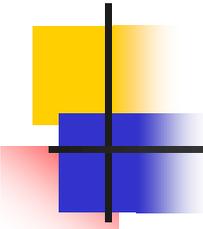
- How common is this?



False Alarms Give Us Hope

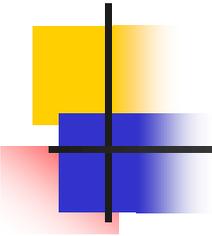
- In Experiment 1, there was a false alarm rate of about 10%
- In Experiment 2, there was a false alarm rate of 5%
- Navigation and image-based elements had a lower false alarm rate in both experiments

- Is this an acceptable error rate?



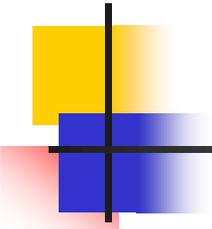
Other Side of Coin

- Experiment 1 confirmed 55% of responses
 - 28% hit rate (said they saw when they really did)
 - 27% correct rejection (said they did not see, when they did not look)
- Experiment 2 confirmed only 34% of responses
 - 12% hit rate
 - 22% correct rejections
- If you want to be sure they saw something, ask a more discrete question about awareness



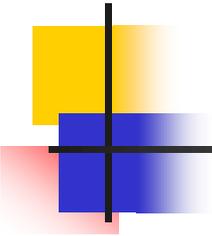
Study Limitations

- Mind's-eye hypothesis
- No tasks were given, only orientation to the home pages
- Did not control for level of familiarity



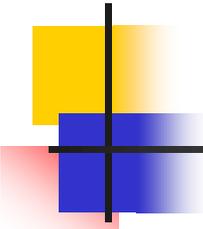
Study Conclusions

- Usability practitioners should feel confident in collecting self-reported awareness measures from participants. They will not draw an incorrect conclusion more than 10% of the time.
- If a practitioner wants to minimize the chance of making an incorrect conclusion, they should use a continuous (5- or 7-point) scale for self-reported awareness
- If a practitioner wants to maximize the likelihood of confirming that a participant did or did not see an element, they should use a discrete set of questions for self reported awareness
- Participants are more reliable in their self-reported awareness for navigation and image elements, than functional elements, regardless of question structure.



Take Home Message

- Think about how you ask an awareness question
- Be careful how you interpret their response
- Eye tracking still **VERY** useful as part of UX research - it all depends on the question you are asking!



Thank You!

Full article:

Reliability of Self-Reported Awareness Measures Based on Eye Tracking, *Journal of Usability Studies*, 5(2), 50-64

http://www.upassoc.org/upa_publications/jus/

Questions or praise

Bill Albert (walbert@bentley.edu)

Donna Tedesco (Donna.Tedesco@fmr.com)