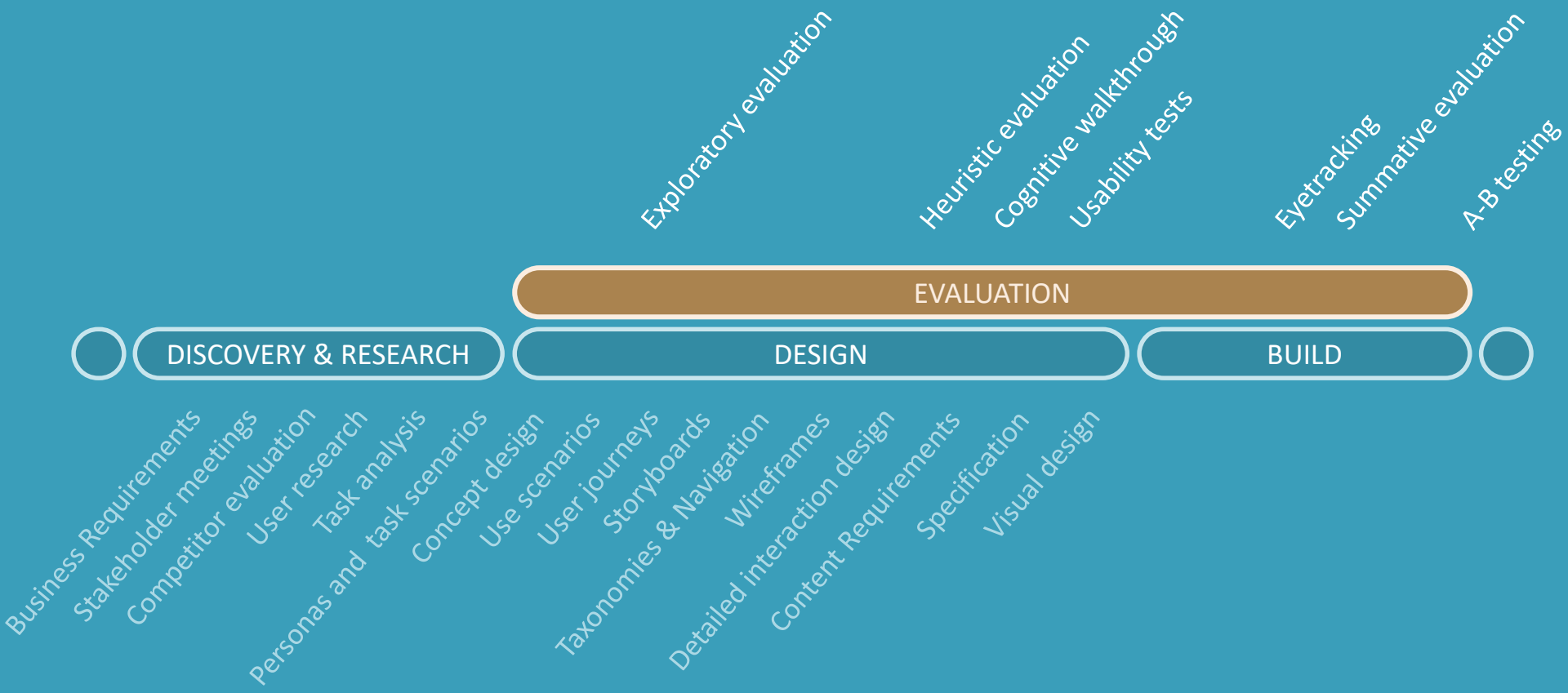
Evaluation

Discuss: *Why evaluate?*


Exploratory evaluation
Heuristic evaluation
Cognitive walkthrough
Usability tests
Eyetracking
Summative evaluation
A-B testing
EVALUATION
DISCOVERY & RESEARCH
DESIGN
BUILD
Business Requirements
Stakeholder meetings
Competitor evaluation
User research
Task analysis
Personas and task scenarios
Concept design
Use scenarios
User journeys
Storyboards
Taxonomies & Navigation
Wireframes
Detailed interaction design
Content Requirements
Specification
Visual design

Evaluating with our without users?

Interviewing and observing users

- Usability testing **with real users is the most fundamental and useful usability method**
- Methods range from **informal testing to precisely controlled studies**
- Testing may be **quantitative** or **qualitative**

Using analytical methods

- You can also conduct a **heuristic inspection** against standard guidelines
- A **cognitive walkthrough** lets you check a user's journey
- Sometimes this is enough for an interim review

Evaluating with or without users?

- **Measuring user behaviour on a finished site**
  - **A-B testing** allows you to show two different versions of a website to users and see which one has better click-through rates
  - Using **eyetracking** techniques can provide an interesting take on what is seen on your site

Evaluating what and when?

- **Start early and low-fidelity** (paper prototypes, basic wireframes, rough design concepts)

- **Once you've got a more complete design, build a prototype** (for example HTML, Powerpoint, Axure, Invision, Proto.io, Balsamic) and test it.

Exploratory/formative evaluation

Early in the process. Test **low fidelity** prototypes to inform conceptual design and ground user pref'nces

Qual

Comparison evaluation

At any stage. Compare designs for suitability or measure task completion rates

Qual & Quant

Assessment evaluation

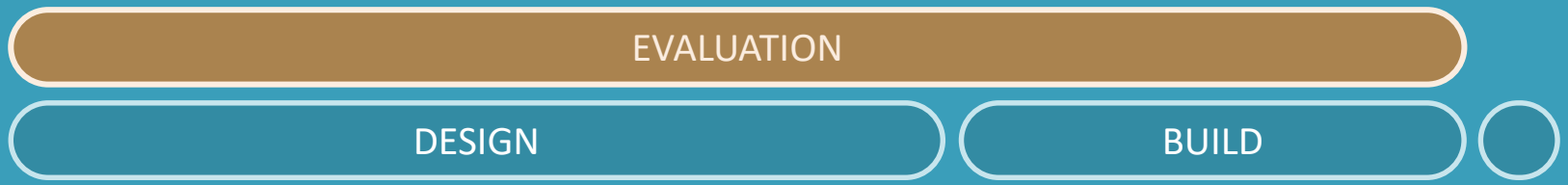
Midway to late, To evaluate for usability problems once a design has been created.

Qual & Quant

Summative/validation evaluation

To verify the system is complete, working and meets usability metrics.

Mostly Quant



The evaluation strategy

- **Why** are we evaluating? – **evaluation goals**
- Which usability **requirements** are we exploring?
- **What** are we evaluating?
- What **type of data** do we want to collect?
- What **constraints** do we have?

Without answering these questions you can't plan your research!

What to evaluate

- Evaluation goals can be identified from different sources:

- From the earlier **user research**
  - Can users understand the proposition?
  - Can users complete goals that are important to them?
- From the **client objectives**
  - Do users comprehend the client's objectives for the website?
- From the **design process**
  - To explore alternative conceptual models, IAs, visual designs, etc
  - To get user data to inform important design decisions

Some common usability evaluation goals

- **Comprehension**
  - do users understand the page/ site and their options?
- **Satisfaction**
  - Does the content and functionality of the page/ site meet user expectations?
- **Anticipation**
  - Is it clear what will happen next in the user journey?
- **Task success rate**
  - How often can users complete common tasks like purchases?

Task: *Think of your own project and formulate an evaluation strategy…*

Task: *Formulate an evaluation strategy…*

- **Why** are we evaluating?
- Which usability **requirements** are we exploring?
- **What** are we evaluating?
- What **type of data** do we want to collect?
- What **constraints** do we have?

Heuristic evaluation

What is a heuristic evaluation?

- Heuristics are rules-of-thumb or guidelines about what makes a website usable
- The most common set of heuristics were developed by Jakob Nielsen in the early 90s
- Many different sets of heuristics exist for dedicated design domains

- They are a low-cost way of making sure that your site will avoid common usability problems by involving other people in critiquing your design

Nielsen's heuristics

- **Visibility of system status**
  - Always keep users informed about what is going on
- **Match between system and the real world**
  - Follow real-world conventions, using natural language & making information appear in a natural and logical order.
- **User control and freedom**
  - Don't box the user in. Support undo and redo.
- **Consistency and standards**
  - Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.
- **Error prevention**
  - Prevent problems from occurring
- **Recognition rather than recall**
  - The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.

Full article on Nielsen's heuristics

Nielsen's heuristics

- **Flexibility and efficiency of use**
  - Support shortcuts for experienced users
- **Aesthetic and minimalist design**
  - Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.
- **Help users recognize, diagnose, and recover from errors**
  - Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.
- **Help and documentation**
  - Even though it is better to design a system that doesn't need documentation, if you have it make it focussed on the users' task and useful

Full article on Nielsen's heuristics

How to do a heuristic evaluation

- Get between 1 and 5 designers or researchers and brief them on the website. Give them a copy of the heuristics you're using
- Working independently, they review the relevant screens of the site against the heuristics
- In a workshop afterwards, they compare issues they found to agree a final set of prioritised usability issues

- You can also use the heuristics while you are designing as a sanity check, but others will always bring a more objective eye

Strengths & weaknesses

- Heuristics are quick to apply and have stood the test of time as principles for design
- But they are only a rule of thumb, they are not infallible
- Heuristics are only as good as the critical thinking of the person who is applying them
- No set of heuristics can address the all the criteria of your website
- But they are a great way to help develop a sense of what makes up a usable website

Task: *Conduct a heuristic evaluation…*

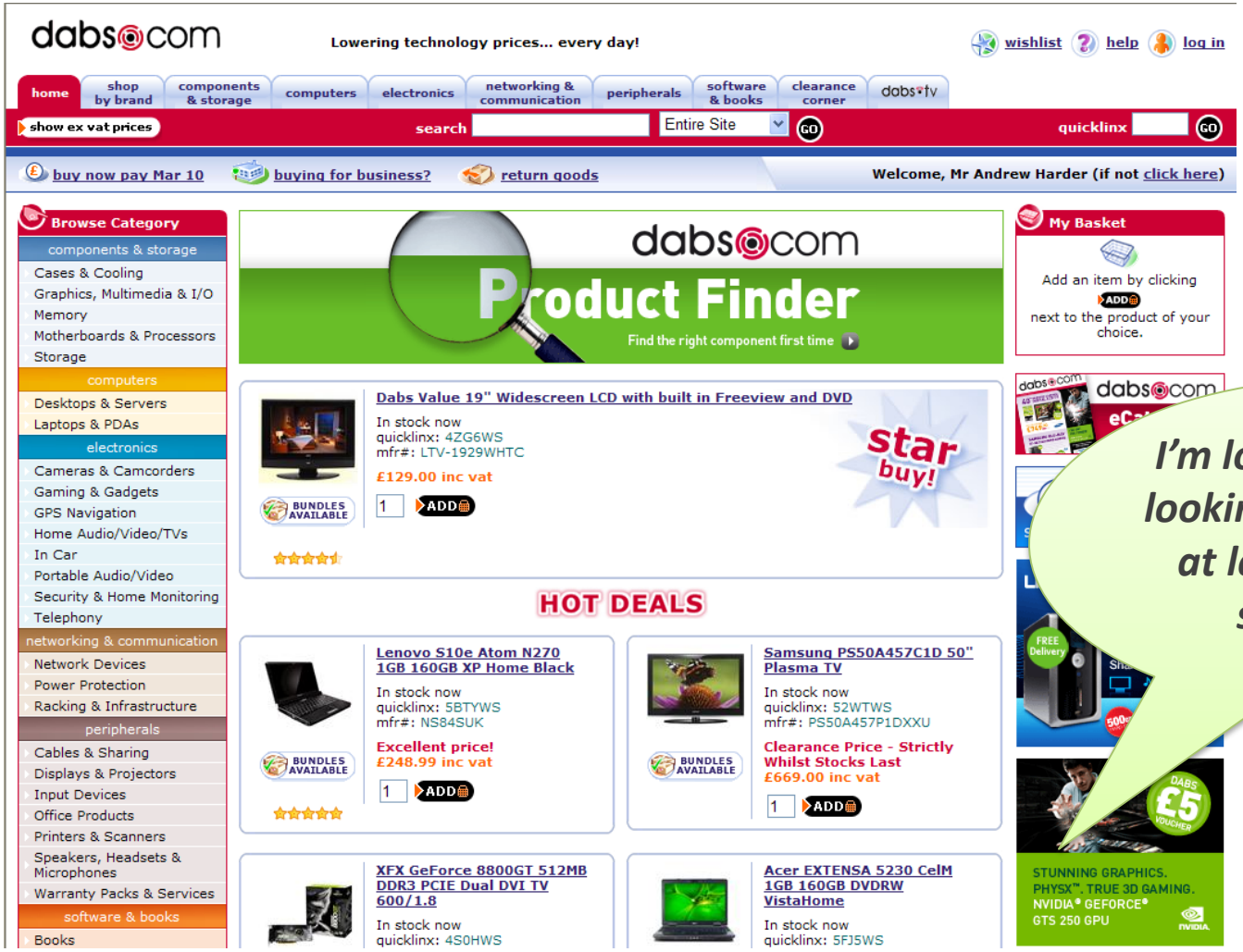
- Conduct a heuristic evaluation of the TfL Journey Planner: https://tfl.gov.uk/plan-a-journey
- Use Nielsen's heuristic evaluation and make notes in these headings
  - **Website feature:** What part of the site is of interest?
  - **User implication:** What is the likely impact on users? Confusion, etc
  - **Priority:** How important is the issue
  - **Recommendation:** What should be done about the problem?
- Remember to include good as well as bad points

Cognitive Walkthrough

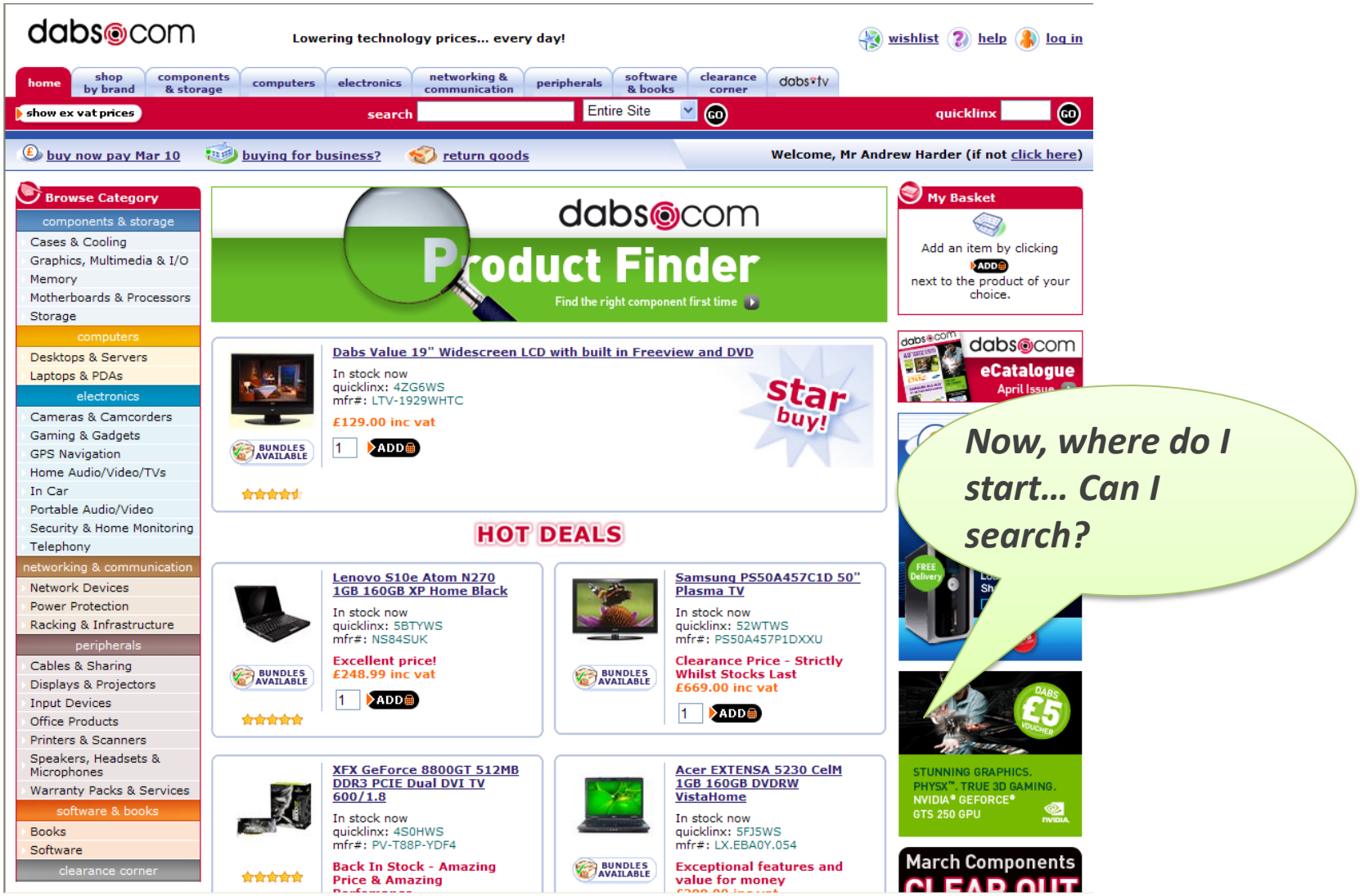
What is a cognitive walkthrough?

- A cognitive walkthrough is a complementary analytical technique that focuses on how easy a site is to learn
- This is a good exercise to ensure your site has a good information scent and that core tasks (like purchase) are well-supported
- Cognitive walkthroughs focus on three key questions

Will the user understand what their next action has to be to complete their task?

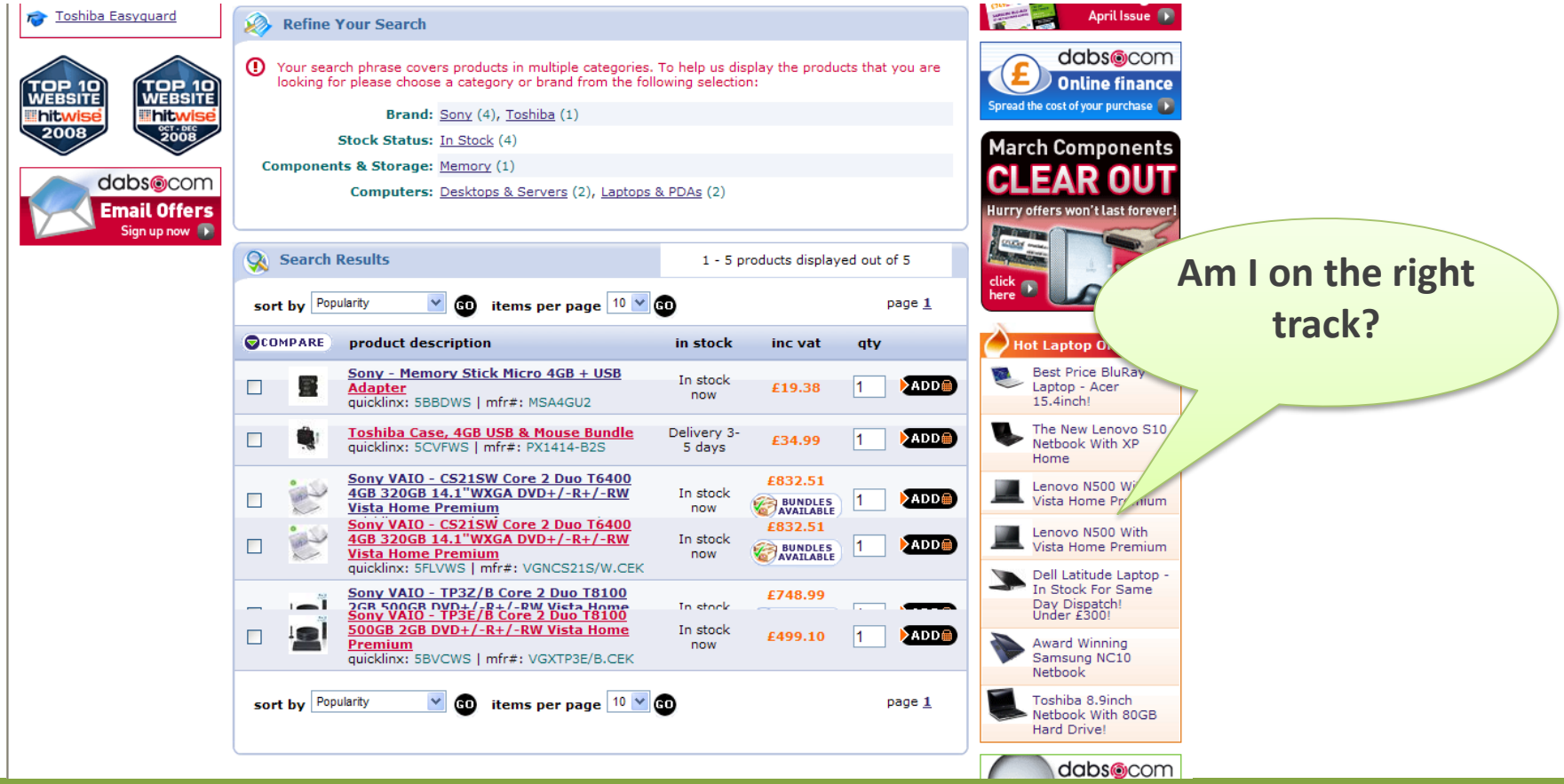


Will the user see how to take their next action?



Will the user understand that they have successfully completed, or failed to complete, the right action?

- Now, what does this mean?



Applying cognitive walkthroughs

- As another analytical method, you use the same approach as heuristics:
- Get a group of designers or researchers together with a prototype of your site and let them loose
- You then get their feedback and agree a list of common usability problems they've identified

Strengths & weaknesses

- Like the heuristic evaluation, a cognitive walkthrough is **relatively quick and cheap**
- It complements heuristic evaluation by **focussing on the users journey** through the website to complete key tasks
- This means that key features like buying will be surfaced
- Because it focuses on learnability, it **won't help you cater to the needs of expert users,** or help identify broader non-transactional objectives like building brand identity

Task: *Conduct a cognitive walkthrough…*

- Do you own cognitive walkthrough
- Use the three key questions and make notes on the same headings
  - **Website feature:** What part of the site is of interest?
  - **User implication:** What is the likely impact on users? Confusion, etc
  - **Priority:** How important is the issue
  - **Recommendation:** What should be done about the problem?

Usability tests

What do we mean by usability test?

- A usability test is a one-on-one session where a representative user interacts with a prototype design with a researcher facilitating and taking notes
- Information you can get out includes:
  - Whether the **proposition is clear and of real value** to the user
  - Whether there are **usability issues that inhibit successful** use of the website
  - Whether key tasks can be quickly and readily completed by users
- A usability test can be **qualitative**, through open-ended interview questions, or **quantitative** by measuring task success and time rate

Quantitative testing

- A quantitative test can help inform the final stages of design by **verifying the performance of the site** along the key measures
  - **Task completion rate and time**
  - **Satisfaction rating**
- The researcher sets a few **standards tasks, times the participant and writes a survey to capture results** at the end
- When it identifies small problems, quant testing can identify where tweaks need to be made to the design
- But when it identifies bigger problems, quant testing typically doesn't provide enough guidance for a redesign

Qualitative testing

- Run **more like an open-ended interview,** qualitative testing lets the participants explore the prototype, exploring their own motivations & interests as well as your test objectives
- Qual testing is quite strong at **discovering a more holistic understanding of your user's mental model, perceptions** and **motivations**
- But you need to **be careful with interpreting** what your users say
- And it is often **hard to get an overall picture** of how well the current design performs

How to do a usability test

- Mix quantitative and qualitative objectives to suit your evaluation objectives
  - Regardless of what you want to learn and where you are in the design process, it is almost always useful to include both qual and quant methods

- Set up the usability test
  - Users
  - Client observations
  - Write an interview script/ discussion guide
  - Write a brief survey for your quantitative objectives

Users – who and how many?

- Should be as **representative** as possible
- For an intranet, it's easy to find appropriate users!
- For an Internet site, need to have a **sample of users with similar demographic distribution** to the intended user population
- Normally you'll get those from the **specialist market research recruiter** your usability consultants work with

- Jakob says you only need to **test with 5 users**, but not everyone agrees!
- A normal test is conducted with 5-10 users doing the same things

Observing sessions

It's important that stakeholders turn up to observe!

- It help them understand that **users may see their product with different eyes**
- It creates a **common reference point** and shared experience across the team
- Lots of observers **see more** than just one (and help with note taking)
- It's great fun! (or rather depressing sometimes!)

Warning: Avoid observers in the same room, it's intimidating!

Basic tips for interviewing users

- **Demographic questions** – what kind of person are you?
- **Task questions** – how would you achieve this? What are you doing? Why? Is that what you expected to happen?

  DON'T: ask questions that could get a yes/no answer
  DON'T: ask leading questions ("You like this, don't you?")
  DO: deviate from the script
  DO: tell them you didn't make the prototype. (Impartial**)**

- **Remember: comprehension, anticipation, satisfaction, task success**

Think aloud

- A useful technique for qualitative interviews is to ask users to “think aloud”
- Get immediate feedback on what users are **doing,** where they **looking,** what they are **reading,** what they are **understanding**
- Some people will find this hard to do, but keep encouraging them to keep talking.
- There are two magic phrases: “uh-huh?” and “why is that?”

- While they are thinking aloud, users will be **paying more attention** to what they are doing, reducing mistakes
- Unsuitable for measuring task time at same time

Retrospective protocol

- A different option to think-aloud is to ask users to proceed through their tasks first, then explain what their experience was like afterwards

- This means you can still time tasks accurately

- But users will still often **post-rationalise** what they did and explain away problems that they had

Measuring performance

In usability studies you can measure many things, though the 3 fundamental things to measure are:

System efficiency

…for example time to complete task, loading time

System effectiveness

…for example task completion, number of errors

User satisfaction

…for example preferences, uptake etc

Other aspects you can think of?

When to ignore what users say…

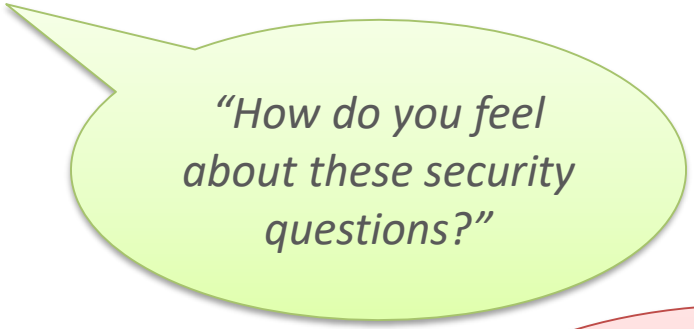
The most trustworthy of user feedback is their behavior during tasks

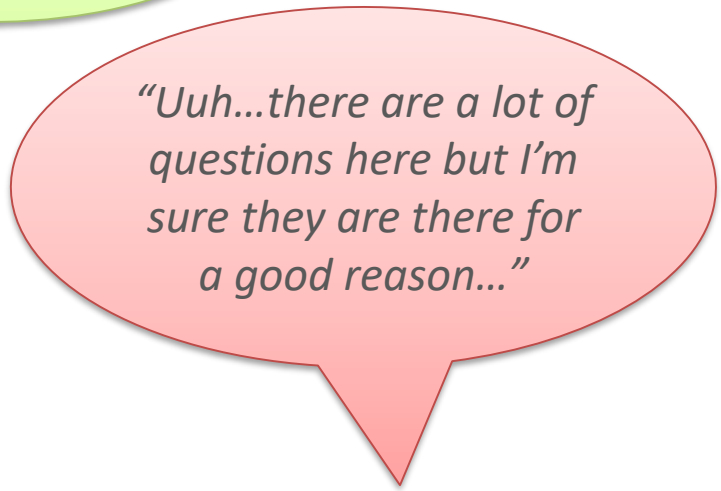
Beware when asking for their opinion

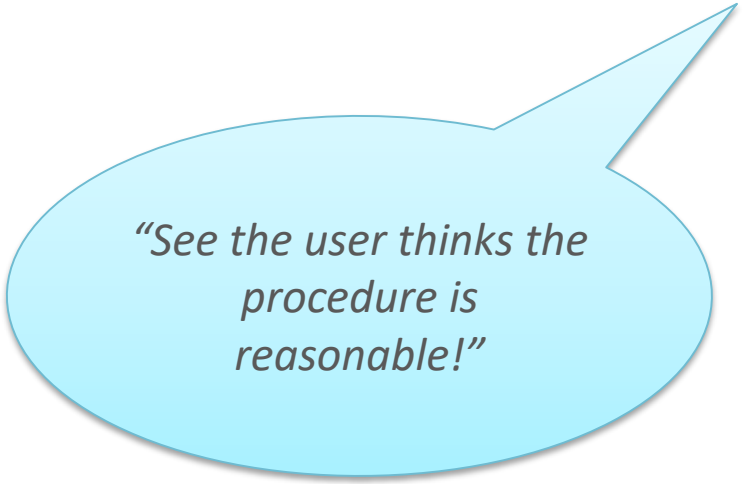
- Humans **aren't good at introspection**
- It's an awkward situation: **they will say whatever they think you expect** them to
- Often they'll **blame it on themselves**
- **They aren't designers:** they don't realise the impact of their suggestions

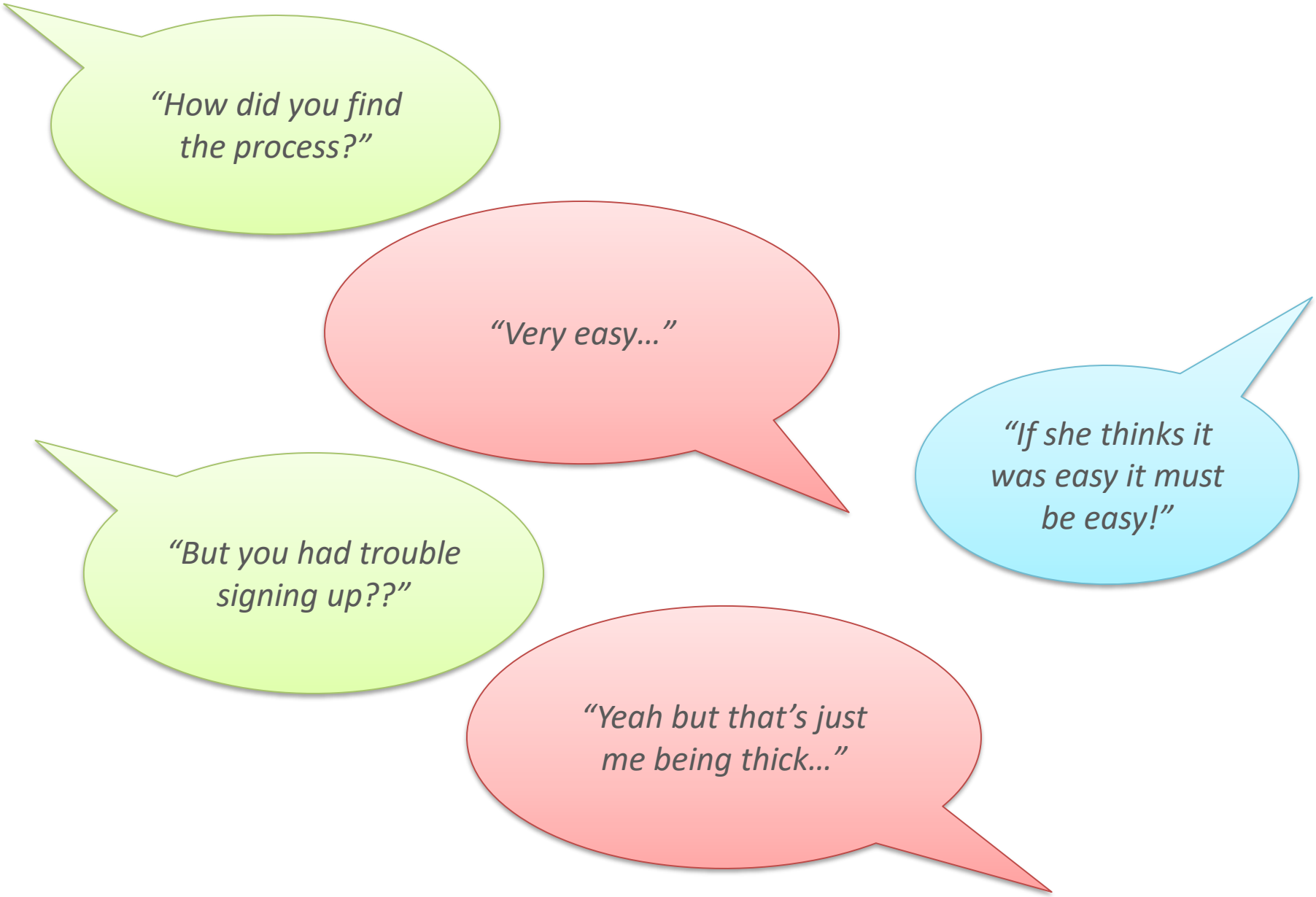
This means

- **Avoid what-if** questions. They are not effective.
- **Always interpret** what users say carefully.


"How do you feel about these security questions?"


"Uuh…there are a lot of questions here but I'm sure they are there for a good reason…"


"See the user thinks the procedure is reasonable!"


"How did you find the process?"
"Very easy…"
"If she thinks it was easy it must be easy!"
"But you had trouble signing up??"
"Yeah but that's just me being thick…"

Exercise

- Choose one of the phones in the session
- Set some evaluation goals
- Set your interview script to last for about 5 minutes
  - Introduction
  - Explore current user attitudes & behaviours
  - Complete some tasks (from user or yourself)
  - Wrap-up questions
- Get into pairs and participate in each other's usability test
- Report back!

Analysing and documenting findings

- **If an issue happens to only one of the six users, is it important?**
  …It's your call
- **What was the impact?**
  Catastrophic or just an annoyance?
- **How persistent was the problem?**
  Was it a one off or would it occur all the time?
- **Problem severity is a judgment call based on your expertise in user behaviour & usability theory**

- A **spreadsheet** or a **PowerPoint** showing findings and severity rating is the most common deliverable

Eyetracking & A-B testing

Eyetracking

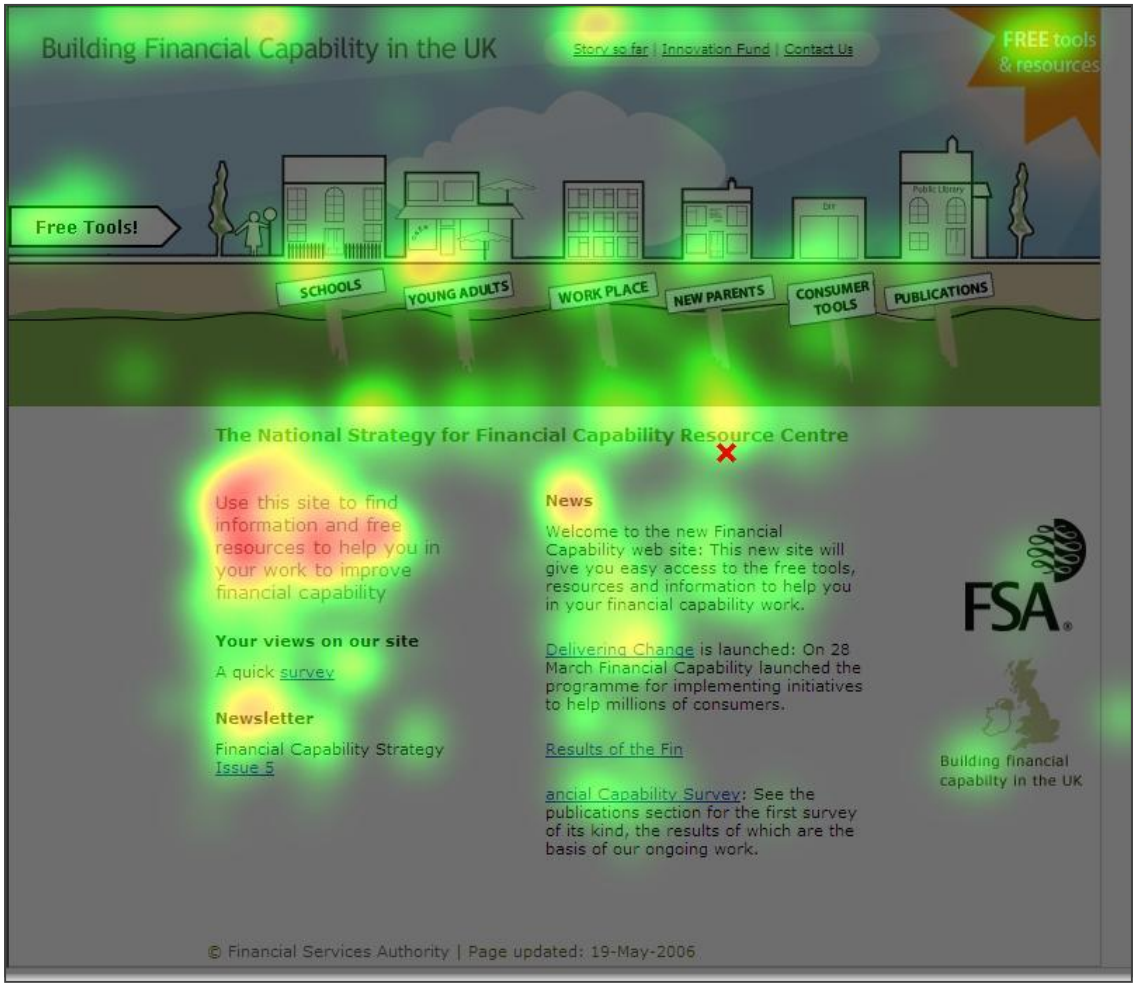
Eye-tracking records a reflection of invisible infra-red light off the cornea to **show where people are looking** on a screen.

The software records:

- The spots where people look (**fixations**)
- The eye movements (**scan path**)
- The **length of time** people look at a specific area of interest (fixation duration)
- The **mouse clicks**
- The **pages that they are looking** at

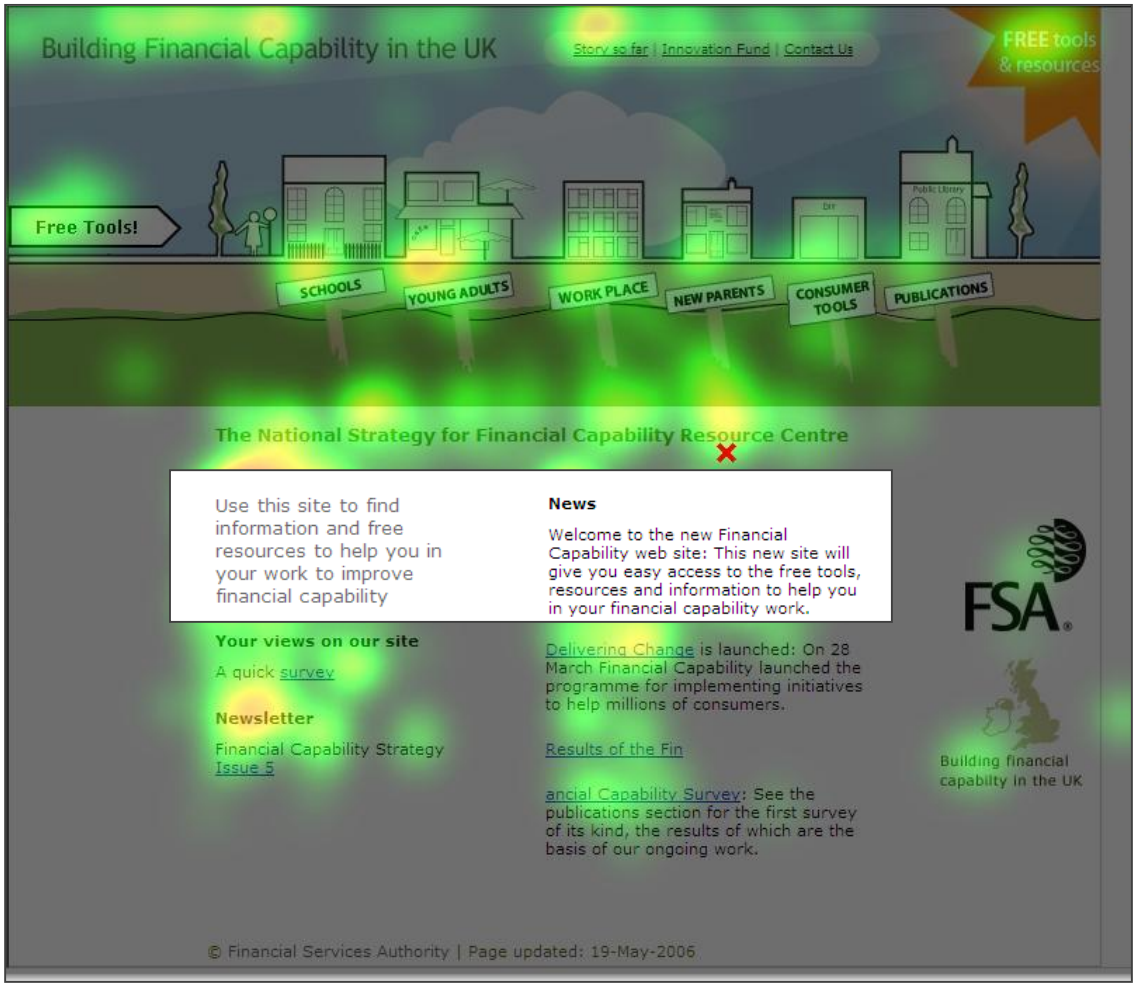
Example: *An eye tracking recording*

Example: *An eye tracking recording*



Discuss: *What does this tell you?*

Example: *An eye tracking recording*



Eyetracking

Downsides:

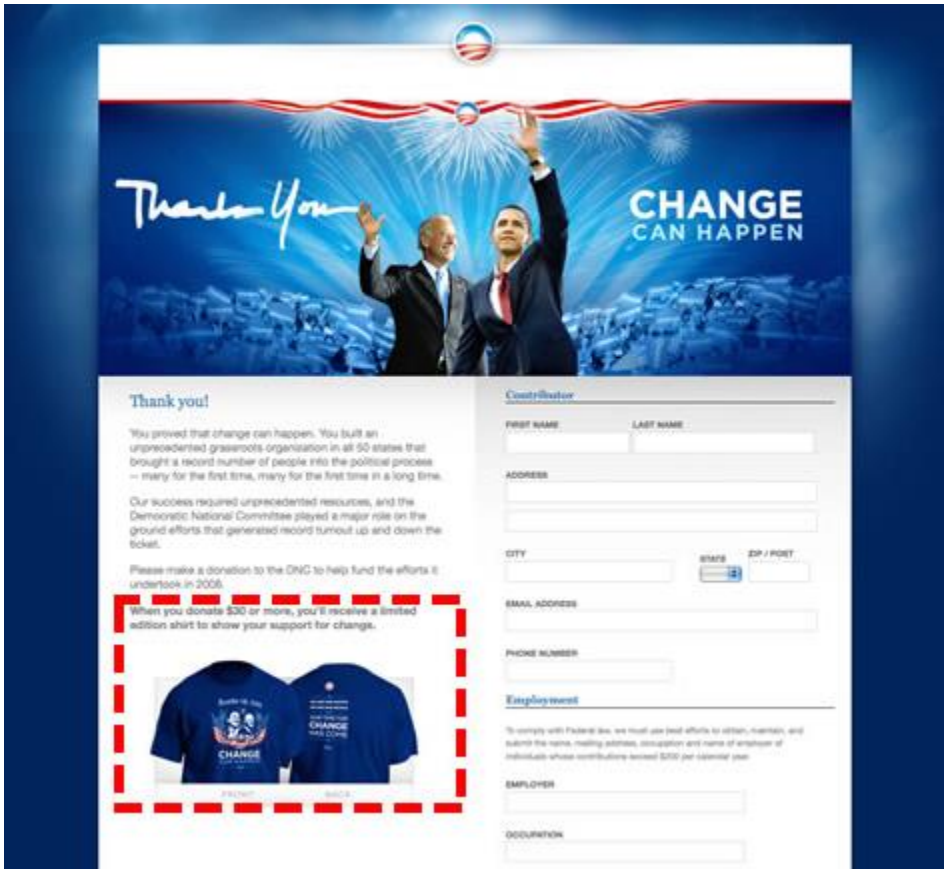
- Eye-tracking requires **expensive equipment, trained operators** and involves **additional set up** (at least ½ a day extra) and analysis time (at least one day extra)
- It only tells you **where users look,** but **not what the see,** why they are looking there and how they feel

Eyetracking needs to be **done in combination with in-depth interviews** and proper user testing.

What is A-B testing?

- A-B testing is where two versions of a website are created that differ by one controlled variable
- These sites are then shown to alternative users, and the results are analysed to see which version has better performance
- This is supported through online analytic tools like **Google Analytics Experiments**

Case study: *BarackObama.com*



When you donate $30 or more, you'll receive a limited edition shirt to show your support for change.

Control: dark blue t-shirt

When you donate $30 or more, you'll receive a limited edition shirt to show your support for change.

Variation A: single t-shirt

When you donate $30 or more, you'll receive a limited edition shirt to show your support for change.

Variation B: modified lighter blue

When you donate $30 or more, you'll receive a limited edition shirt to show your support for change.

Variation C: dual white t-shirts

Please make a donation to the DNC to help fund the efforts it undertook in 2008.

Variation D: no incentive

From [Harry Brignull's blog]

…beyond launch

The beauty of the web is that it's easy to adjust things after launch

Use web metrics to **measure uptake** post launch

Things you can find out:

- What users click on
- Their way through the site
- Where they come from
- Drop-out rates
- Entry and exit points

- More information here: **14 free tools that reveal why people abandon your website**

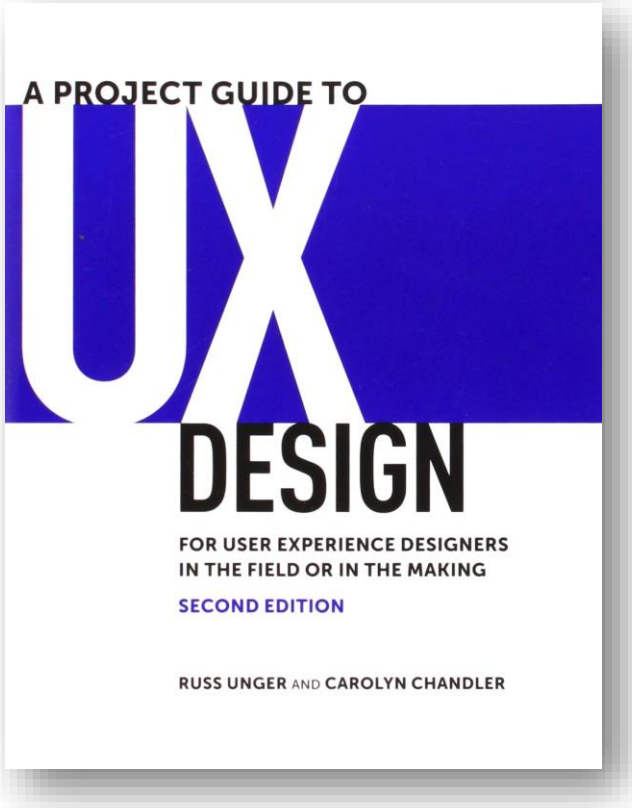
Summary

Today we talked about:

- The psychological basis of usability problems and behaviour on the web:
  - Sensation, perception and gestalt
  - Memory and attention
  - Implications for design
  - User behaviour on the web
- **Usability evaluation methods**
  - How to set goals for your evaluation
  - Analytical methods: heuristics and cognitive walkthroughs
  - User methods: qualitative and quantitative usability tests
  - Eye-tracking and A-B testing

Further reading…

Chapter 14: Design testing



Your project…