

# Refining Designs

# Administrivia

- No class Thursday
  - Screencast on designing UI will be posted
- Deliverables (particularly for final)

# Keeping what works, eliminating what doesn't

- The Sunday edition of a Toronto newspaper contained a fascinating, disturbing story on high-school students who are quite literally incapable of signing their own name. Alas, that would seem destined to include most children: The art of writing longhand is no longer required teaching in most Ontario schools. For today's teenagers it's at best a distant memory; for tomorrow's it will be something akin to hieroglyphics.

– Andrew Coyne, June 25, 2013

# Is Handwriting Useful?

# Is Handwriting Useful?

- Typing is file retrieval, remembering where a letter is. With handwriting, you create the letters anew each time, using much more complex motor skills. Whether it's the flowing motion of the arm, or the feel of the page under your hand, or the aesthetic satisfaction of a well-turned “f”, it seems to engage the more intuitive, right-brain aspects of cognition. Tapping into your intuition is a critical part of writing, or indeed of thinking.

# Is Handwriting Useful?

- But it isn't only its fluidity that makes handwriting a useful aid to thinking. It is also the constraints it imposes. As with many older technologies, its virtues consist partly in its defects. (The problem with radio is you can't watch it; the great thing about radio is you don't need to.) Text on a computer is infinitely corrigible: We commit to nothing, either in words or sentence structure. This frees us to make an incomprehensible mess of things.
- Handwriting, to the contrary, forces us to make an investment. The words are there on the page; we can't change them, except to scratch them out. It inclines us thus to compose the sentence in our heads first — and the sort of sentence you can compose and keep in your head is likely to be shorter and clearer than otherwise. Your readers will generally thank you.
- That's one of the reasons the Beatles' songs were so memorable: they had to compose them in their heads. Neither Lennon nor McCartney could read or write music. So if they weren't finished a song by the end of a day, they had to remember it until the next. If they couldn't remember it, chances are it wasn't worth remembering.

# Is Handwriting Useful?

- The imagination needs to tug on a leash. Nothing is so inhibiting as a blank screen, precisely because the possibilities are infinite. Likewise, anyone who has done improv knows the first thing you do is ask for an audience suggestion — “give me an occupation.” Only after you have narrowed your choices can you begin to expand on them. That’s what putting words down on paper does.
- How we write, in other words, affects what we write. You compose in a different way using pen and ink than you do on a computer. You think in a different way. It may even be that you are, to that extent, a different person, much as we take on a different personality when we speak a foreign language.

# Contextual Design: Stages

- Interviews and observations
- Work modeling
- Consolidation
- Work redesign
- User environment design
- Prototypes
- **Evaluation**
- Implementation

# Evaluation

# Evaluation

- Evaluation for many purposes
- Two forms
  - Quantitative
    - Data involves numerical measures that can be contrasted
  - Qualitative
    - Data is narrative and observational in form
- Can combine
  - Mixed methods
    - Data involves both observation and numerical data

# Goals of evaluation (2)

- To assess extent and accessibility of systems functionality
  - Does system do enough? Can users access functions?
- To assess users' experience of interaction
  - Do they like it? Do they understand it?
- To identify specific problems with system
  - Is something done wrong? Can aspects be improved?
- To understand real world
  - How do users use technology? Can design be improved, can work be automated, can we help a potential user group?
- To compare designs
  - Best/better/worse Essential features
- To engineer toward a target
  - Is design good enough?
- To check conformance to a standard
  - Microsoft design guidelines, Mac interface guidelines

# Quantitative Evaluation

- Postivist/Postpositivist claims and testing
- Experimental method
  - Hypothesis
  - Typical measures
  - Test
  - Evaluate results
- Confounds
  - Example

# Hypothesis

- State something that you believe to be true
- Must be disprovable in a finite amount of time
  - Can design an experiment to test
  - The experiment will be of reasonable duration
- Bad examples:
  - There is intelligent extra-terrestrial life
  - There is no intelligent extra-terrestrial life
- Good examples:
  - Interface A is faster than interface B
  - Interface A results in lower errors than interface B
  - Users prefer interface A to interface B

# Quantitative Evaluation

- Can be hard to control for confounds
- Solution?
  - Punt
  - Usability engineering
  - Define metrics
    - Time to accomplish a task
    - Error rate
    - User satisfaction
    - Etc.
  - Keep re-engineering until you reach metrics
  - Note that metrics can interact

# Quantitative Evaluation

- Generally useful late in design
  - Given two systems, can we evaluate their relative performance
  - Need careful metrics
- Also used for novel interaction techniques
  - Given a new way of selecting, is it faster, less error prone, etc.
- Not typically used in design

# Evaluation

- Evaluation for many purposes
- Two forms
  - Quantitative
    - Data involves numerical measures that can be contrasted
  - Qualitative
    - Data is narrative and observational in form
- Can combine
  - Mixed methods
    - Data involves both observation and numerical data

# Testing Low-Fidelity Prototypes

- Low-fidelity prototypes are tested in unique ways
  - No system, only rough screen shots
- Goal is to understand “what user is thinking”
  - Need techniques that prompt for this
- Common approaches
  - Person down the hall testing
  - Walkthroughs
  - Thinkalouds

# Person down the hall testing

- Common in the real world; also, basically, goal of last poster session
- When people come to your poster
  - Select someone to walk through the interaction
  - Others watch
  - Collect feedback
- In real world
  - Walk colleague through task, how users work now, and how you are changing work
  - Then show prototypes

# Walkthroughs

- A series of sketches
- Walk user representatives through different screen shots
- Ask users what they would do on each screen
- Advantages
  - Fast overview of system
  - Very useful for early stage sketches
- Disadvantages
  - Feedback limited by no “doing”
  - Risk of over-control of execution by experimenter
- Can augment walkthroughs with “think-aloud” protocol

# Thinkalouds

- Two methods
  - Retrospective
    - Capture video of users using system
    - Watch video with users
    - Users comment on their actions and present their thinking
    - Very common with Difficult-to-evaluate systems like ATC
    - Can introduce post-hoc rationalizations
  - Concurrent
    - Very typical during design
    - You will do this

# Concurrent Thinkalouds

- Observe user using your prototype
- Encourage them to “think-aloud”
  - Express what they are thinking and wondering at each moment
- When user is not having problems they work fast
  - Faster than they think
- When user is having problems, they slow down
  - Think aloud can reveal aspects of bad mental models, poor affordances, insufficient constraint, poor feedback, etc.
- Sometimes, when under heavy load, user will pause
  - Essential to continue to encourage them to think-aloud, but in a friendly way
- Tasks can be specified (“Could you schedule a reservation?”) or open-ended (user chooses what he/she would like to do with system)
- Informal technique – creating an informal atmosphere will result in more successful session

# Goals of evaluation

- Design versus implementation
  - Formative evaluation is used during development
  - Summative evaluation is used for finished product
- Can help to align models
  - Designer's model
  - User's mental model

# Conducting concurrent think-alouds

- Settle on task
  - Vertical or horizontal testing?
- Settle on exactly what you want to tell user
  - You want to give appropriate level of direction
  - If using Anoto pen, need to communicate how technology works
  - If using a traditional interface, need to communicate purpose of system
- Think about how much help you want to give
  - You want an honest assessment
- Two people maximum at think-aloud
- The interface, not the person, is under scrutiny
  - How they work is how they work
  - You want an interface that will be easily incorporated into work practice
  - Let them know that you will be providing only limited help, and apologize for this in advance

# Conducting concurrent think-alouds

## (2)

- One of you take the lead and greet the person
  - Put them at ease, describe process, give them information on what you are testing
  - Pleasant expression
- Person who greets should observe
  - Maintain pleasant expression
  - Set up audio recording
  - Get notebook ready and ask them to start (the task you give or the tasks they typically would do)
  - Take notes as they work (supplements audio recording)
  - Prompt during silences
    - ASK: What are you thinking now?
    - NOT: Why did you do that?

# Conducting concurrent think-alouds

## (3)

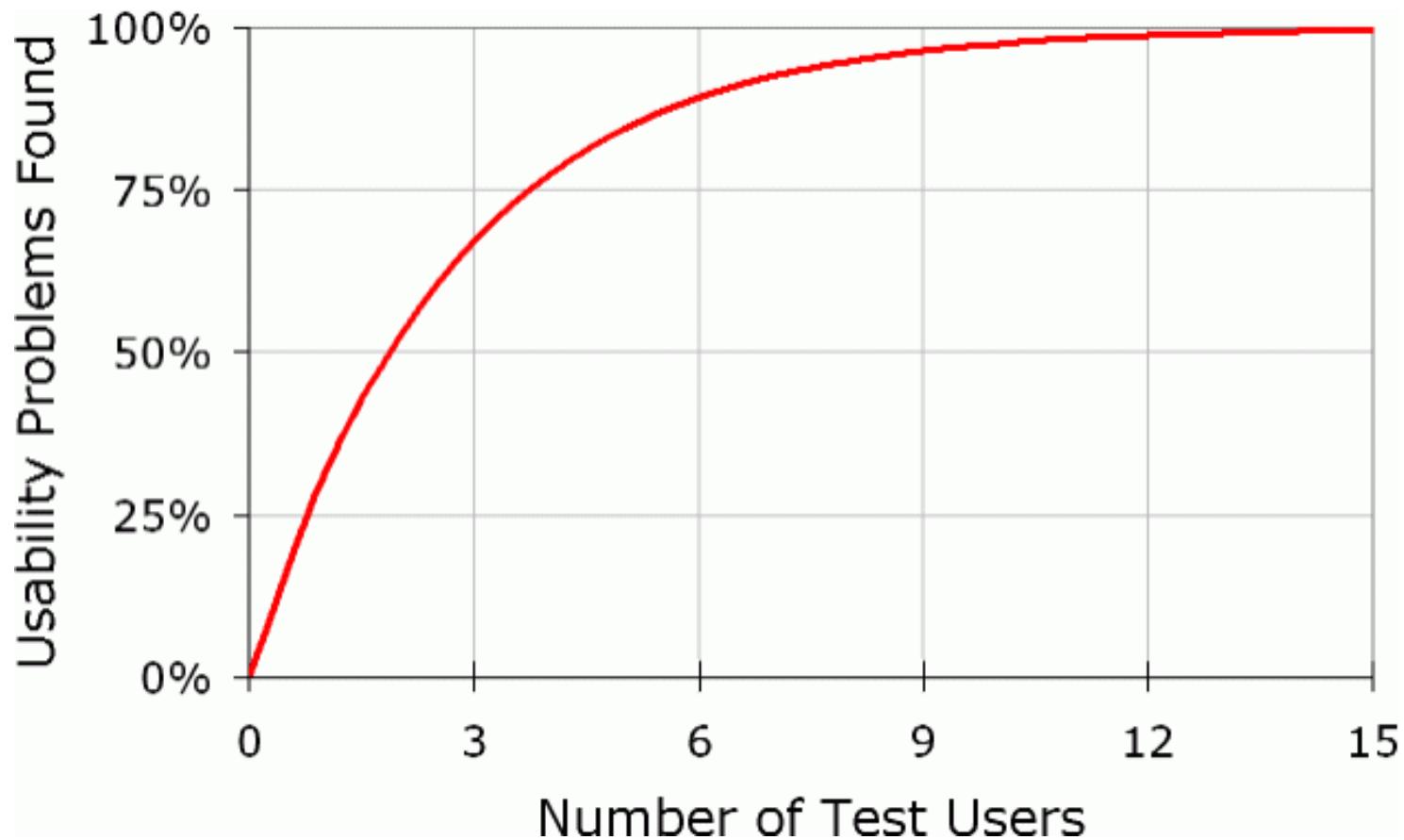
- After they finish, debrief
  - Look to your notes for points you would like clarification on
  - Ask them for overall impressions of the system
    - Biology example
- Thank your users
- After session
  - Get together with your group asap
  - Walk through your notes, use audio, and make an affinity diagram of data
  - Look for themes you can use to improve prototype
- Iterative on prototype (if possible) and conduct walkthrough with other participant

# Conducting concurrent think-alouds

## (4)

- Advantages
  - Not limited to paper prototypes
    - Mathbrush
  - Rapid, high-quality qualitative feedback
  - Data is as rich as with contextual inquiry
    - Observations, hearing
  - Can interact with subject to get complete information
  - Can help subject if it becomes necessary
  - Flexibility in initiative
  - Doing, so less opportunity to give rote positive assessment
- Disadvantages
  - Limited sample?

# Recall: Why you only need to test with five users

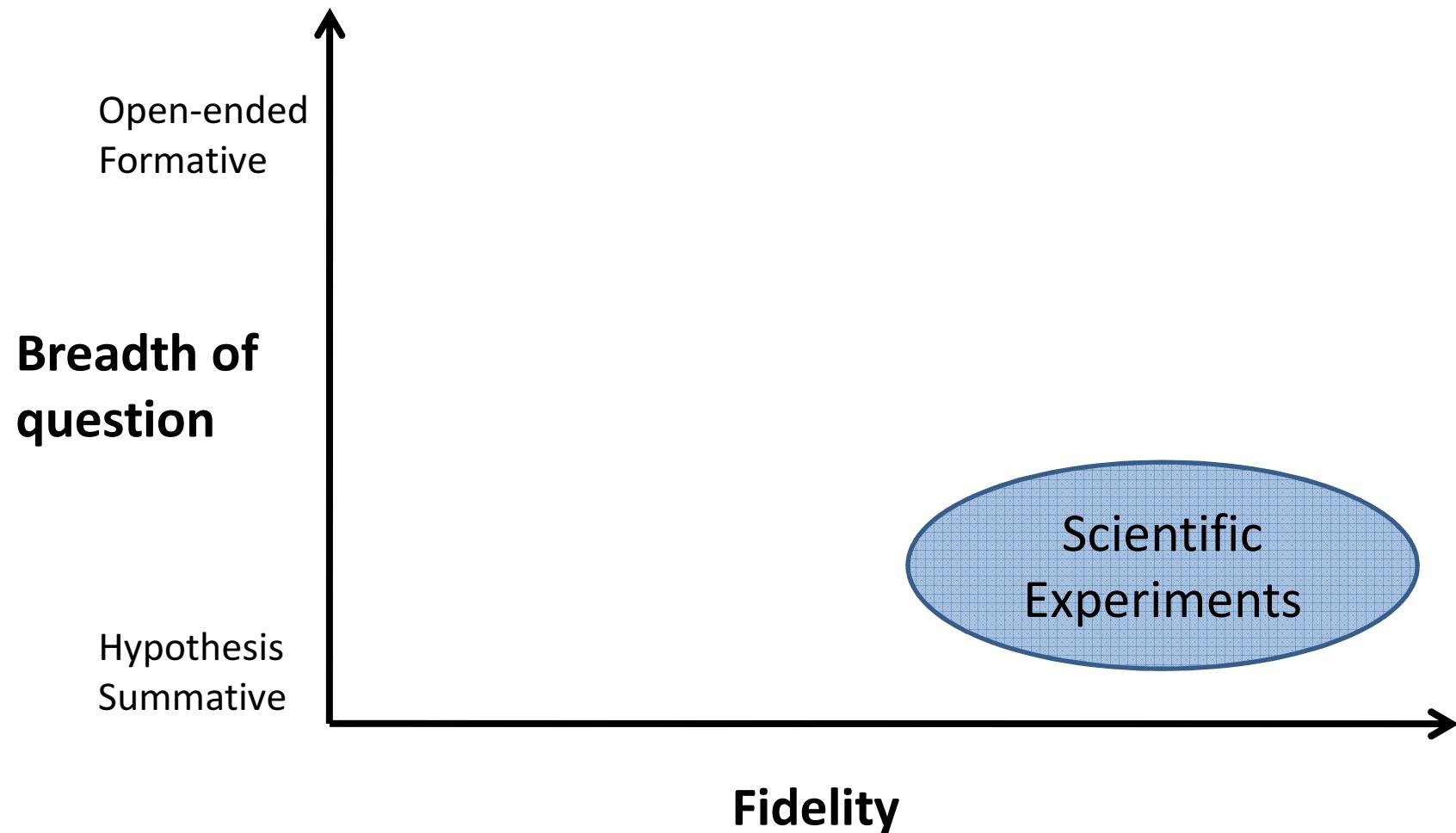


**But recall the assumption that any usability problem typically affects 31% of users**

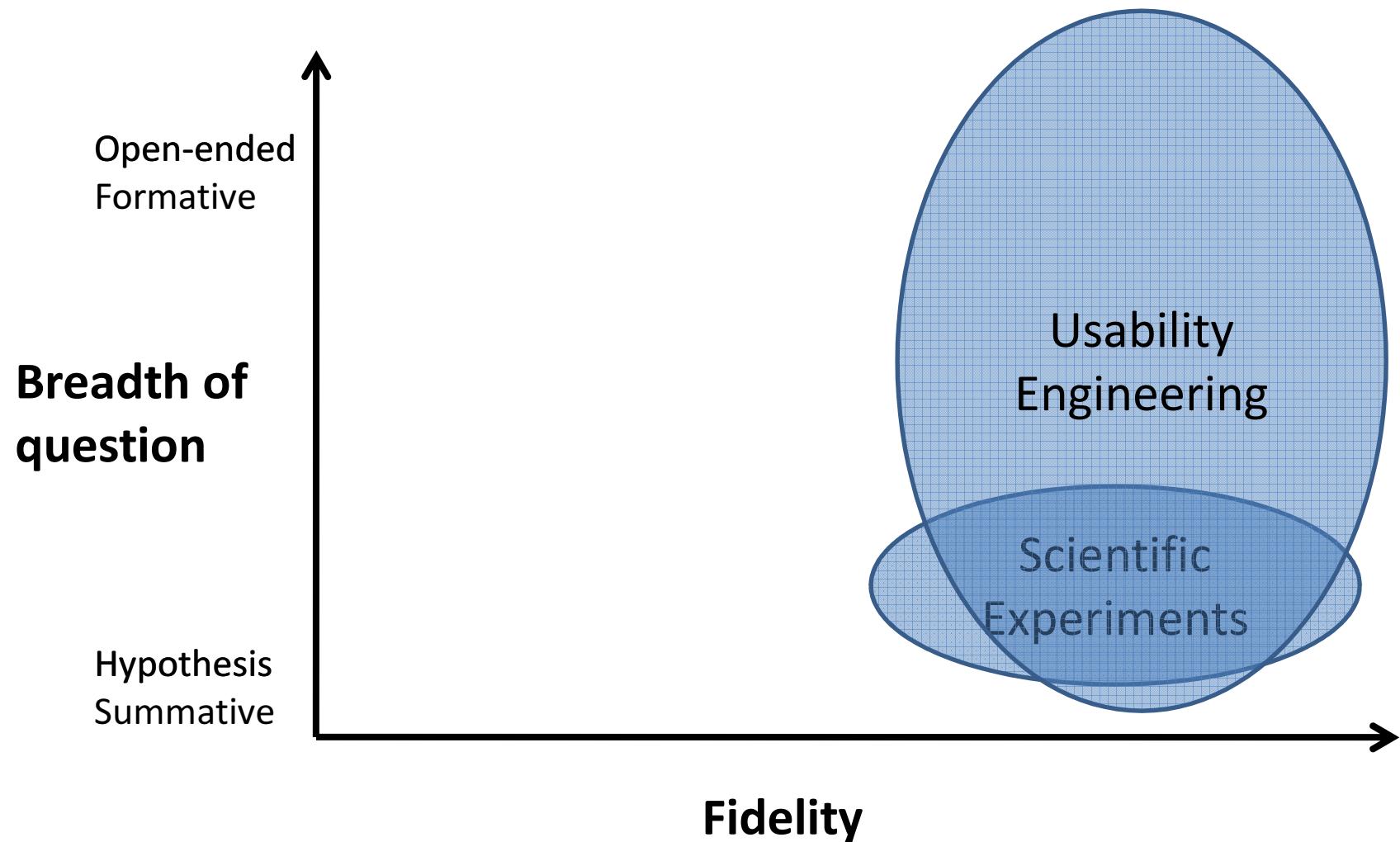
# Refining Designs

- Bring sketching paper to evaluation sessions for prototypes
- Evaluation is ‘sweet-spot’ in contextual design for transition to participatory design

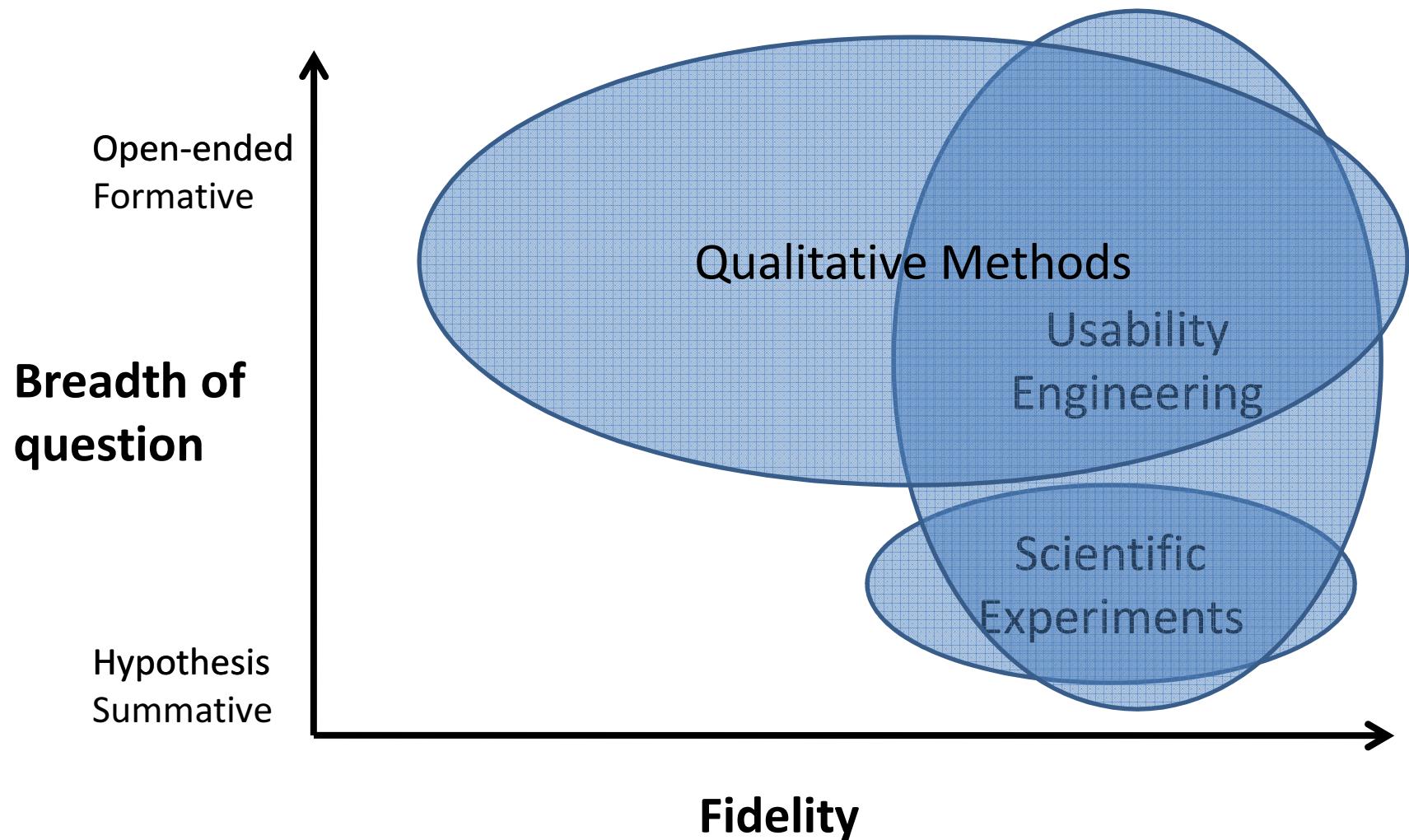
# A Design Space for Evaluation



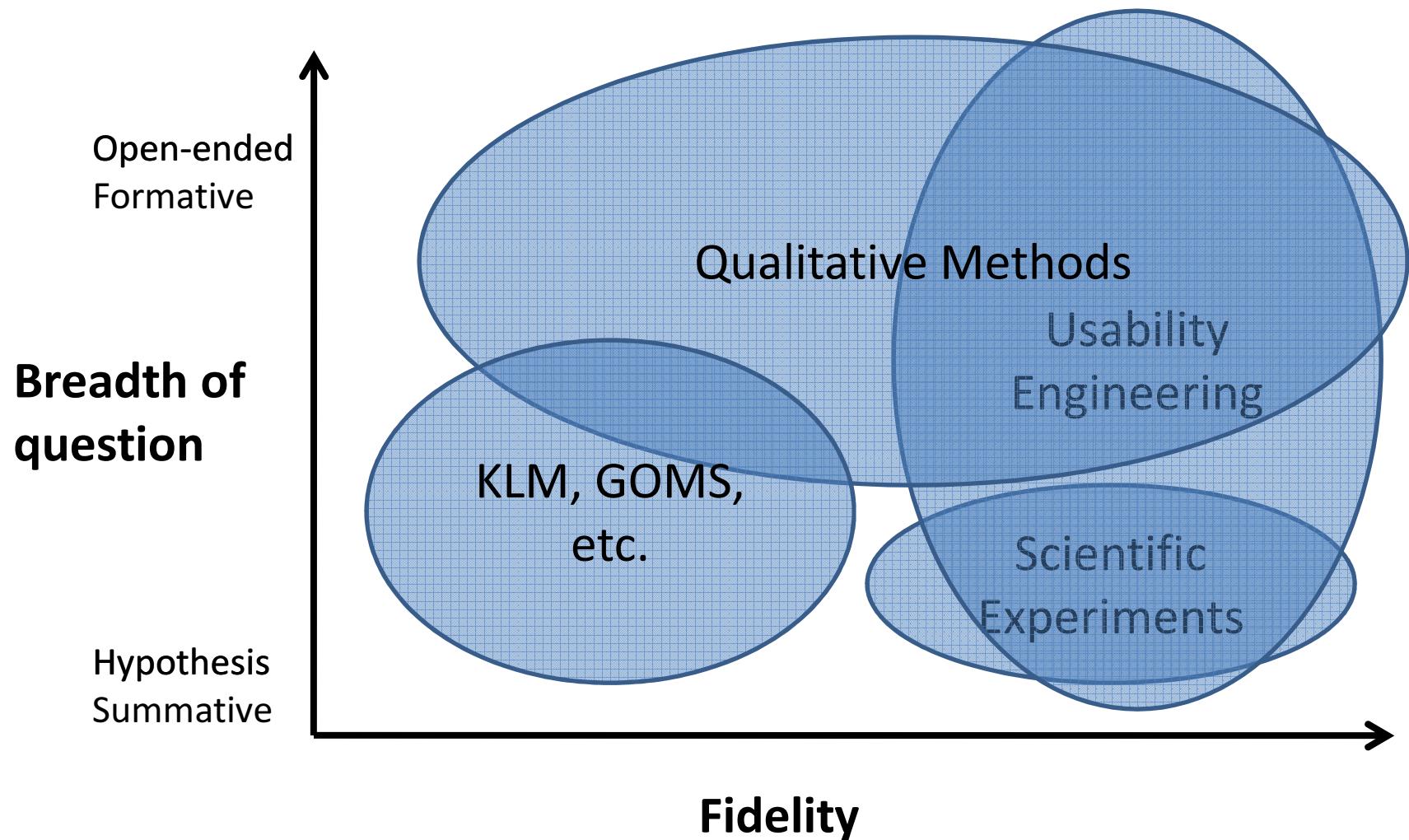
# A Design Space for Evaluation



# A Design Space for Evaluation



# A Design Space for Evaluation



# Experimental Biases in the RW

- Hawthorne effect/John Henry effect
- Experimenter effect/Observer-expectancy effect
- Pygmalion effect
- Placebo effect
- Novelty effect

# Hawthorne Effect

- Named after the Hawthorne Works factory in Chicago
- Original experiment asked whether lighting changes would improve productivity
  - Found that anything they did improved productivity, even changing the variable back to the original level.
  - Benefits stopped studying stopped, the productivity increase went away
- Why?
  - Motivational effect of interest being shown in them
- Also, the flip side, the John Henry effect
  - Realization that you are in control group makes you work harder

# Experimenter Effect

- A researcher's bias influences what they see
- Example from Wikipedia: music backmasking
  - Once the subliminal lyrics are pointed out, they become obvious
- Dowsing
  - Not more likely than chance
- The issue:
  - If you expect to see something, maybe something in that expectation leads you to see it
- Solved via double-blind studies

# Pygmalion effect

- Self-fulfilling prophecy
- If you place greater expectation on people, then they tend to perform better
- Studied teachers and found that they can double the amount of student progress in a year if they believe students are capable
- If you think someone will excel at a task, then they may, because of your expectation

# Placebo Effect

- Subject expectancy
  - If you think the treatment, condition, etc has some benefit, then it may
- Placebo-based anti-depressants, muscle relaxants, etc.
- In computing, an improved GUI, a better device, etc.

# Novelty Effect

- Typically with technology
- Performance improves when technology is instituted because people have increased interest in new technology
- Examples: Computer-Assisted instruction in secondary schools, computers in the classroom in general, etc.

# Controlling for Biases?

- Cannot fully
  - More an awareness issue
- Approach any test data with some skepticism
- Assume subjects are trying to be helpful, so any errors must be pretty serious
- Aggressively seek contradictory data