DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF TORONTO

CSC 318S

THE DESIGN OF INTERACTIVE COMPUTATIONAL MEDIA

Lecture 15 — 11 March 1998

SYSTEM AND INTERFACE EVALUATION 2

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

15.1 Demonstrations

Method
- Demonstrate system to:
  - Any warm body you can capture
  - The “powers-that-be”
  - Potential customers
  - Potential businesss partners
- Take detailed notes

Role
- Elicit reactions to user's model, functionality, interface

Advantages
- Get feedback at an early stage of prototype or system construction
- You're going to have to give demos anyway — why not learn from them?

Disadvantages
- System still rough, which introduces noise into process

Examples
- Pick your favorite project!

These are respondent strategies, specifically, judgment studies, usually less formal than usability inspection methods, about to be discussed

15.2 Usability inspection methods

Methods
- Heuristic evaluation
  - Judgments by a panel of evaluators (e.g, 3 to 5) of the degree to which an interface satisfies a set of usability guidelines, followed by discussion and analysis
- Cognitive walkthroughs (CSC428)

Roles
- Studies that don't involve users (in contrast to demos, usability testing, etc.)
- Elicit *expert opinion* about user's model, functionality, look and feel of the interface, etc.

Advantages
- Structured method of using accumulated wisdom of experts

Disadvantages
- Doesn't take advantage of real insights from real users

Example — Heuristic evaluation with 10 usability guidelines (Nielsen, BGBG, Fig. 2.7, p. 83)

- Visibility of system status
- Match between system and the real world
- User control and freedom
- Consistency and standards
- Error prevention
- Recognition rather than recall
- Flexibility and efficiency of ue
- Aesthetic and minimalist design
- Help users recognize, diagnose, and recover from errors
- Help and documentation

15.3 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)

15.4 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

15.5 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:

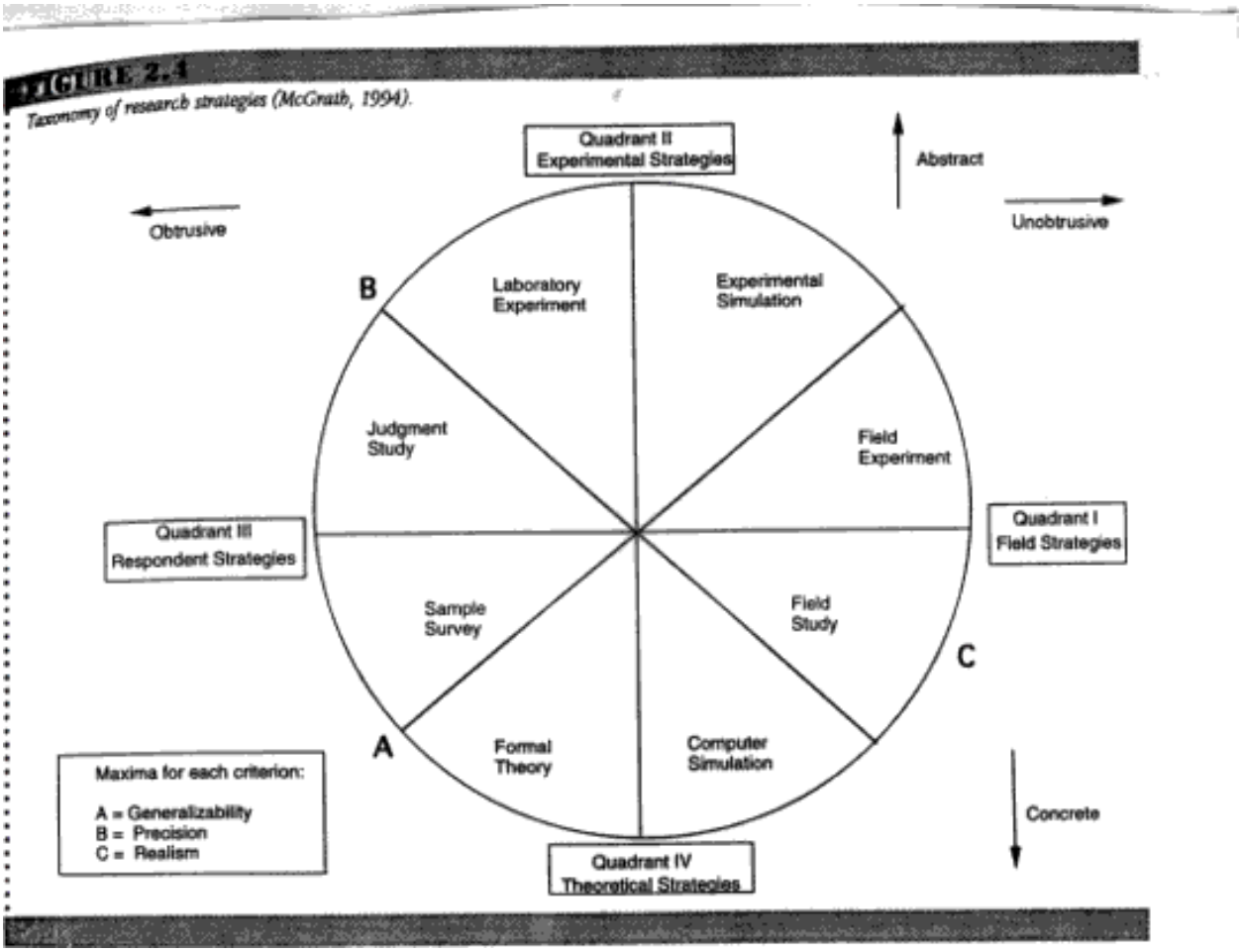
Generalizability of evidence over *populations* of actors (A)

Precision of measurement of the *behaviours* (B)

Realism of the situation or *context* (C)

McGrath research strategies diagram (Fig 15.1) shows tradeoffs

Figure 15.1.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

15.6 Research strategies in the development process

Figure 15.2. 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

15.7 Ethical issues

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

Use of research protocols and consent forms

Difficult issue — Uses of video data