



Experiential Elicitation using Bound Queries

Bowen Hui & Craig Boutilier

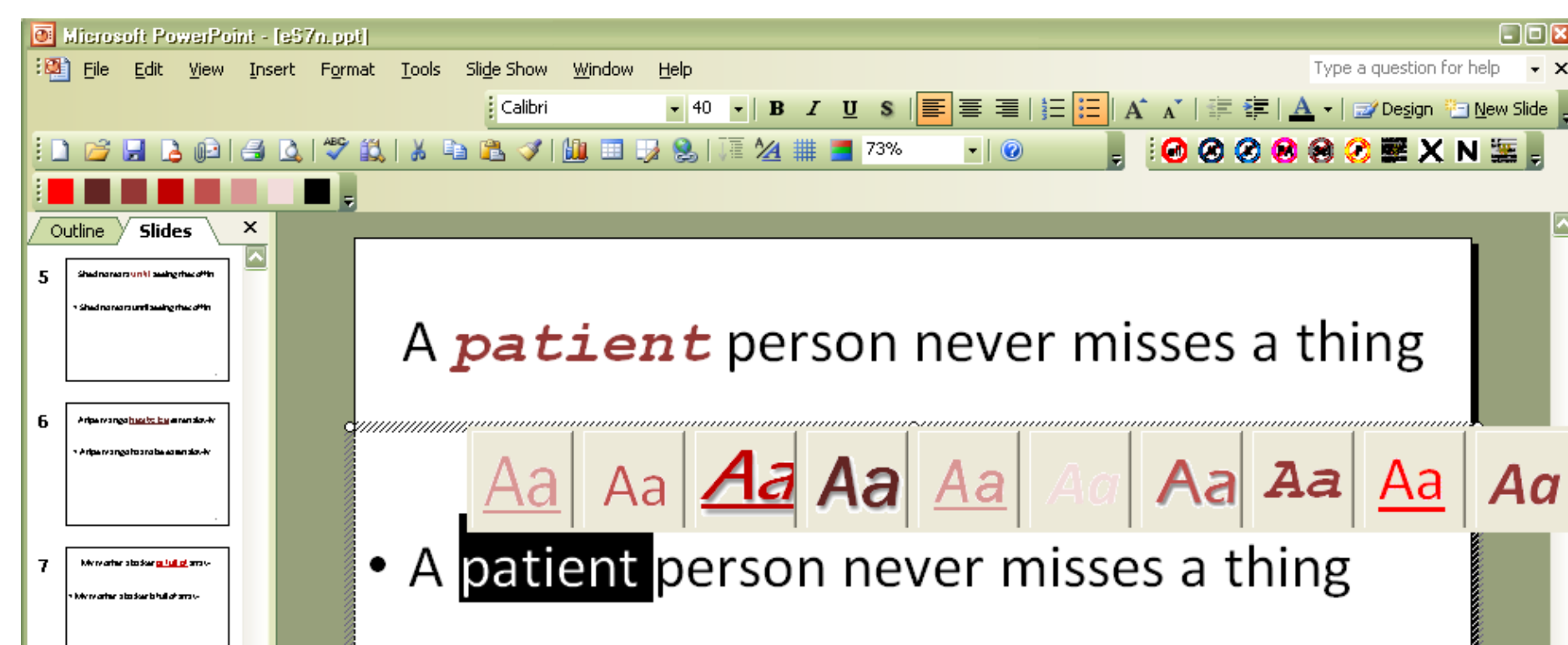
university of toronto

department of computer science

Decision-Theoretic Interface Customization

- Domain: authoring PowerPoint slides
- Goals: repetitive highlighting tasks
- Assistance: toolbar suggestions
- Value and costs of suggestions
- Partially observable Markov decision process (POMDP)

Highlighting in PowerPoint



Value and Costs of Help

- Toolbar, t
- Highlighting goal, g
 - Complexity of goal
- Quality of toolbar, $Q(t|g) = \max Q(i|g)$
 - Quality of icon, $Q(i|g)$
- Neediness, $N(g)$
- Length, $L(t)$
- Suggestion utility, $U(N,L,Q)$

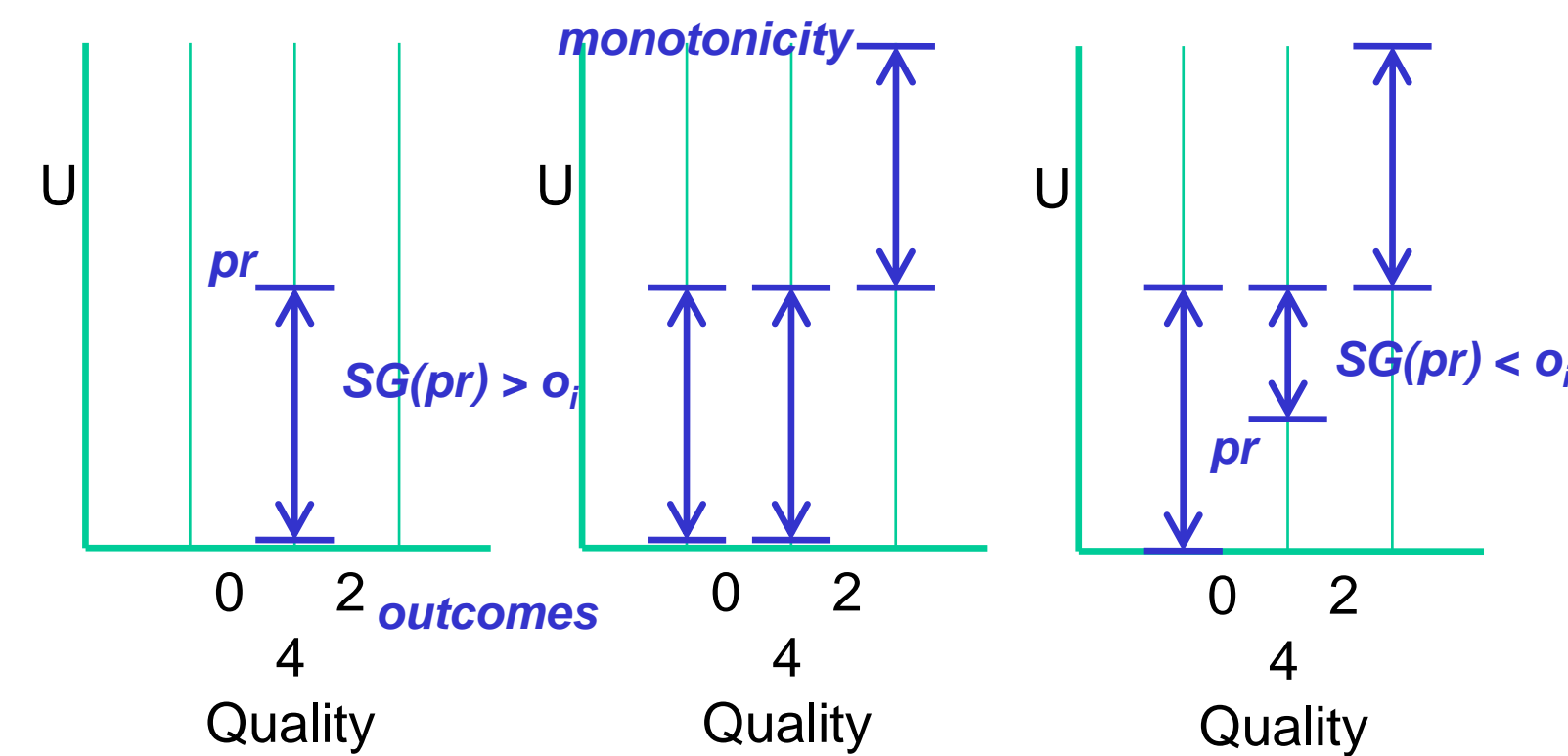
Preference Elicitation

- Outcomes, O
- Utility function, $u: O \rightarrow \text{Reals}$
 - $u(o_i) > u(o_j)$ iff o_i is preferred to o_j
 - $u(o_i) = u(o_j)$ iff indifferent between o_i, o_j
 - o^- is best outcome s.t. $u(o^-) = 1$
 - o_+ is worst outcome s.t. $u(o_+) = 0$
- **Strength** of preferences
- Standard gamble, $SG(pr) = [pr, o^-; 1-pr, o_+]$

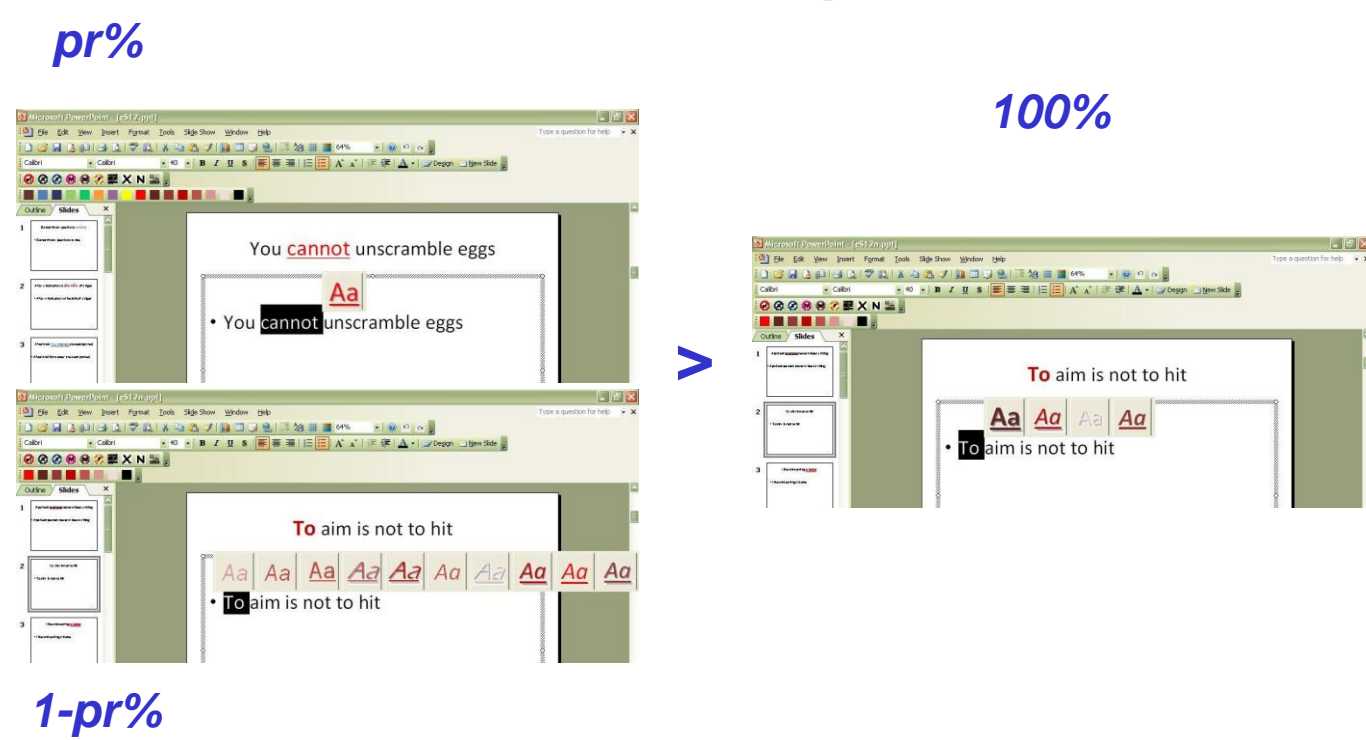
Query Type	Question	Range of Responses
SGQ(pr, o_i)	What is pr s.t. $SG(pr) = o_i$?	$pr \in [0,1]$
Bound(pr, o_i)	Given pr , is $SG(pr) > o_i$?	Yes/No

Bound(pr, o_i) in Theory

- Constraints allow incremental refinement



Bound(pr, o_i) in Practice



- Extremely informative, but...
 - Impossible to answer confidently
 - SG difficult to interpret
 - Difficult to distinguish differences in pr
 - Sequential costs/benefits underestimated

Experiential Queries

- Outcomes describe interface configuration
- Experience via task completions
- Simulate pr with repeated tasks
 - k tasks per alternative
 - Each query involves $2k$ tasks
- Discretize $pr \in [0, .1, .2, \dots, 1.0]$

Experiment Set-Up

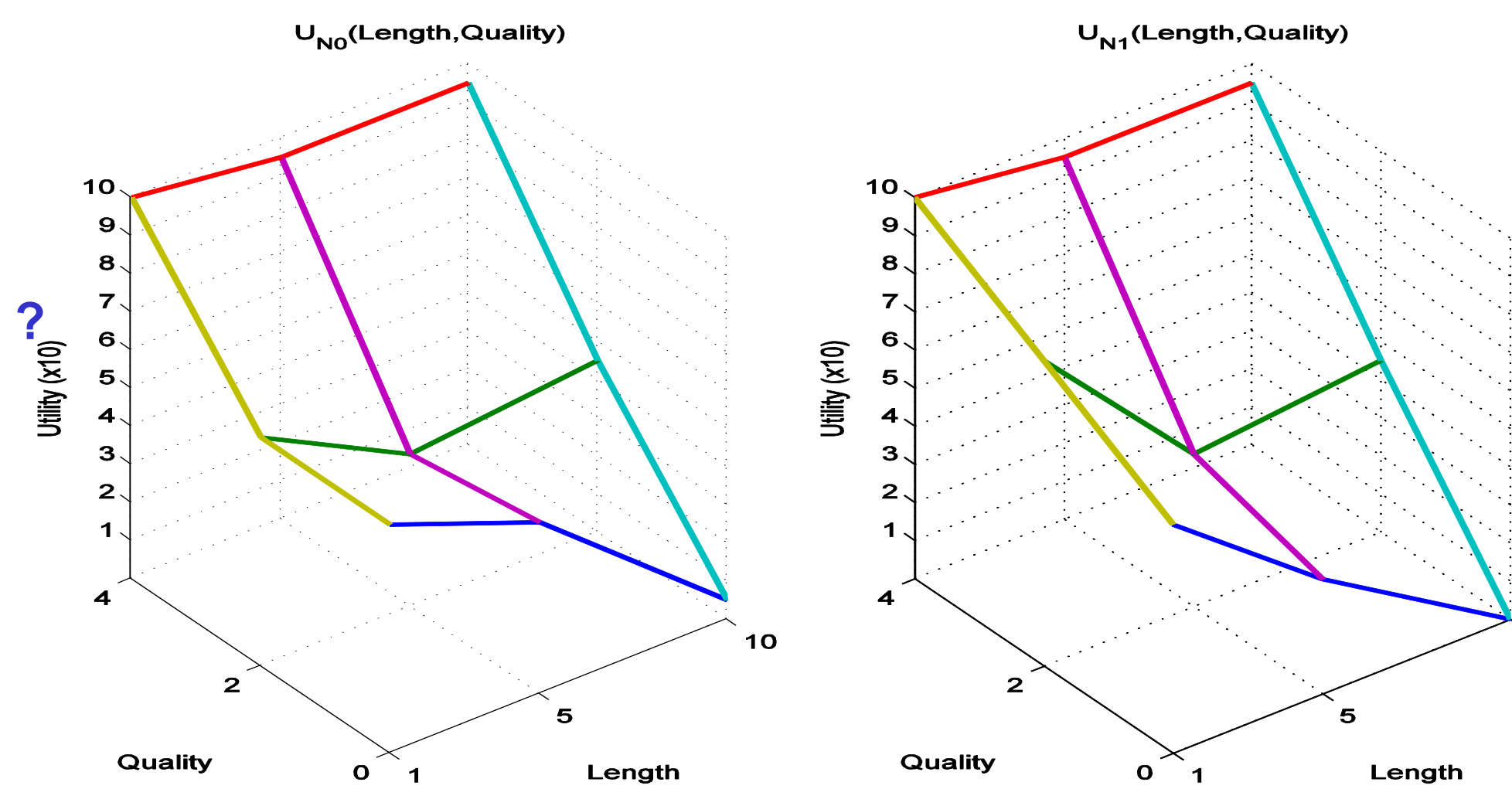
- Controlled highlighting task in PowerPoint
- Simulated N into interface
- Sampled from $U(N,L,Q)$
 - Neediness, $N = 0$ (low), 1 (high)
 - Length, $L = 1, 5, 10$
 - Quality, $Q = 0$ (wrong), 2 (partial), 4 (perfect)
 - $o^- = N0, L1, Q4$
 - $o_+ = N1, L10, Q0$
- Treat options as adaptive vs. static system
- Elicited until small regions ($pr \pm 0.05$)

Experiential vs. Conceptual

- Conceptual: Imagine task completions
- Experiential: Carry out task completions

Structural Results

- Value of non-perfect help (Q2, even Q0)
- Monotonically non-increasing in L
- Monotonically non-decreasing in Q
- Variations in N (user feature)
- Curvature of partial functions
- **Non-additive decomposition**



Methodological Comparison

Experiential	Conceptual
2 hours	30 minutes
Easy to administer	Difficult to explain (outcome mixture, repeated scenario)
Tired easily	Not easily tired
Generally consistent	Often inconsistent

Quantitative Comparison

- H_0 : experiential $\mu \neq$ conceptual μ
- T^2 shows significance ($p < 0.01$)
- Therefore, reject H_0
- Component-wise t test with independent means
- Experience enables users to perceive value of automated help in repeated scenarios

Improving Experiential Elicitation

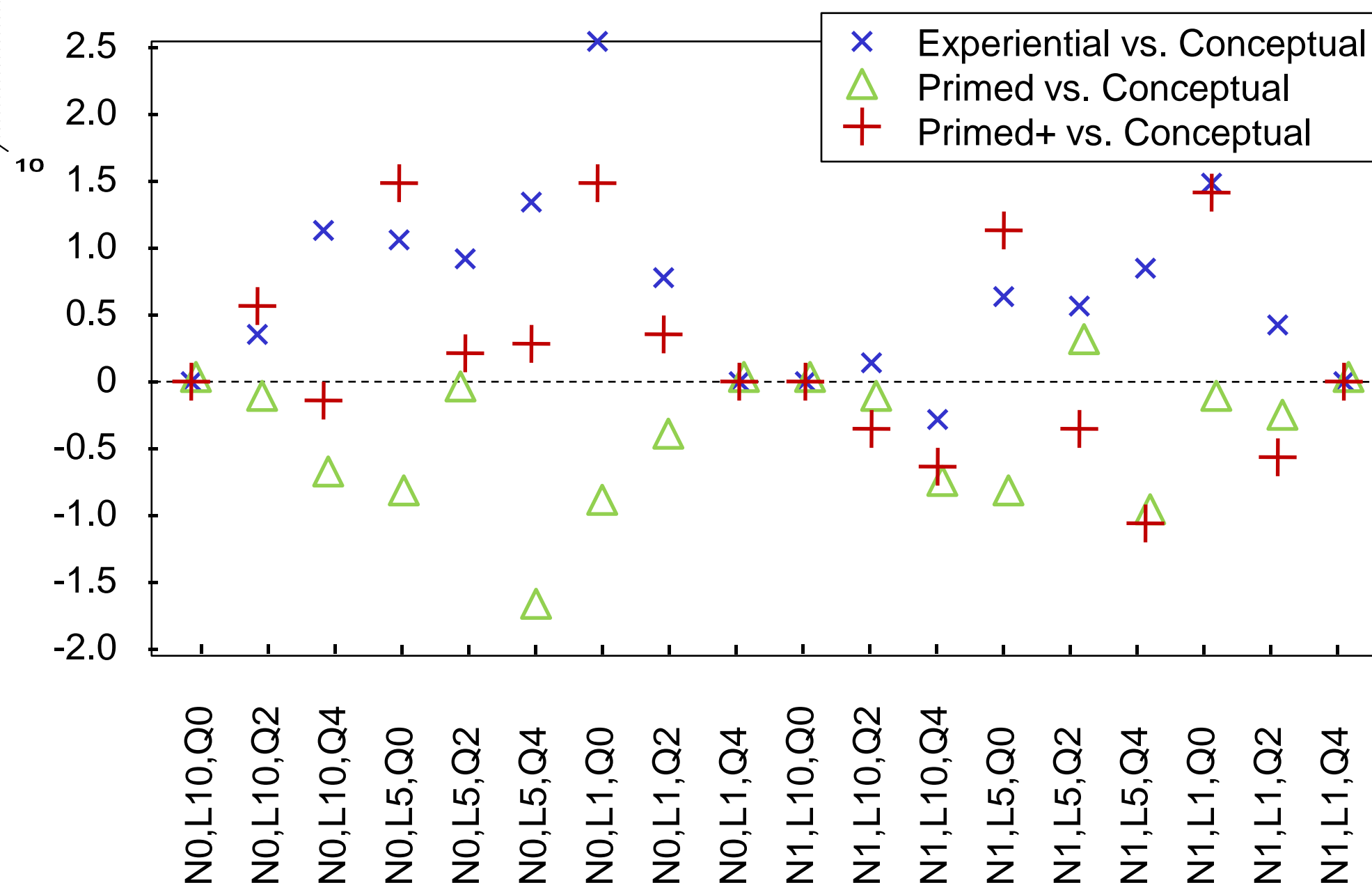
- Reduce time (thus, reduce effort)
- **Primed**: Training session
 - Familiarity with interface and help parameters
- **Primed+**: Training + 5 experiential queries

Methodological Comparison

Primed	Primed+	Conceptual
30 minutes	60 minutes	30 minutes
Easy to administer	Easy to administer	Difficult to explain (outcome mixture, repeated scenario)
Not easily tired	Not easily tired	Not easily tired
Often inconsistent	Experiential queries primed future responses	Often inconsistent

Quantitative Comparison

- H_{10} : primed $\mu \neq$ conceptual μ
- H_{20} : primed+ $\mu \neq$ conceptual μ
- T^2 shows significance ($H1:p < 0.01$; $H2:p < 0.05$) therefore, reject H_{10} and H_{20}
- Component-wise t test with independent means
 - Primed+ approaches Experiential



Contributions

- **Experiential elicitation** for interface customization
 - Uses SGQs with real users
 - Provides repeated experience
- **Primed+** as an efficient approximation

Future Work

- Learn parametric form for $U(N,L,Q)$
 - Quadratic in L and Q ?
- Model general utility function
 - Occlusion, bloat, disruption, interruption, etc.
- Understand experiential "affordance"
 - User expectations in richer domains