EVALUATION

PROBLEM-DRIVEN VS TECHNIQUE-DRIVEN

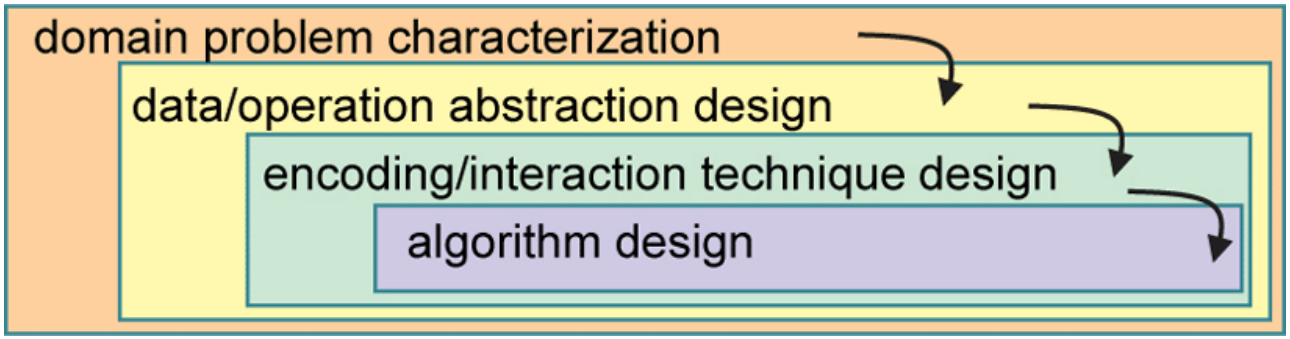
Problem-driven

- top-down approach
- identify a problem encountered by users
- design a solution to help users work more effectively
- sometimes called a design study

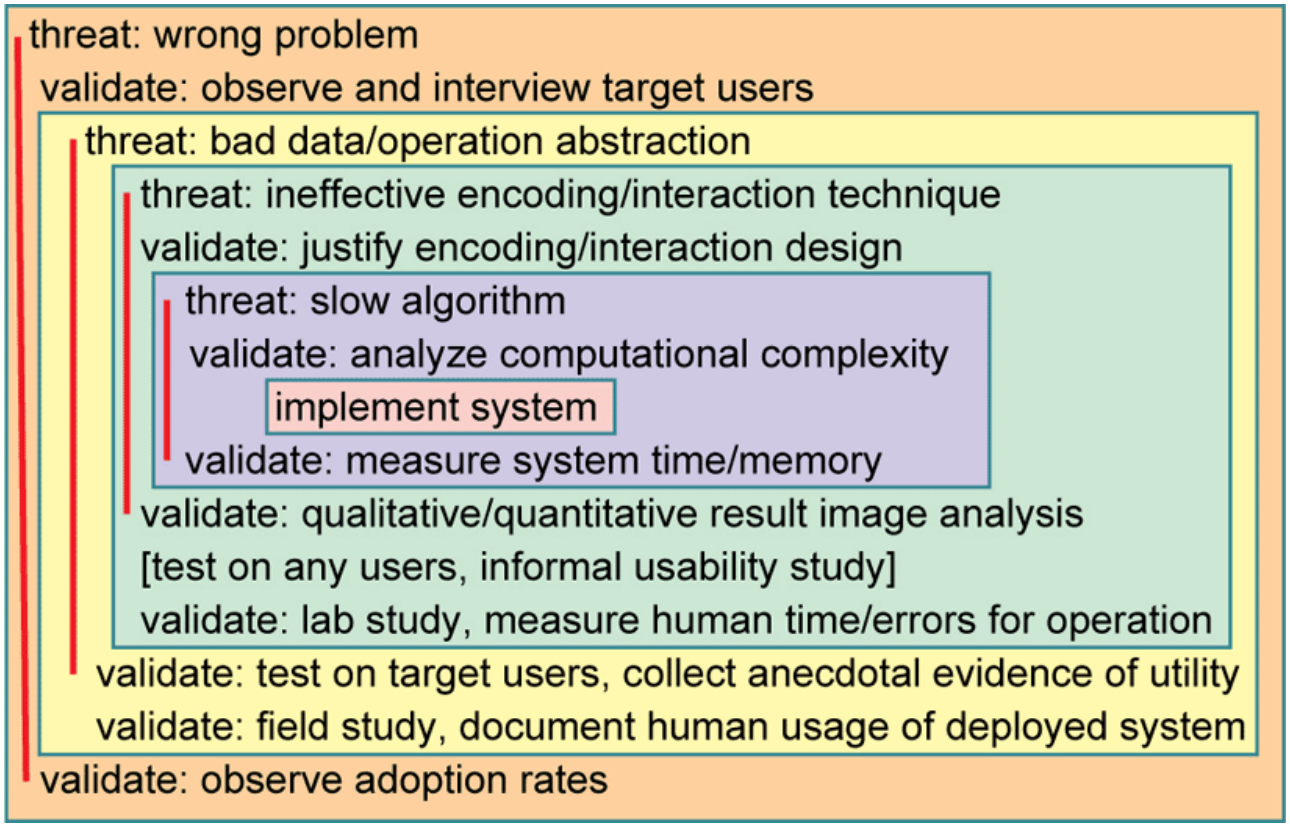
Technique-driven

- bottom-up approach
- invent new visualization techniques or algorithms
- classify or compare against other idioms and algorithms

THE NESTED MODEL

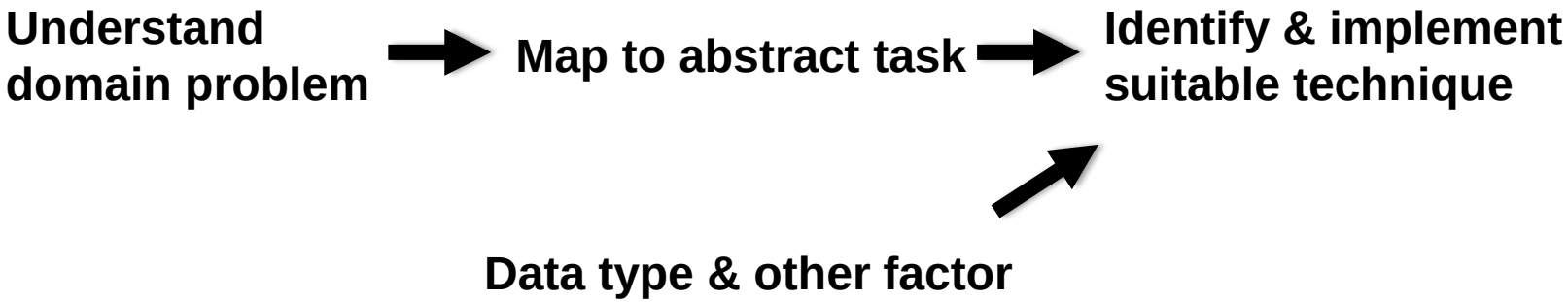


Design



Threats & evaluation

DESIGN PROCESS



DOMAIN CHARACTERIZATION

Details of an application domain

Group of users, target domain, their questions, their data

- varies wildly by domain
- must be specific enough to continue with

Cannot just ask people what they do

- introspection is hard

DOMAIN PROBLEM CHARACTERIZATION

Infinite numbers of domain tasks

Can be broken down into simpler abstract tasks

We know how to address the abstract tasks

Identify task – data combination:
solutions probably exist

EXAMPLE: FIND GOOD MOVIES

I want to identify good movies in genres I like

Domain: general population, movie enthusiasts

DATA & TASK ABSTRACTION

The what-why, map into generalized terms

Identify tasks that users wish to perform or already do

Find data types and good model of the data

Sometimes must transform the data for a better solution

this can be varied and guided by the specific task

EXAMPLE: FIND GOOD MOVIES

What is a good movie for me?

- Highly rated by critics?
- Highly rated by audiences?
- Successful at the box office?
- Similar to movies I liked?
- Specific genres?

Data sources: IMDB, Rotten Tomatoes, …

ENCODINGS & INTERACTIONS

The design of visualization techniques

- Visual encodings
- Interactions

Ways to create and manipulate the visual representation of data

Decisions on these may be separate or intertwined

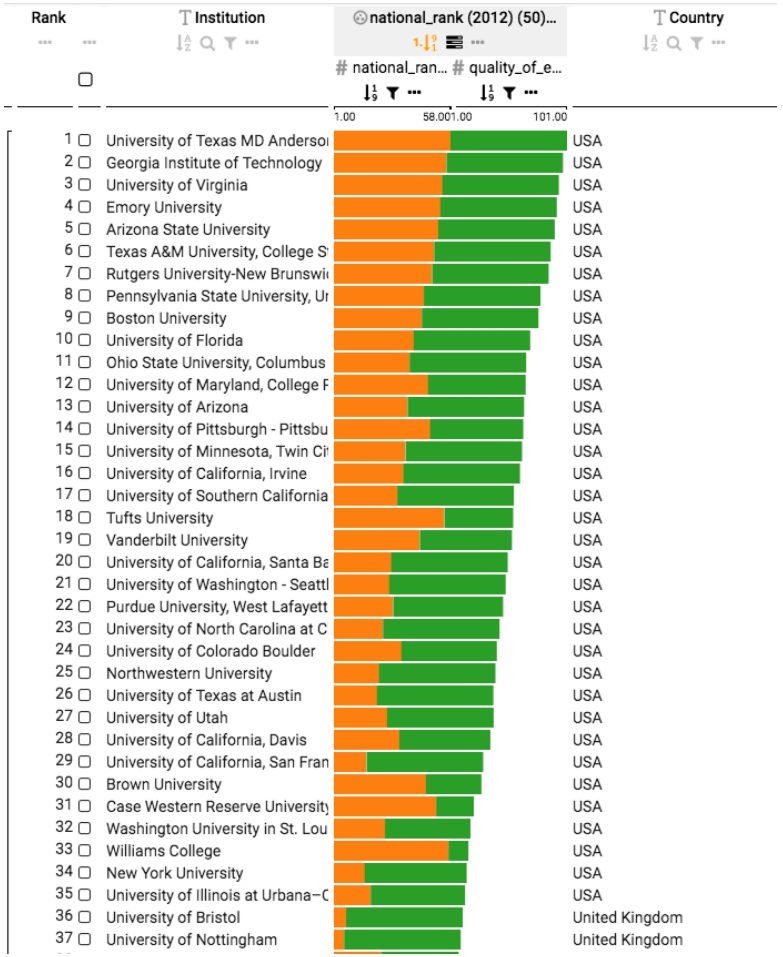
Visualization design principles drive decisions

EXAMPLE: FIND GOOD MOVIES

Combination of audience ratings and critics ratings, filtered by genre

Idiom:

- **stacked bar chart for ratings**
- **filter interface for genre**



TASKS

Analyze

- High-level choices
- Consume vs produce

Search

- Find a known/unknown item

Query

- Find out about characteristics of item
- By itself or relative to others

HIGH-LEVEL ACTIONS: ANALYZE

Consume

Discover vs present

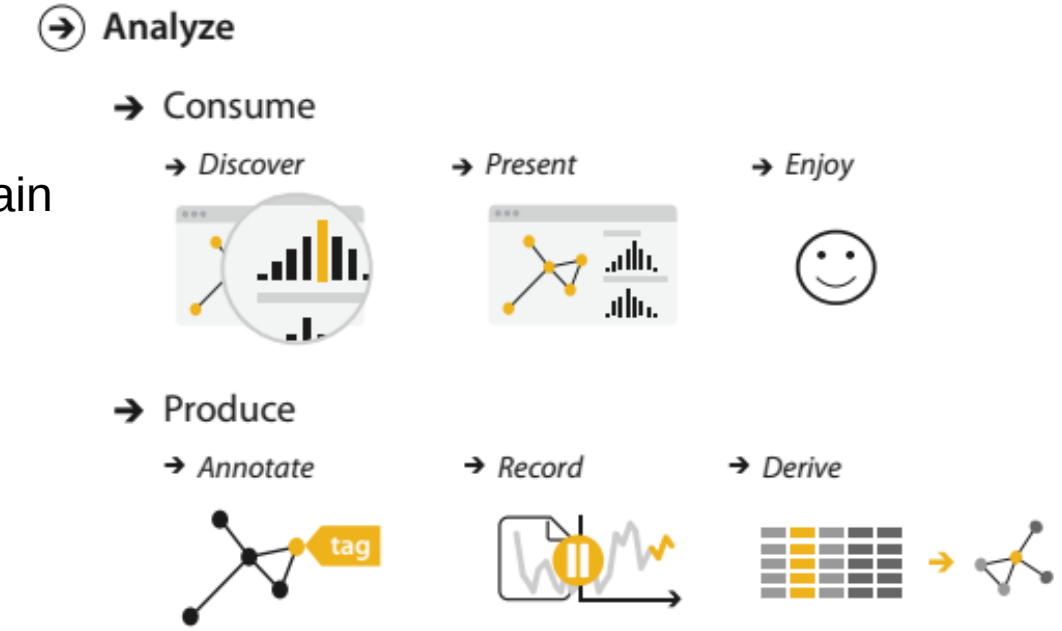
Classic split: explore vs explain

Enjoy: casual, social

Produce

Annotate, record

Derive: crucial design choice



MID-LEVEL ACTIONS: SEARCH, QUERY

Search: what does user know?

Target, location

Search

	Target known	Target unknown
Location known	 *Lookup*	 *Browse*
Location unknown	 *Locate*	 *Explore*

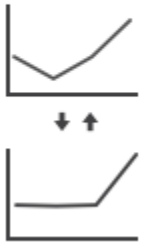
How much of the data matters?

One, some, all

Query

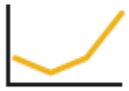
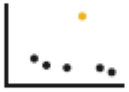
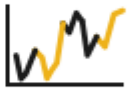
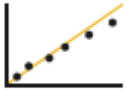
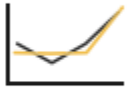
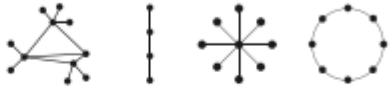
→ Identify

→ Compare



→ Summarize

LOW LEVEL: TARGETS

- **All Data**
  - → Trends 
  - → Outliers 
  - → Features 
- **Attributes**
  - → One
    - → *Distribution* 
    - → *Extremes* 
  - → Many
    - → *Dependency* 
    - → *Correlation* 
    - → *Similarity* 
- **Network Data**
  - → Topology 
  - → *Paths* 
- **Spatial Data**
  - → Shape 

DESIGNING VISUALIZATIONS

WHAT IS DESIGN?

creating something new to solve a problem

can be used to make buildings, chairs, user interfaces, etc.

design is used in many fields

many possible users or tasks

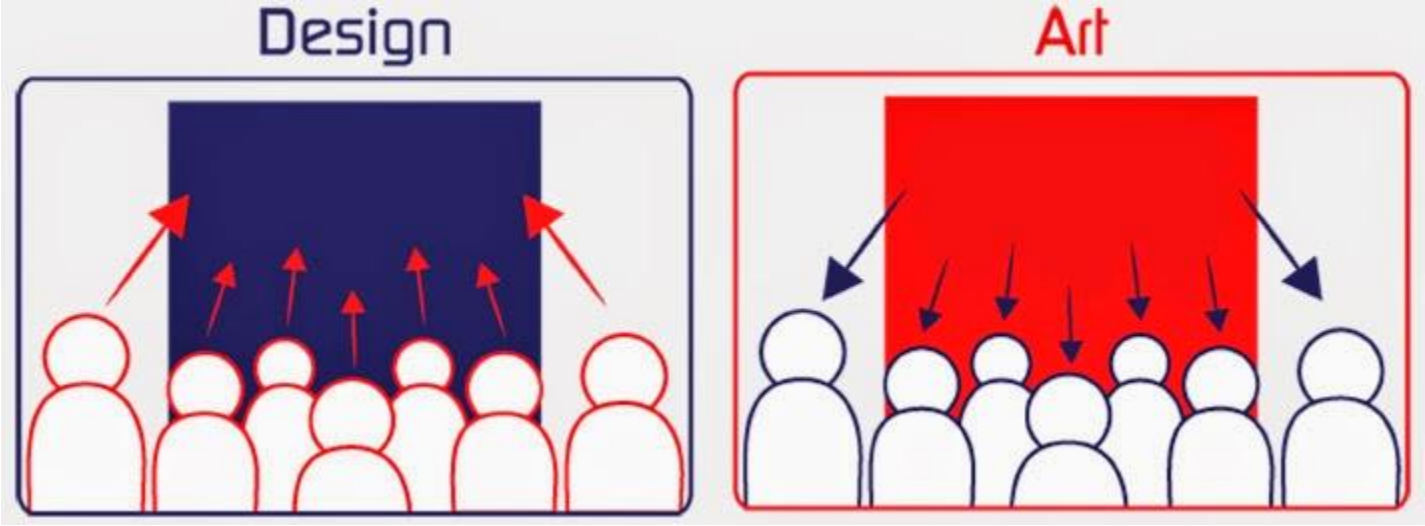
WHAT IS DESIGN NOT?

just making things pretty

art – appreciation of beauty or emotions invoked

something without a clear purpose

building without justification or evidence



FORM & FUNCTION

Commonly: “Form follows function”

Function can constrain possible forms

Form depends on tasks that must be achieved

“The better defined the goals of an artifact, the narrower the variety of forms it can adopt”

- Alberto Cairo

WHEN DO WE DESIGN?

Wicked problems

- No clear problem definition
- Solutions are either good enough or not good enough
- Multiple solutions exist, not true/false
- No clear point to stop with a solution

Examples of non-wicked (“tame”) problems

- Mathematics, chess, puzzles

WHY DOES DESIGN MATTER FOR VIS?

Many ineffective visualization combinations

Users with unique problems & data

Variations of tasks

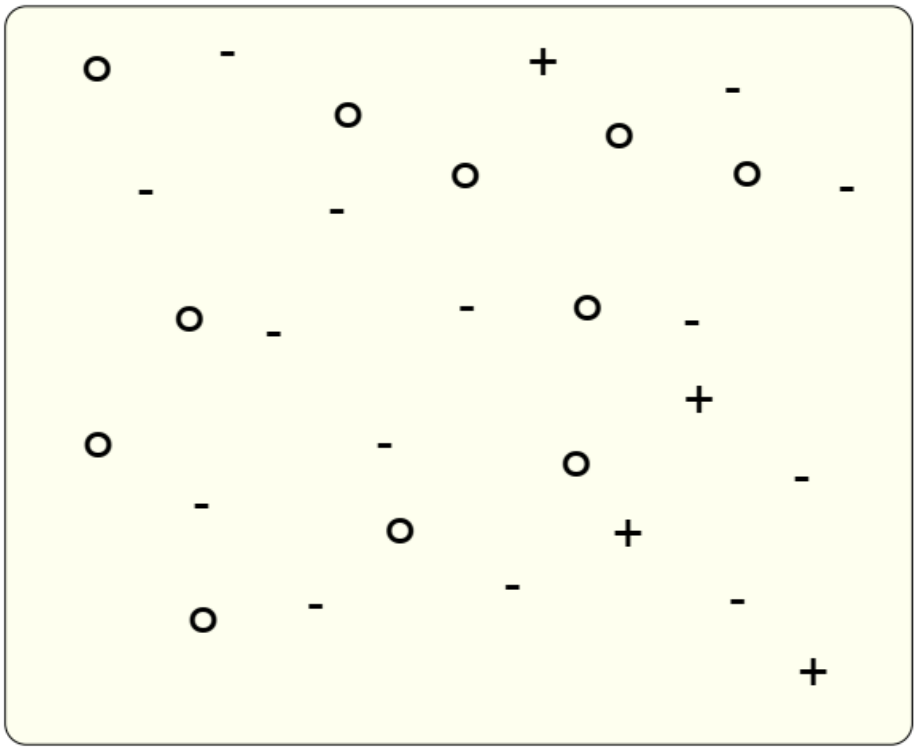
Large design space

Pitfall

Premature Design Commitment

by T. Munzner

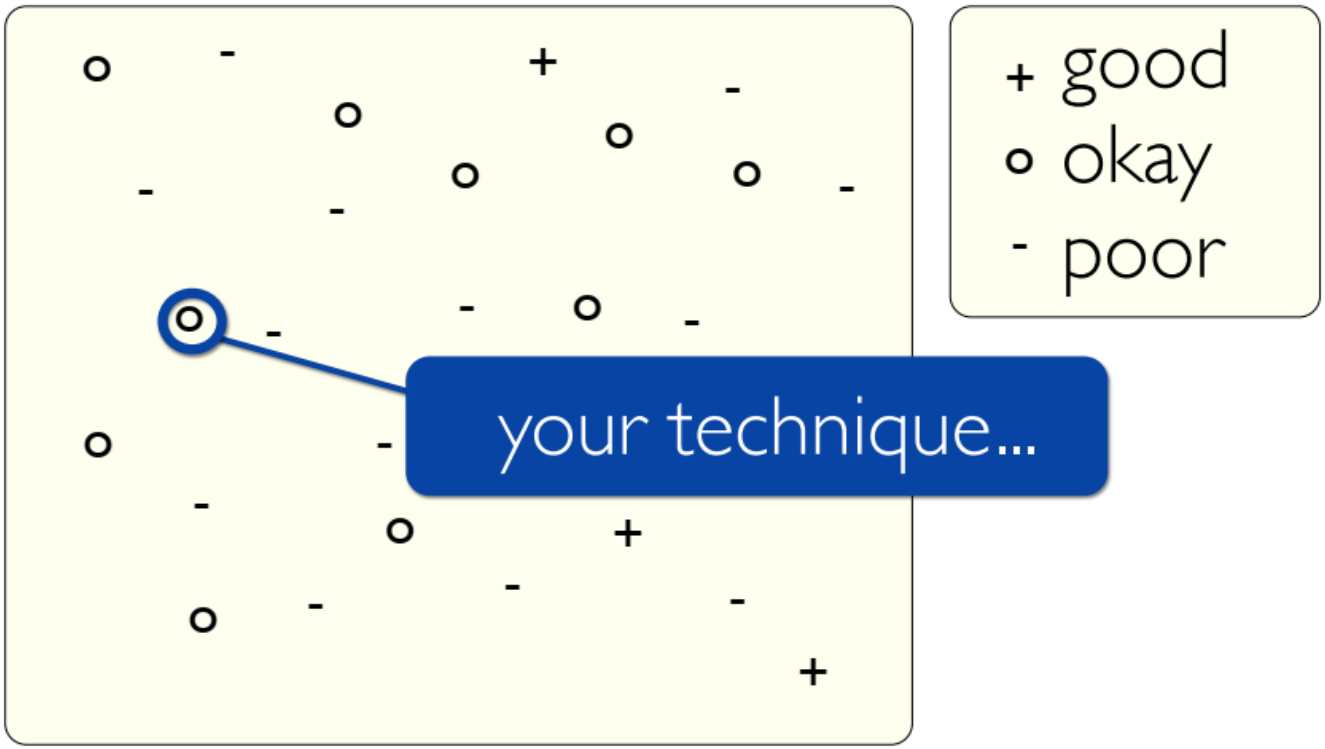
METAPHOR
Design Space



+ good
o okay
- poor

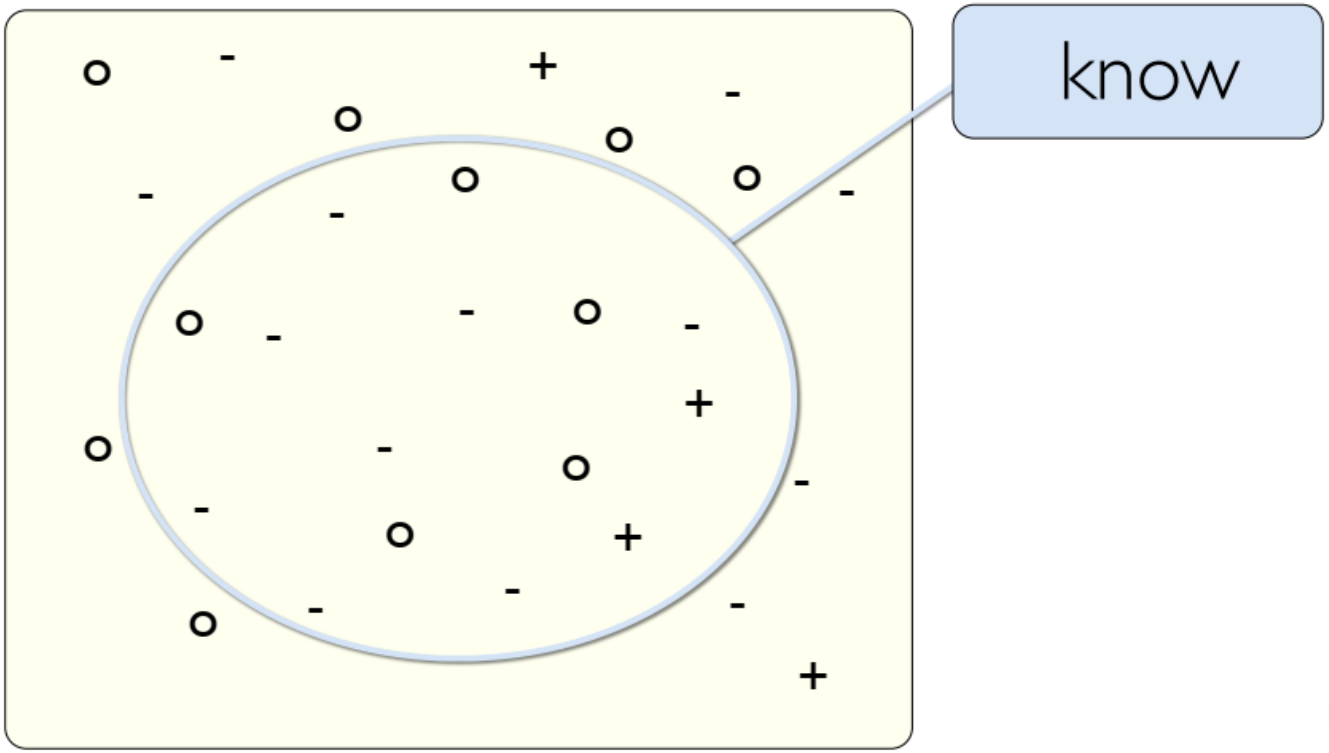
by T. Munzner

METAPHOR
Design Space



by T. Munzner

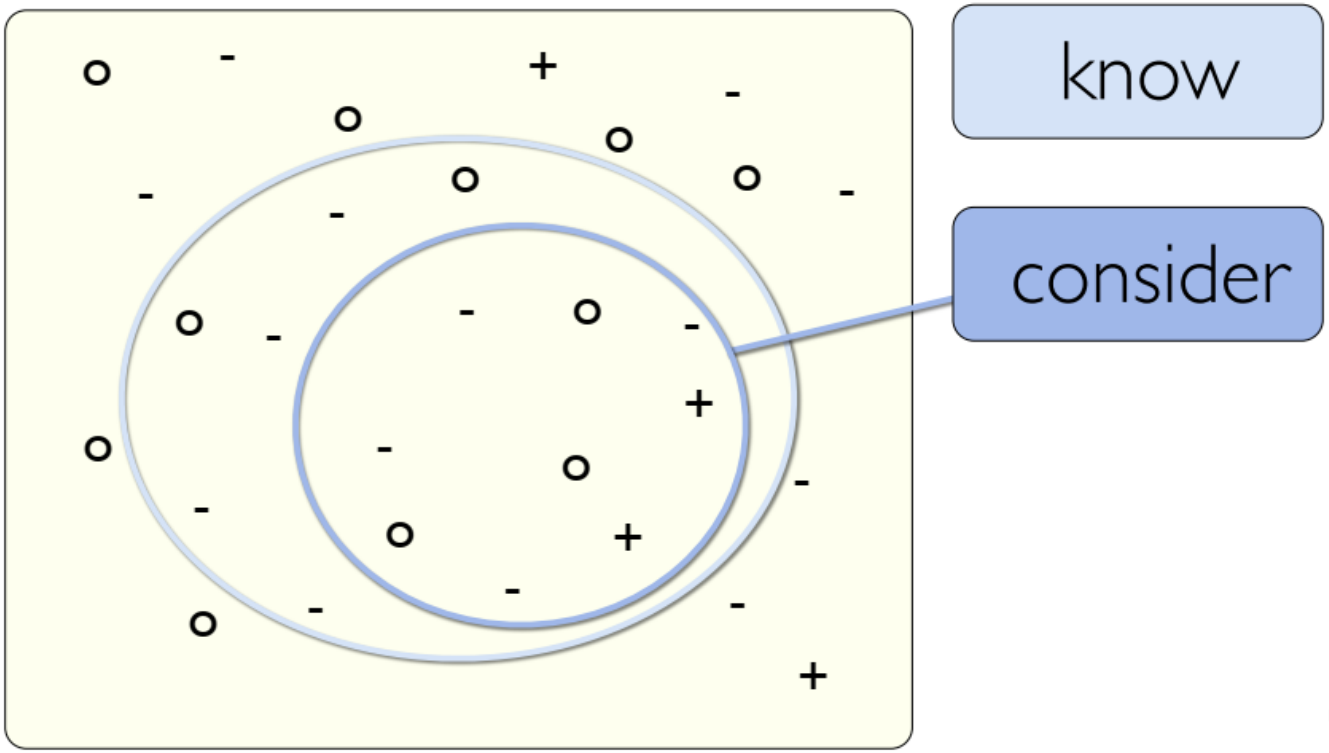
METAPHOR
Design Space



by T. Munzner

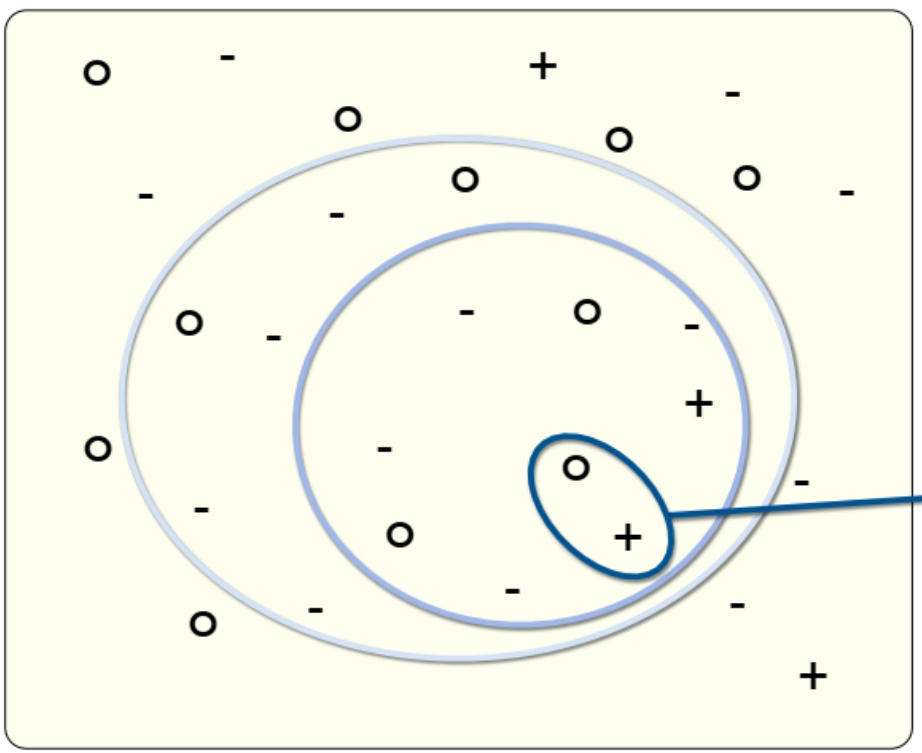
METAPHOR

Design Space



by T. Munzner

METAPHOR
Design Space



know

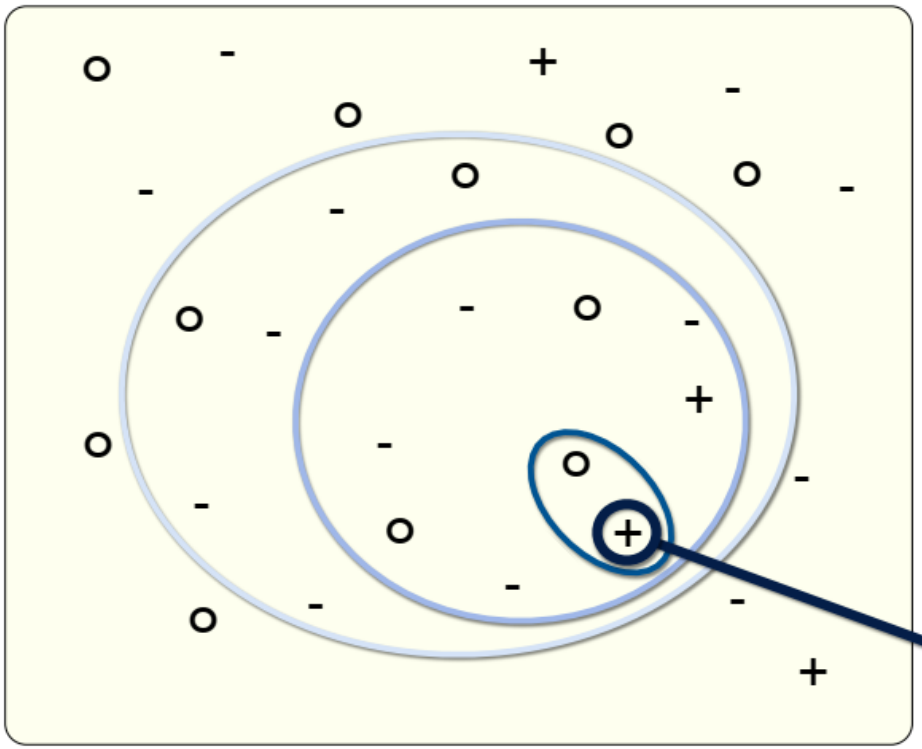
consider

propose

by T. Munzner

METAPHOR

Design Space



know

consider

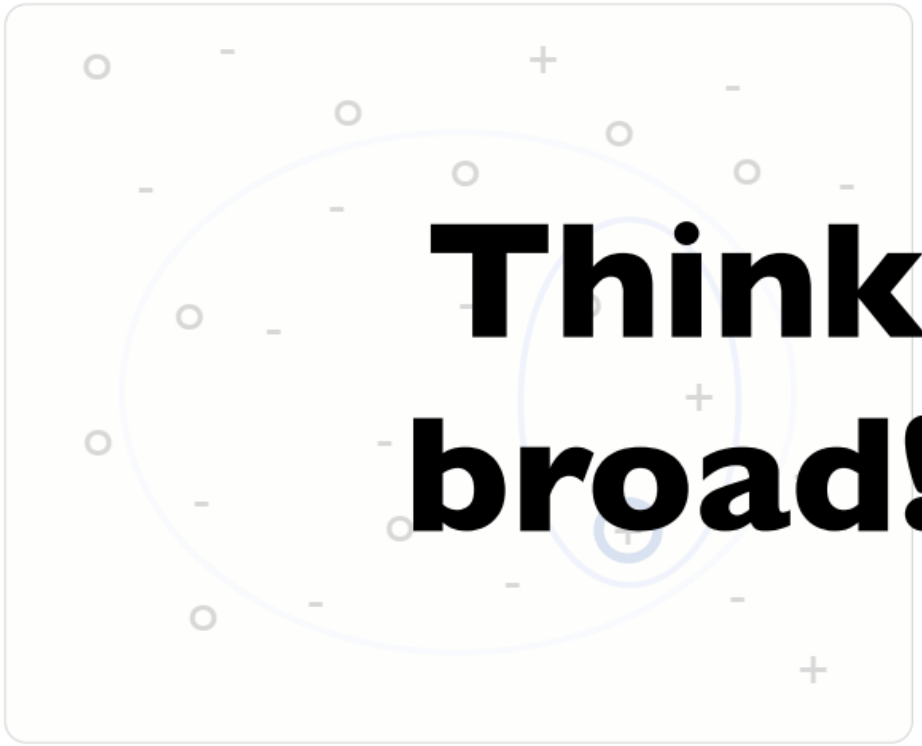
propose

select

by T. Munzner

Metaphor

Design Space



Think broad!

+ good
o okay
- poor

consider

propose

select

by T. Munzner

40

EVALUATION METHODS

Controlled experiment

Laboratory, crowd-sourced

Interviews/questionnaires

Unstructured, structured, semi-structured

Field observation, lab observation

Video/audio analysis

Coding/classification of user behavior (speech, gestures)

Log analysis

Algorithmic performance measurement

EVALUATION METHODS

Heuristic evaluation

Judge compliance with recognized metrics/usability methods (heuristics)

Usability testing, e.g., thinking aloud tests

Wizard of Oz

Human simulates response of system

Test functionality before it's implemented

Eye tracker evaluation

Expert evaluation

Insight-based evaluation

Case studies

QUANTITATIVE VS QUALITATIVE EVALUATION

Quantitative methods

Objective metrics, measurements

Use numbers/statistics for interpreting data

Qualitative methods

Subjective metrics

Description of situations, events, people, interactions, and observed behaviors, the use of direct quotations from people about their experiences, attitudes, beliefs, and thoughts

Focused on understanding how people make meaning of and experience their environment or world

INTERNAL VS EXTERNAL VALIDITY

Internal validity – can you trust your experiment

High when tested under controlled lab conditions

Observed effects are due to the test conditions (and not random variables)

External validity – is your experiment representative of real world usage

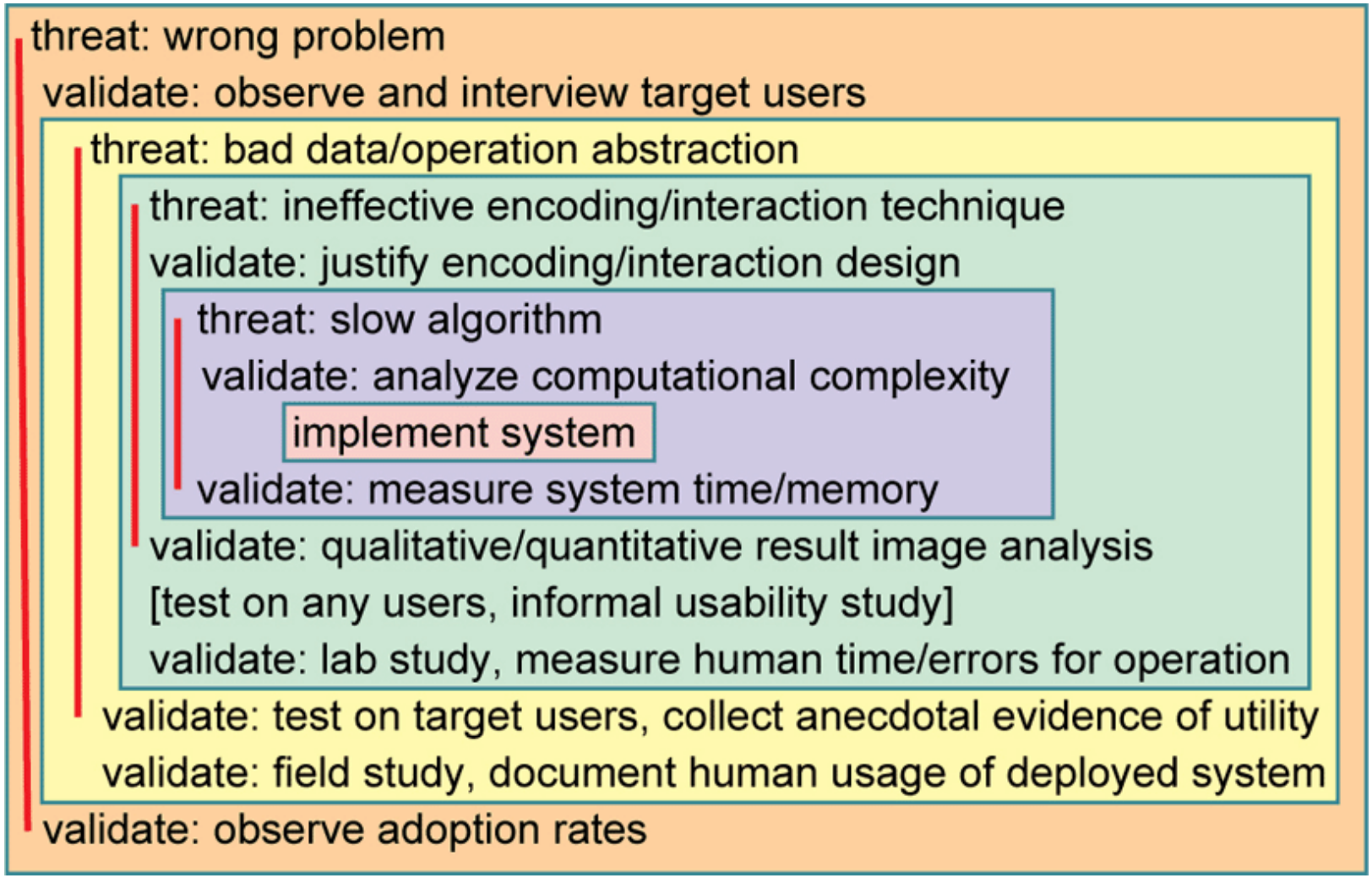
High when interface is tested in the field, e.g. handheld device tested in museum

Results are valid in real world

The trade-off

The more akin to real-world situations, the more experiment is susceptible to uncontrolled sources of variation

SCOPE OF EVALUATION



SCOPE OF EVALUATION

Pre-design

To understand potential users' work env. And workflow

Design

To scope a visual encoding and interaction design space based on human perception and cognition

Prototype

to see if a visualization has achieved its design goals, to see how a prototype compares with the current state-of-the-art systems or techniques

Deployment

to see how a visualization influences workflow and work processes, to assess the visualization's effectiveness and use in the field

Re-design

to improve a current design by identifying usability problems