Defining and Eliciting Optimality Criteria

WHAT IS GOOD?
Optimise a system based on measurements of the user (or hypothetical user)

Measurements of the human-system loop
Criterion types

- Performance, objective
  - Performance measures: efficiency, error rates, action times
  - (e.g. ITR, empowerment)

- Generally easy to measure via instrumentation
Information theoretic

- Information theoretic approaches
  - Transfer rate
    - bits/second is a universal measure
  - Empowerment
    - Control over the environment
  - Novelty
    - In inputs and outputs
    - Entropy measures
Criteria Types

- Human centered, subjective
  - Ergonomics
  - Aesthetics
  - Satisfaction
  - Cognitive load
  - Delight

- Generally hard to measure
  - Noisy, expensive acquisition methods

- Scaling and weighting
  - Comparable units consistent across trials
• Absolute estimation of objective function
• Local gradients
  ○ Randomised trials (e.g. AB) to estimate gradient

• Discrete (e.g. keyboard layouts, menu organisation)
• Continuous (e.g. cursor gain)

• Pre-design (e.g. minimise number of operations)
• Post-design (e.g. maximise text entry rate with a keyboard)
Approaches to elicitation

- Simulation approaches
  - Fast, but are they realistic?

- Mass experimental approaches
  - Can the experiment be packaged up?
Approaches to elicitation

- **Meta-metric learning**
  - Optimising the optimisation...

- **Active learning**
  - Identify weakest parts of the model
    - E.g. in probabilistic methods
Issues

- Cost of elicitation vs. benefit
- High-dimensional objective functions
  - Symmetry analysis
- Getting users to understand when things can be better
• Learning curves
  ○ Is data at t=1hr useful?

• Consistency/robustness of measurements
  ○ Do we converge to population optima?
  ○ Stability of criteria is important
Criteria

- **Measurability** (is this an accessible variable?)
- **Scalability** (can this be scaled up?)
- **Stability** (is this criterion noisy?) (over users, over time…)
- **Smoothness** (are there discrete jumps in the landscape? Is it spiky? Are there plateaus?)
- **Gradient** (can we measure the function, its derivative, or just comparisons)?
- **Continuity** (Continuous scale of performance like ITR, or discrete measurements like a Likert scale)
- **Meaning** (can designers understand the impact criterion?)
- **Phase** (pre-users: designers or with end-users?)
- **Weighting** (can this meaningfully be combined with other variables?)
- **Learning effect** (is a longitudinal study required?)
- **Uncertainty** (can we quantify the uncertainty involved?)
- **User cost** (how much user effort is needed? Is this exhausting to measure?)
- **Economic** (how does this affect stakeholders)
- **Designer empowerment** (does this still afford designers control?)
- **Computational complexity** (how expensive will be to optimise this function)
- **Theoretical plausibility** (how plausible is this function given theoretical models? How well will it generalise?)
- How do these map onto mathematical optimisation approaches?
- How do they map onto existing design practice?