DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF TORONTO

CSC 428F/2514F

# HUMAN-COMPUTER INTERACTION

Lecture 17

## UNDERSTANDING USERS AND USER NEEDS; RESEARCH AND EVALUATION METHODOLOGIES 2


17.1 An HCI research and evaluation taxonomy.....2
17.2 HCI research and evaluation strategies.....4
17.3 Thinking aloud .....5
17.4 Data capture, video taping, and protocol analysis.....5
17.5 Cognitive walkthroughs .....6
17.6 Controlled experiments.....7
17.7 Quasi-experiments .....8
17.8 Field experiments .....9
17.9 Tradeoffs among empirical methods.....10
17.10 Research strategies in the development process .....11
17.11 Ethical issues.....12

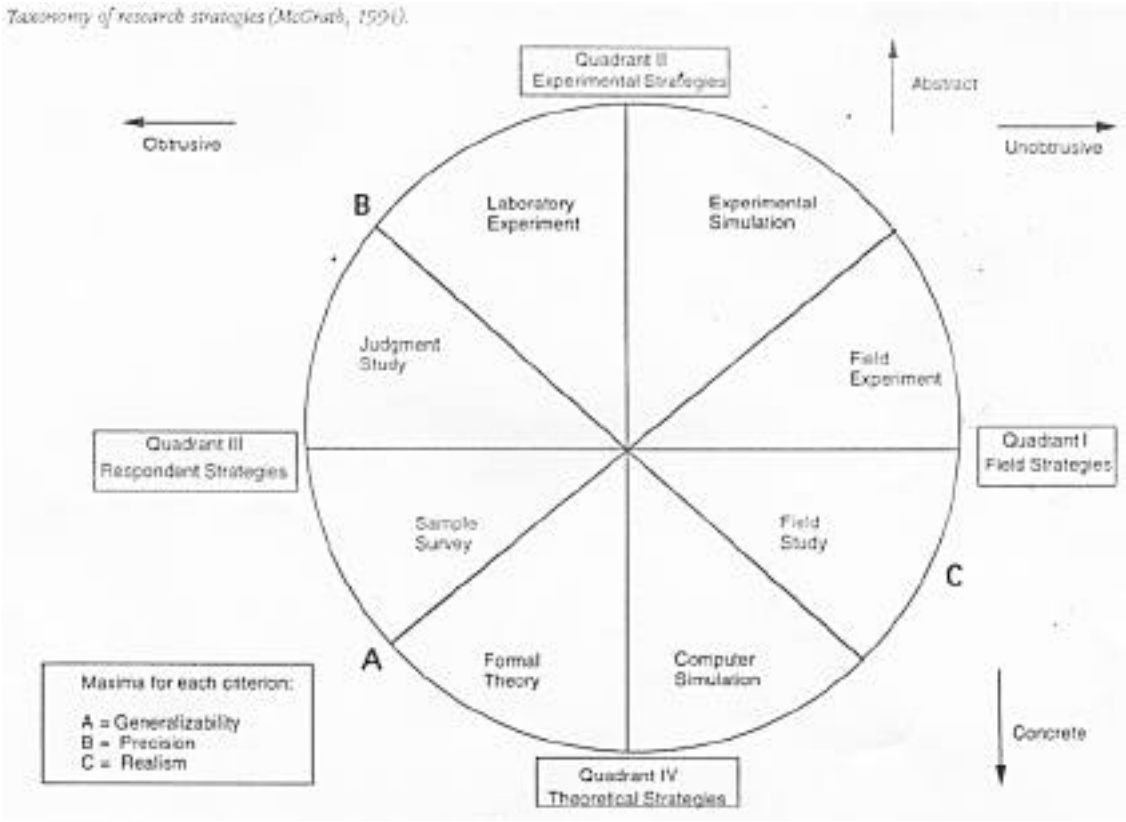


Ronald Baecker
Professor of Computer Science,
Electrical and Computer Engineering, and Management
University of Toronto

## 17.1 An HCI research and evaluation taxonomy

We again present a taxonomy of research strategies due to McGrath (Figs. 4.1, 17.1), then list and organize with respect to McGrath's taxonomy some HCI research and evaluation methods (Figs. 4.2, 17.2); later in this lecture we shall position them in the systems development cycle

*Figure 17.1. Taxonomy of research strategies (BGBG, Fig. 2.4, p. 81)*




Quadrant 1 — Field strategies

- Study systems in real use on real tasks in real work environments
- Field studies — Study systems in situ, disturbing as little as possible
- Field experiments — Observe impact of changing (ideally) one aspect of a work environment

Quadrant 2 — Experimental strategies

Study systems in a laboratory under controlled conditions

Laboratory experiments — Carry out controlled experiments studying impacts of (ideally) one interface parameter

Experimental simulations — Create in laboratory for experimental purposes a real system that is used by real users on (usually) artificially simplified tasks

Quadrant 3 — Respondent strategies

Ask informants to tell us something about themselves and/or their work or about an interface

Judgment studies — Ask respondents about an interface

Sample surveys — Ask respondents about themselves and/or their work

Quadrant 4 — Theoretical strategies

Ask a theory to tell us something about people's work or about an interface

Formal theory — Use a qualitative theory or some equations

Computer simulation — Use and run a computer model

## 17.2 HCI research and evaluation strategies

*Fig. 17.2. HCI research & evaluation methods (based on BGBG, Fig. 2.5, p. 81)*

| |
|---|
| **Field strategies** |
| **(Settings under conditions as natural as possible)** |
| Field studies |
| Ethnography and interaction analysis (Lect. 17) |
| Contextual inquiry (Lect. 6) |
| Field experiments (Lect. 17) |
| Beta testing of products (CSC 454) |
| Studies of technological change (CSC 300) |
| |
| **Experimental strategies** |
| **(Settings concocted for research purposes)** |
| Experimental simulations |
| Usability testing (Tut. 1, Lect. 4, Lect. 17, CSC318) |
| Usability engineering |
| Laboratory experiments |
| Controlled experiments (N&L, Ch. 10) |
| |
| **Respondent strategies** |
| **(Setting is muted or made moot)** |
| Judgment studies |
| Demonstrations (Lect. 4) |
| Usability inspection methods (e.g., heuristic evaluation) (Tut. 2, Lect. 4) |
| Cognitive walkthroughs (Lect. 17) |
| Sample surveys |
| Customer surveys, questionnaires, and interviews (Lect.'s 4-6) |
| |
| **Theoretical strategies** |
| **(No observation of behavior required)** |
| Formal theory |
| Design theory, e.g., Norman's 7 Stages (Lect. 10) |
| Behavioral theory, e.g., color vision (Lect.'s 7-9) |
| Computer simulations |
| Human information processing theory (Lect.'s 19-21) |

We shall cover advanced aspects of user testing (thinking aloud), ethnography and video and interaction analysis, cognitive walkthroughs (a judgment study), experiments and quasi-experiments, and put this all in context

## 17.3 Thinking aloud

Attempt to elicit thought processes of user testing subjects (4.6)

Subjects talking while they are doing
- Problems they are having
- Solutions they are considering
- Why they are having trouble

An interesting variation is pairs of subjects conversing (Co-Discovery Learning, Kennedy paper in BGBG, Ch. 2)

## 17.4 Data capture, video taping, and protocol analysis

Keystroke logging
- Sometimes known as “dribble files”
- Record precise user behaviour
- Record times to carry out actions
- Record user errors

Observation and notetaking
- Critical incidents
- User problems

Audio and video recordings
- Can't record all behaviour in real-time
- Preserve behaviour for review
- Non-verbal behaviour
- Behaviour in context

Data extraction and analysis
- Enormous amounts of “data” in the video tape
- How to convert to information?

Review of video and audio tapes
Construction of *protocols*
Classifying events and counting events
Computer systems for video annotation and analysis

Ethnographic methods, interaction analysis (Suchman & Trygg)
Ethnography...“the careful study of activities and relations between them in a complex *social setting*.”
*Situations of use* in actual work settings
Interaction analysis...“to uncover the regularity and efficacy of peoples' relations with each other and their use of the resources that their environment affords”

## 17.5 Cognitive walkthroughs

A judgment study (using knowledgeable respondents) based on a theory of exploratory learning by inexperienced users without any prior training

Users have goals
Users choose actions to accomplish goals
Users assess progress with respect to goals
Users modify goals or generate new goals based on system responses as they work

So developers and cognitive walkthrough analysts step through specific goals that users make have in working with a system, asking the following questions:

- Will the users try to achieve the right effect?
- Will the user notice that the correct action is available?
- Will the user associate the correct action with the effect trying to be achieved?
- If the correct action is performed, will the user see that progress is being made toward solution of the task?

## 17.6 Controlled experiments

Method
- Manipulate independent variables,system characteristics
- Control for other variables
- Measure dependent variables, user behaviour

Roles
- Understanding causes of user behaviour
- Understanding factors influencing interface quality

Advantages
- Strong statements about causality
- Many experimental designs suitable for varying situations

Disadvantages
- Requires time, planning, may be expensive
- Complex designs (more than 3 or 4 independent variables) are often difficult to interpret
- May legitimize trivial research, and generate results of weak generalization (*external validity*)

Example of a real experiment — Perlman study on menu format (words, numbers) and order (sorted, unsorted), and selection mechanisms (letter, number, compatible, incompatible) (B&B, pp. 451-455)

Example of real experiments — Egan et al. study of searching with print text and electronic text (SuperBook), as a function of whether or not the search term appears in the document heading structure and/or the document text (BGBG, pp. 843-848)

## 17.7 Quasi-experiments

Experiments that lack statistical significance (i.e., not enough subjects or individual variability too great for stat. signific.) or that lack controls, lacks *internal validity*

Typical method
- Measure change of subjects' behaviour as system changes

For example, study system as it evolves over time,
- measure performance of group of subjects *both* before and after an experimental treatment (like modification of user interface, icons, input devices, etc.)

But this is not a controlled experiment
- Same people used: learning is a *confound*
- Subjects know system has been refined:
  - expectation is a *confound*
- Multiple factors changed from version *n* to *n* +1:
  - these factors are *confounds*

Roles
- Understanding effects of system change on user behaviour
- Evaluation at far lower cost than controlled experiments

Examples of quasi-experiment:
- Bewley et al. tests on Star “graphics” (line drawing) functionality (B&B, pp. 662-667)
- Baecker, Small, Mander tests on “animated icons” (BGBG, pp. 444-449) — Confound is learning from test of static icons to test of animated icons
- Perkins et al. iterative design of Freestyle user interface plus tutorial (BGBG, pp. 881-885) — Confound is changing the interface plus the tutorial

## 17.8 Field experiments

Methods
- Typical usage, or critical incidents
- Keystroke recording
- Thinking aloud protocols
- Videotape protocols
- Protocol analysis
- Interviews

Roles
- Testing and enhancing productivity
- Testing and improving speed
- Testing and reducing errors

Advantages
- Observations done is real work settings

Disadvantages
- Usage may be very patterned, not cover full range of system features
- Studies expensive to run

Example: Roberts and Moran studies of text editor usage (B&B, pp. 250-268), testing 4 users using 9 text editors to do 53 benchmark tasks on which they measured:
- The *time* to perform basic editing tasks by experts
- The *error* cost for experts
- The *learning* of basic editing tasks by novices
- The *functionality* over a wide range of editing tasks

## 17.9 Tradeoffs among empirical methods

Internal validity
- Degree of confidence that we have found "the" explanation for our results, that is, we do not know of other confounding explanations —
- We achieve this by increasing precision and direct control over experiment

External validity (generalizability)
- Degree to which our research applies to other phenomena than just the "experiment" —
- Achieving this by increasing range, or scope, of phenomena studied

Tradeoff between internal validity (soundness) and external validity (generalizability, relevance, realism)
- Controlled experiments for internal validity
- Breadth of naturalistic observation for external validity

"Credible empirical knowledge requires consistency or convergence of evidence across studies based on different methods." (McGrath, in BGBG, p. 155)

Different strategies and methods have different advantages and disadvantages — cannot simultaneously maximize:

*Generalizability* of evidence over *populations* of actors (A)

*Precision* of measurement of the *behaviours* (B)

*Realism* of the situation or *context* (C)

McGrath research strategies diagram (Fig 17.1) shows tradeoffs

## 17.10 Research strategies in the development process

*Figure 17.3. Possible uses of evaluation methods in a sample development process (BGBG, Fig. 2.9, p. 88)*

| |
|---|
| *Information collection* |
| Interviews and questionnaires |
| Contextual inquiry |
| Interaction analysis |
| |
| *Concept design* |
| Interviews |
| Heuristic evaluation |
| Usability testing |
| Controlled experiments |
| |
| *Functionality and interface design* |
| Heuristic evaluation |
| Usability testing |
| Theory-based evaluations |
| Human information processing simulations |
| |
| *Prototype implementation* |
| Usability testing |
| Heuristic evaluation |
| |
| *Deliverable system implementation* |
| Usability testing |
| Quasi-experiments |
| |
| *System enhancement and evolution* |
| Interaction analysis |
| Interviews and questionnaires |
| Field experiments |

## 17.11 Ethical issues

Basic principles
- Do no harm
- Voluntary participation
- Informed consent
- Right to privacy

Difficult issue — Uses of video data