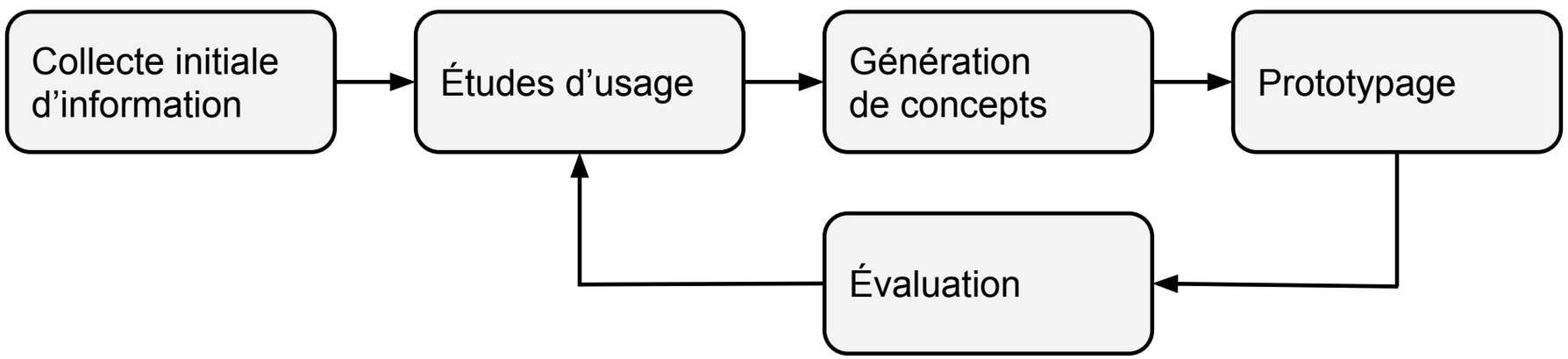
INF03 Expérience Utilisateur

6. Évaluation et tests

Aurélien Tabard

Rappels des dernières séances – 1

- Historique de l'UX : Design et Informatique.
- Définition de l'UX :
  - *All the aspects of how people use an interactive product: the way it feels in their hands, how well they understand how it works, how they feel about it while they're using it, how well it serves their purposes, and how well it fits into the entire context in which they are using it.*"
    Alben, L. 1996, Quality of Experience. Interactions, 3 (3), 11-15
- Le cycle de conception UX



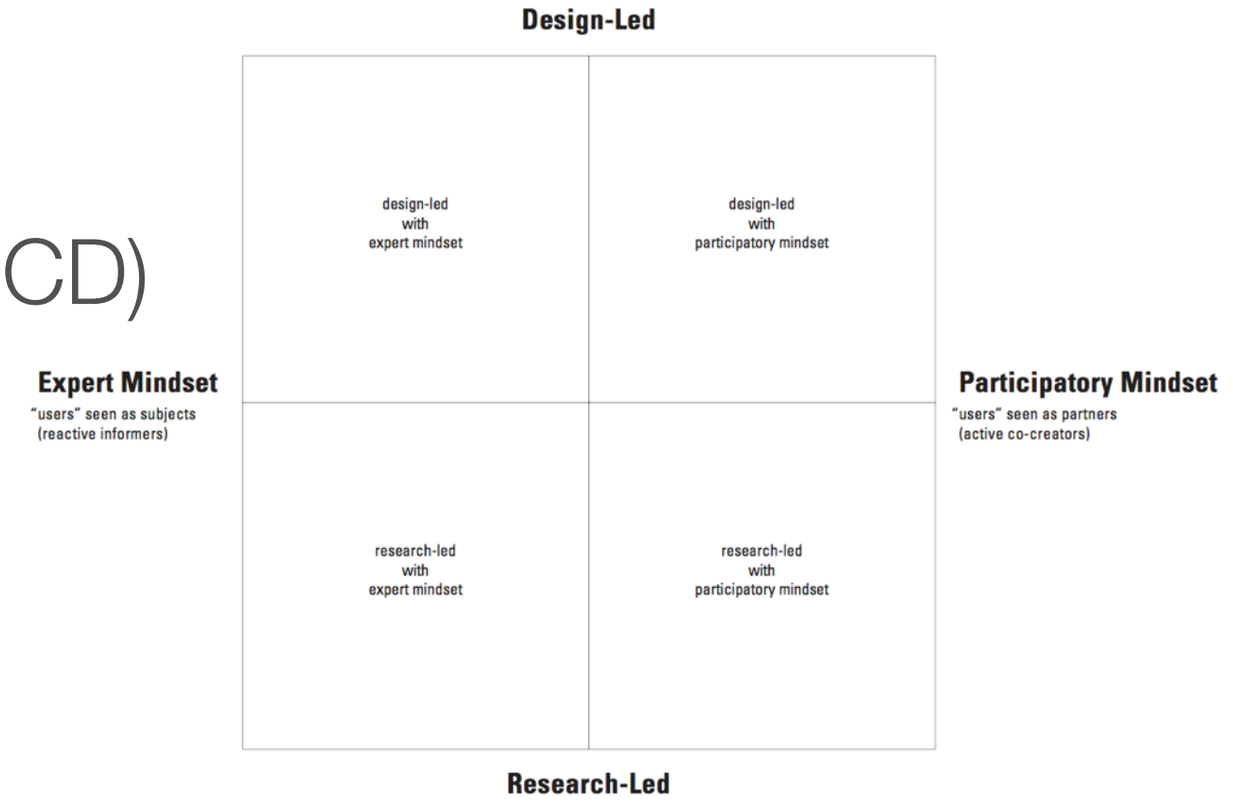
Rappels des dernières séances – 2

- Recherche utilisateur
  Pourquoi, origines, approches
- Comment faire des entretiens
- Projet : Embarquement Cozycloud

Rappels des dernières séances – 3

Les approches UX :

1. Design centré utilisateur (UCD)
2. Design systémique
3. “Genius” design



Point sur les entretiens

Rappels des dernières séances – 4

1. Méthodes d'analyse d'entretiens et d'observations
   - La théorie ancrée
   - Les diagrammes d'affinité
   - L'analyse de tâche
2. Génération d'idées / brainstorming

Rappels des dernières séances – 4

Pourquoi prototyper

Situer la conception dans un contexte

Prototyper

- Prototypes papier,
- Prototypes vidéo,
- Wireframes,
- Prototypes animés (Axure / power-point)

Plan des prochaines séances

Évaluation et tests :

- Introduction
- Approches d’évaluation
- Méthodes analytiques
- Méthodes empiriques
- Évaluation 2.0 : passer à l’échelle
- Design expérimental

Plan de la séance

Évaluation et tests :

- **Introduction**
- **Approches d’évaluation**
- **Méthodes analytiques**
- Méthodes empiriques
- Évaluation 2.0 : passer à l’échelle
- Design expérimental

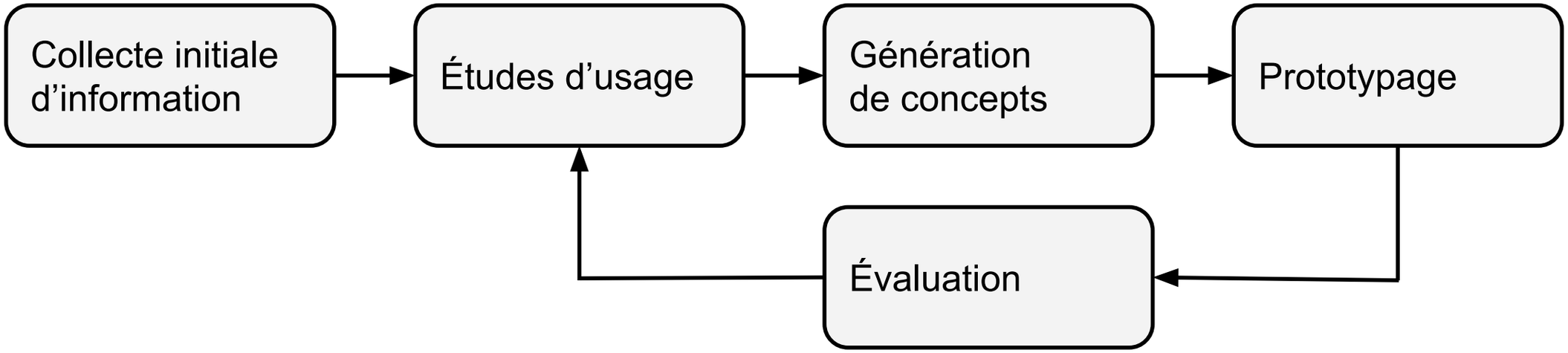
Évaluation et tests

- **Introduction**
- Approches d’évaluation
- Méthodes analytiques
- Méthodes empiriques
- Évaluation 2.0 : passer à l’échelle
- Design expérimental

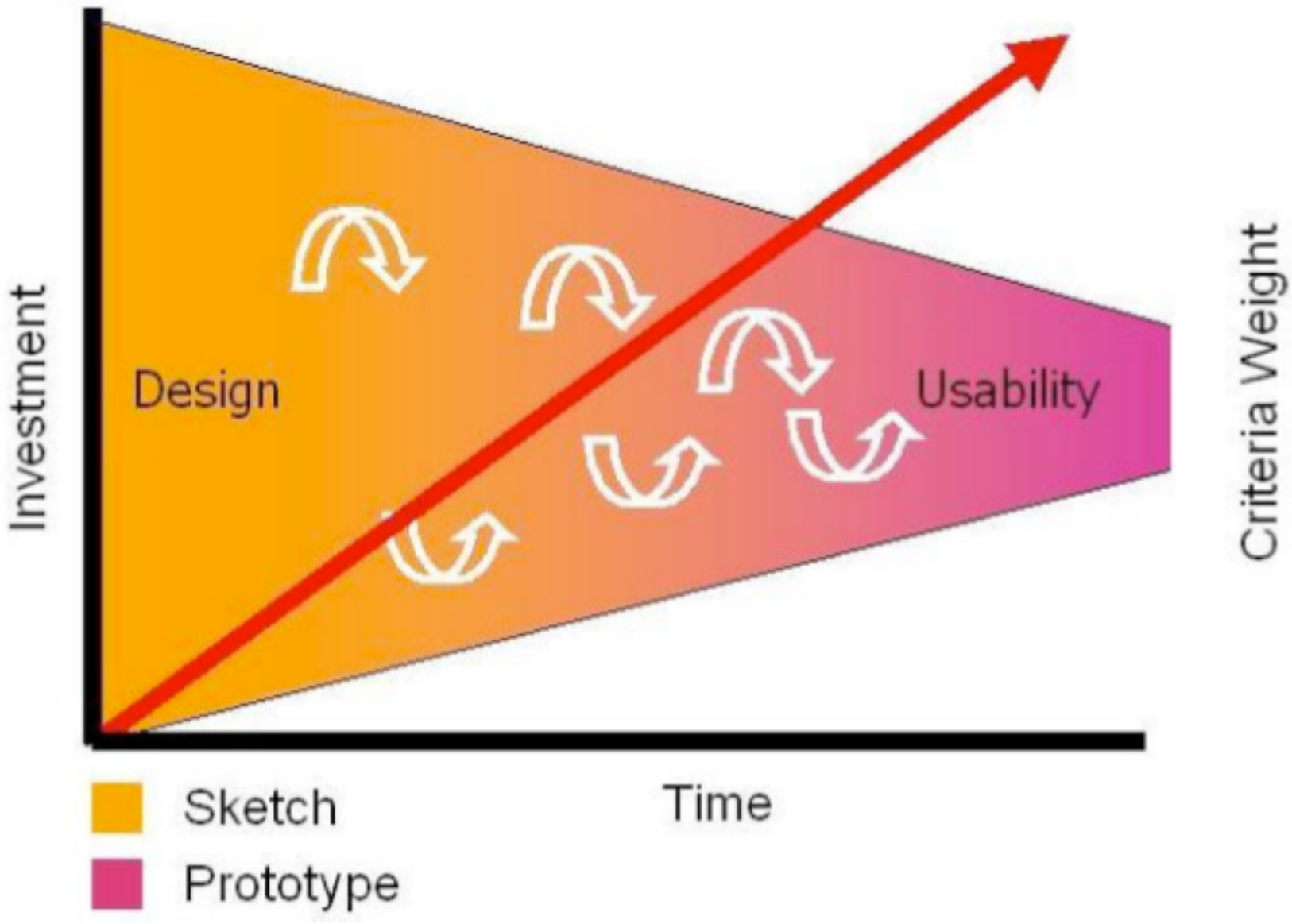
Le rôle de l'évaluation dans le processus UX

- Une partie du cycle itératif : *design-build-evaluate*
- Une comparaison entre ce qui est "construit" et ce qui était prévu
- Un endroit pour réfléchir sur cette différence et la prochaine phase de conception

Quand évaluer



Le processus de conception



Être agile

Rater rapidement pour réussir plus tôt :

- Itérations sur des prototypes basse fidélité
- Design en parallèle : construire et tester plusieurs prototypes
- Explorer des alternatives

Augmenter progressivement la fidélité

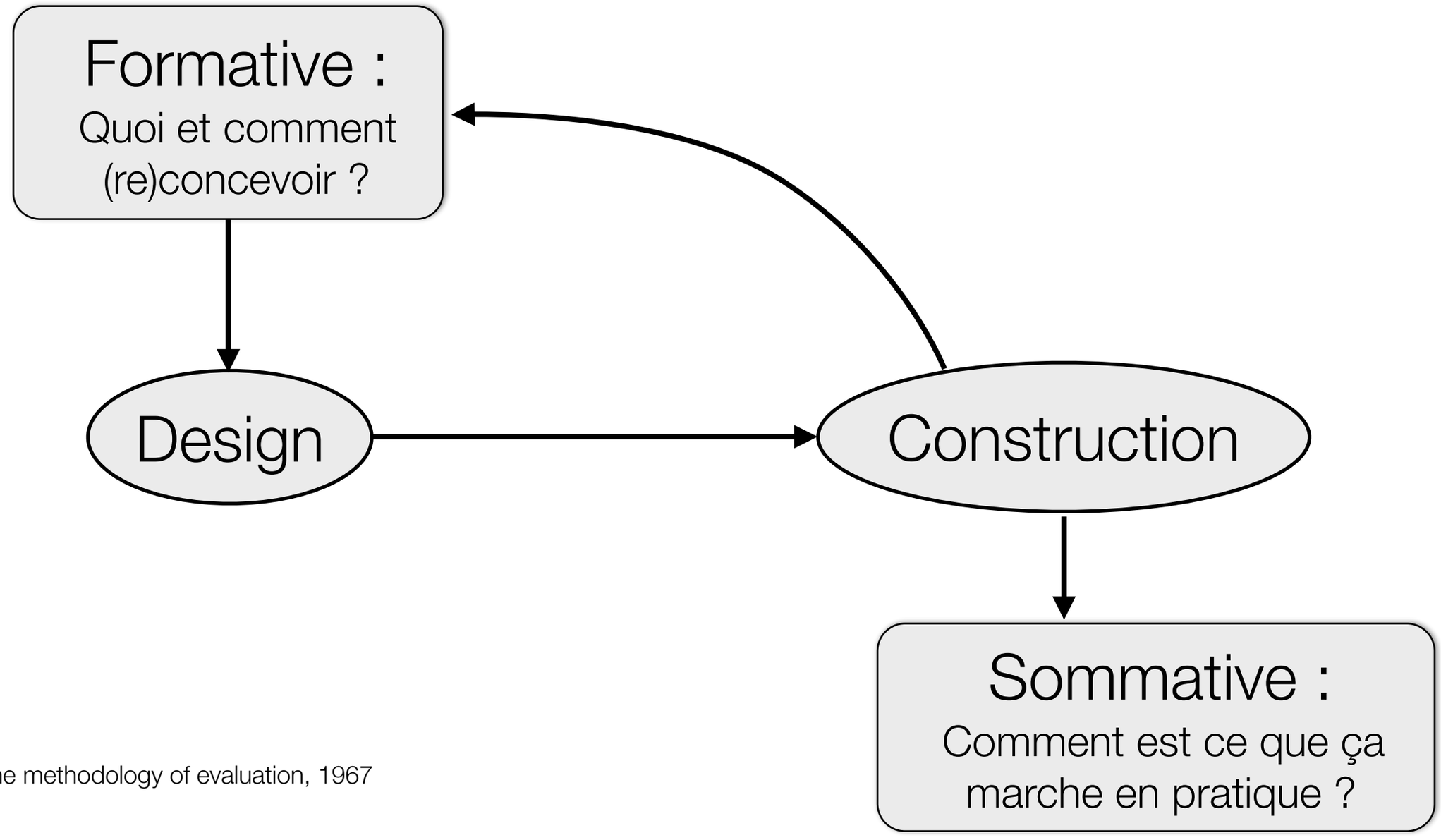
Affronter la réalité, concevoir pour des cas d'utilisation pas des spécifications

Évaluation et tests

- Introduction
- **Approches d'évaluation**
  - Méthodes analytiques
  - Méthodes empiriques
  - Évaluation 2.0 : passer à l'échelle
  - Design expérimental

Qu’est ce qu’on évalue ?

Évaluation formative ou sommative ?



M. Scriven: The methodology of evaluation, 1967

Évaluation Analytique vs. Empirique

“If you want to evaluate a tool, say an axe, you might study the design of the bit, the weight distribution, the steel alloy used, the grade of hickory in the handle, etc., or you may just study the kind and speed of the cuts it makes in the hands of a good axeman.”

[Scriven, 1967]

Des méthodes complémentaires

L'évaluation empirique permet de comprendre le contexte des propriétés de l'objet

- Si la hache ne coupe pas bien, que faut-il changer ?

Améliorer

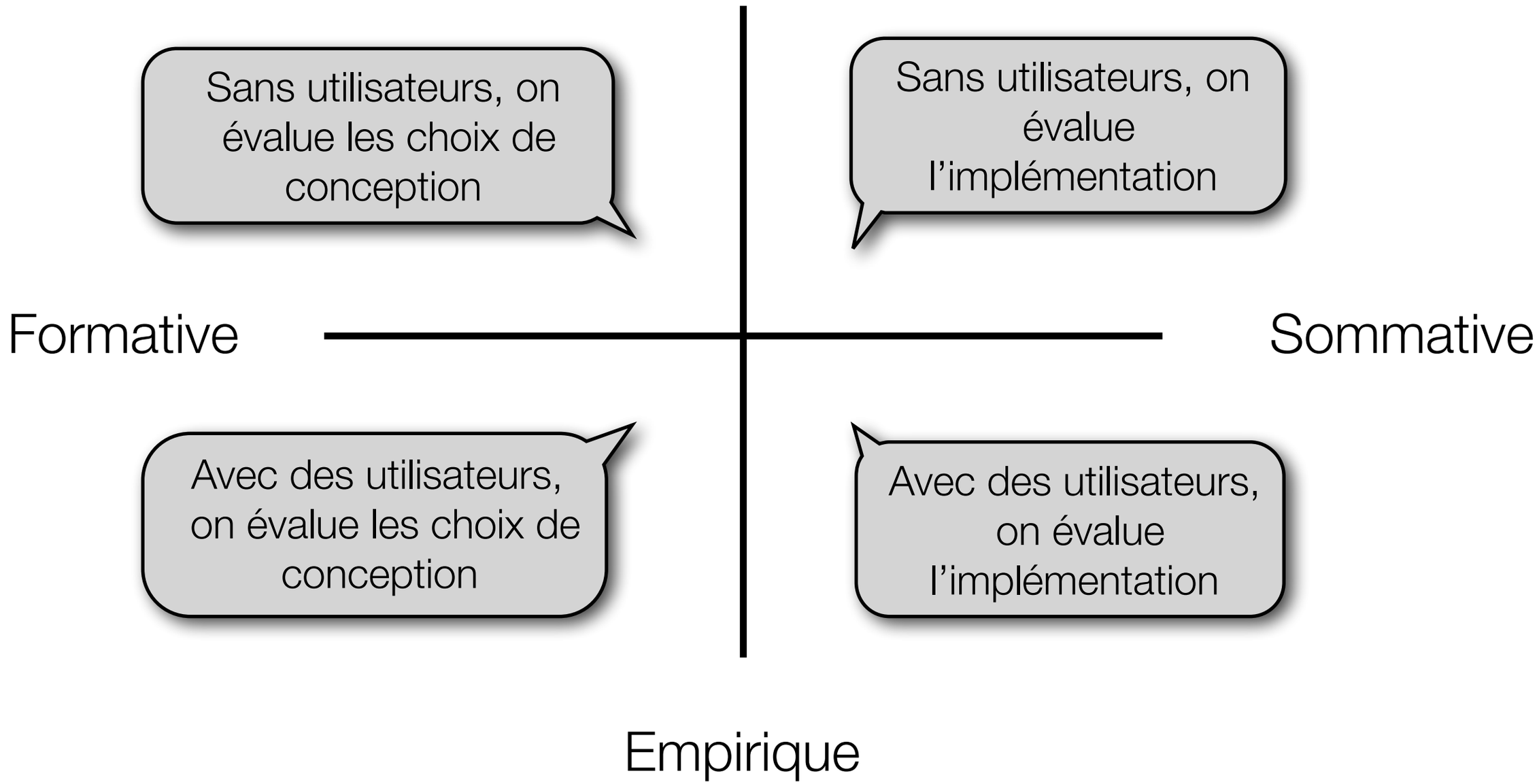
L'évaluation analytique opriétés importantes

- Pourquoi la hache a un manche spécial ?

Dans les deux cas :

- Productions de faits qui doivent être interprétés

Des approches orthogonales



Une évaluation sans critère ne sert à rien !

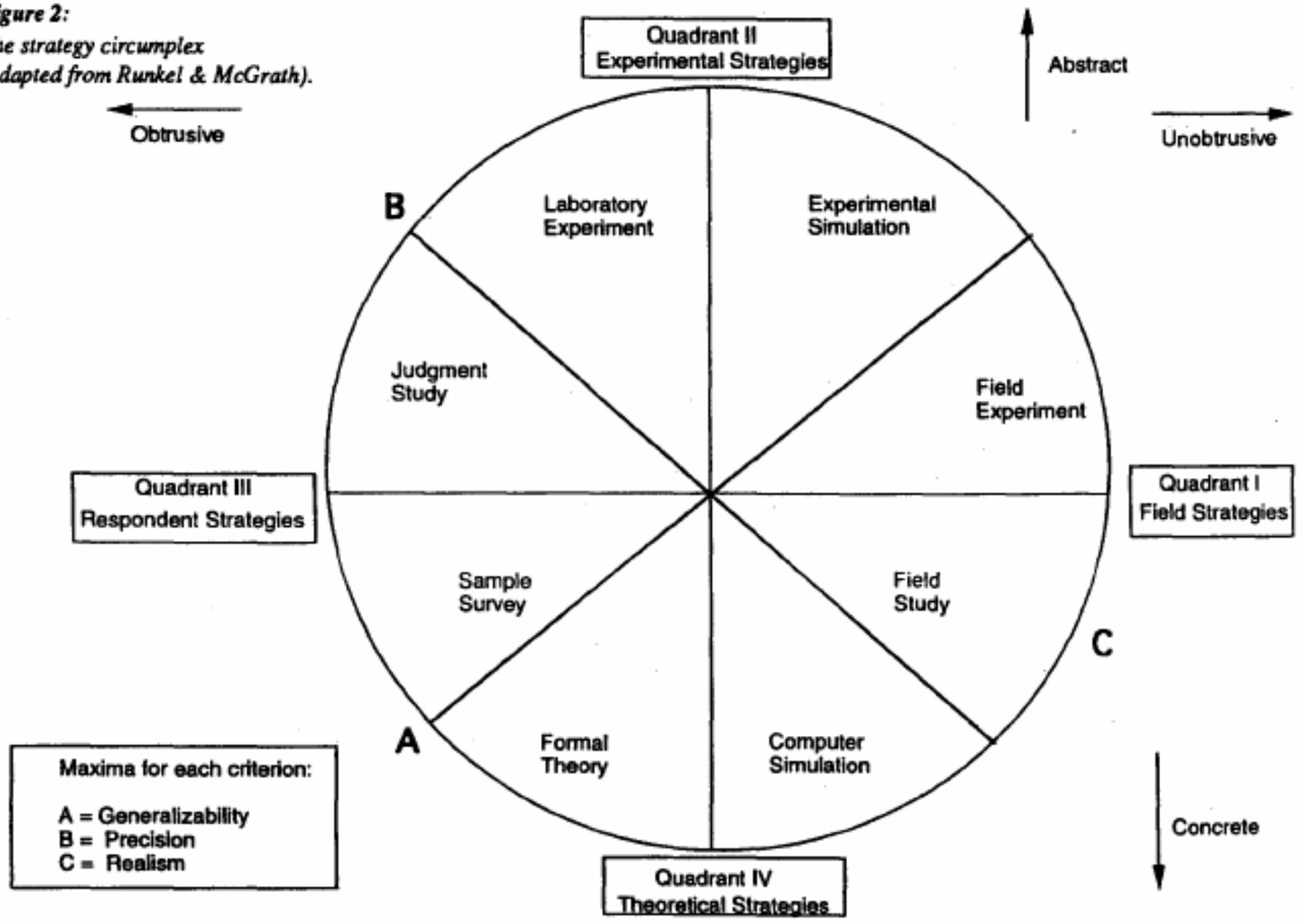
If faut définir ce qu'on veut savoir avant d'évaluer !

Critères possible (parmi bien d'autres) :

- Test informel d'une idée contre une autre
- Analyse statistique de la performance moyenne
- Acceptation par un groupe d'utilisateur réaliste
- Vérification de critères heuristiques / ergonomiques
- Mesure de métriques d'utilisabilité lié au design

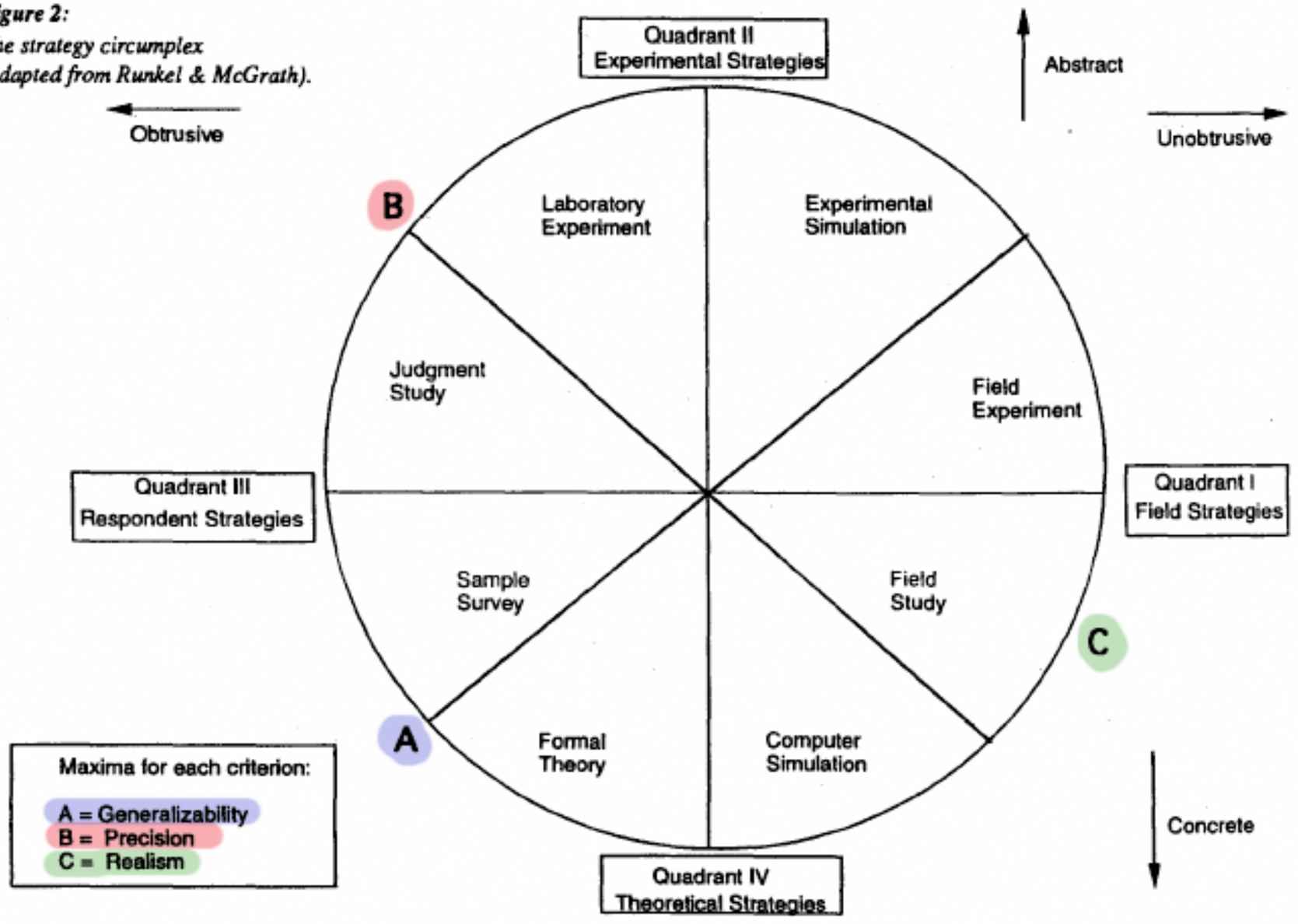
Taxonomy of Methods [McGrath et al. 1994]

Figure 2:
The strategy circumplex (adapted from Runkel & McGrath).



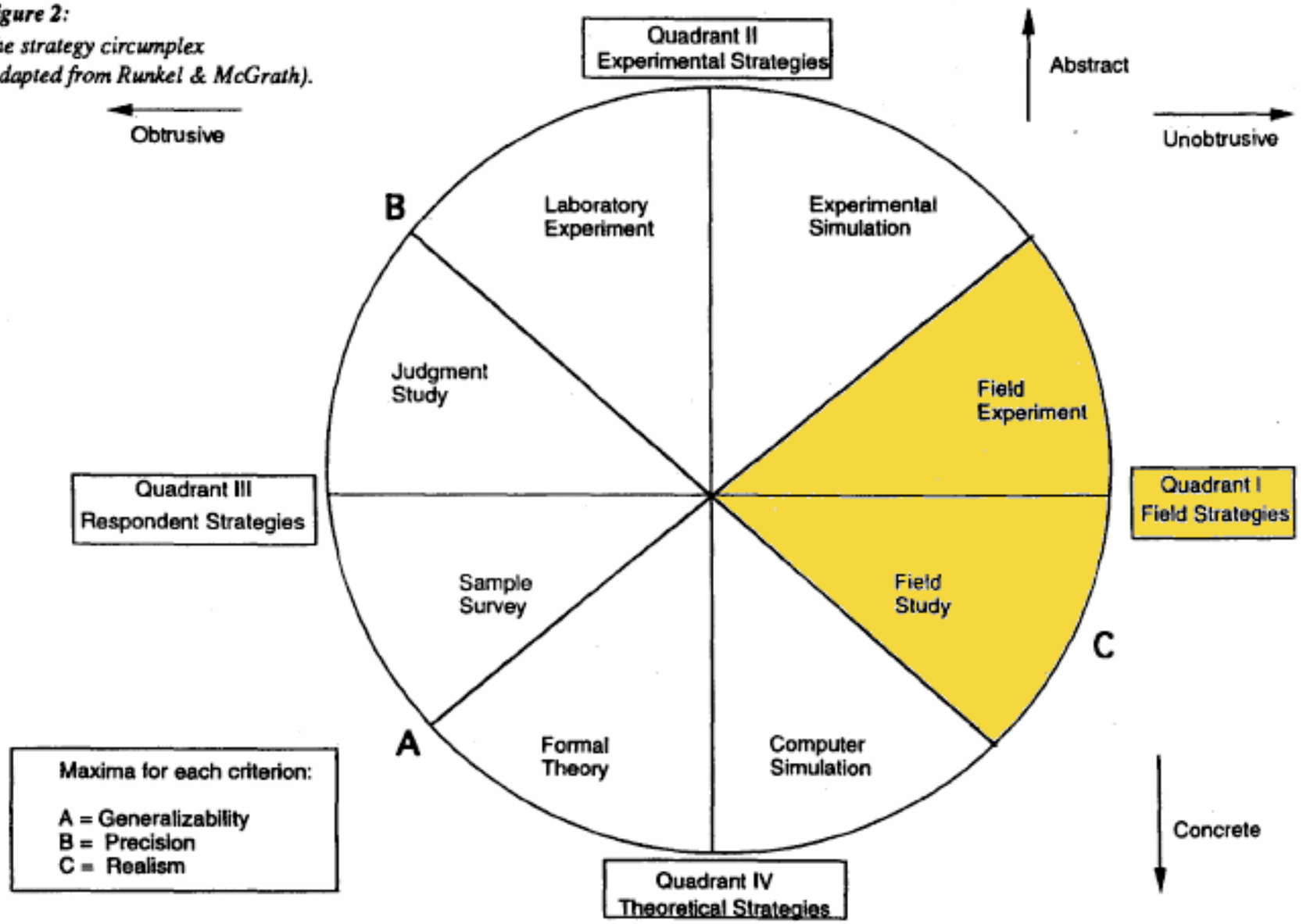
Taxonomy of Methods [McGrath et al. 1994]

Figure 2:
The strategy circumplex (adapted from Runkel & McGrath).



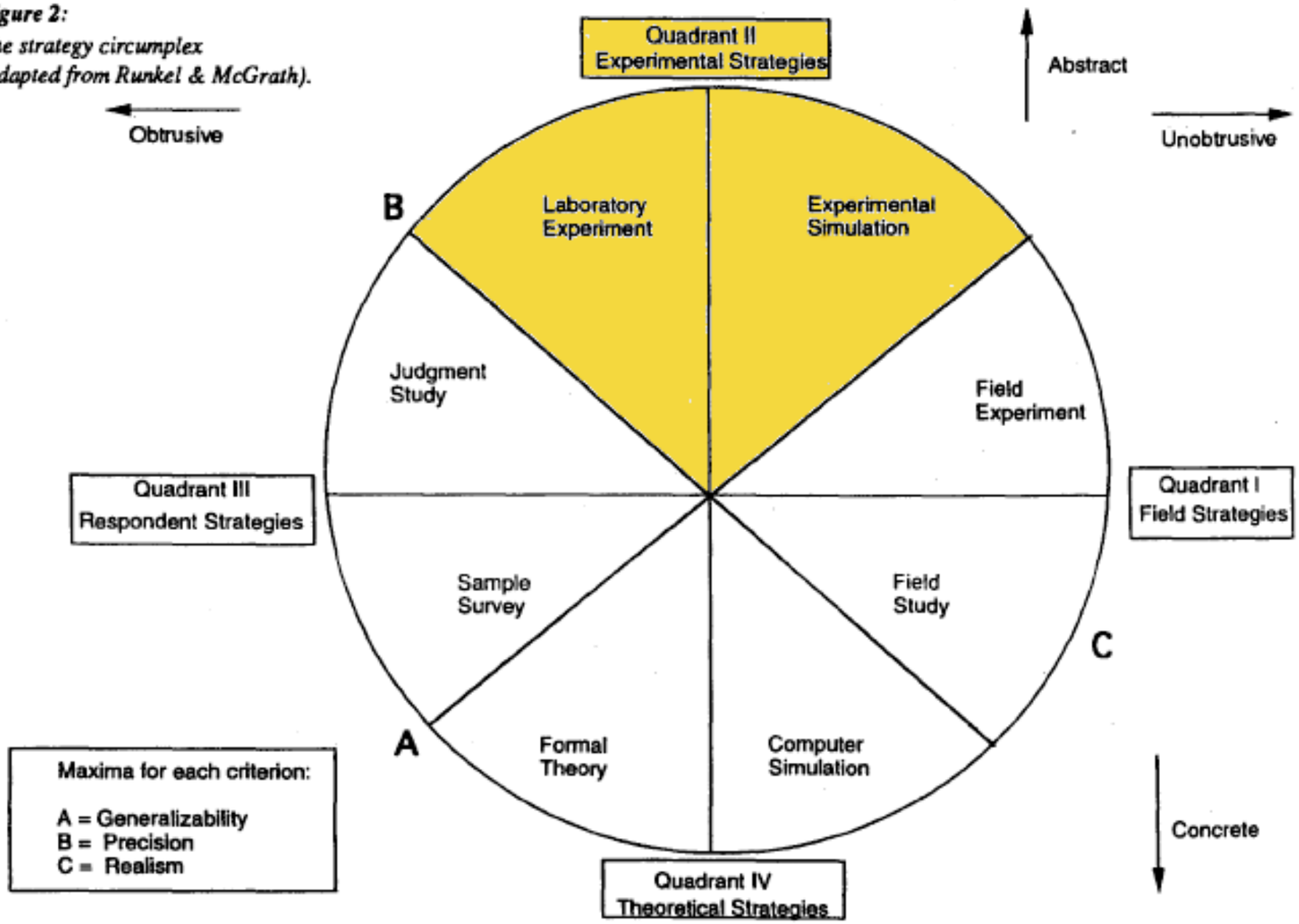
Taxonomy of Methods [McGrath et al. 1994]

Figure 2:
The strategy circumplex (adapted from Runkel & McGrath).



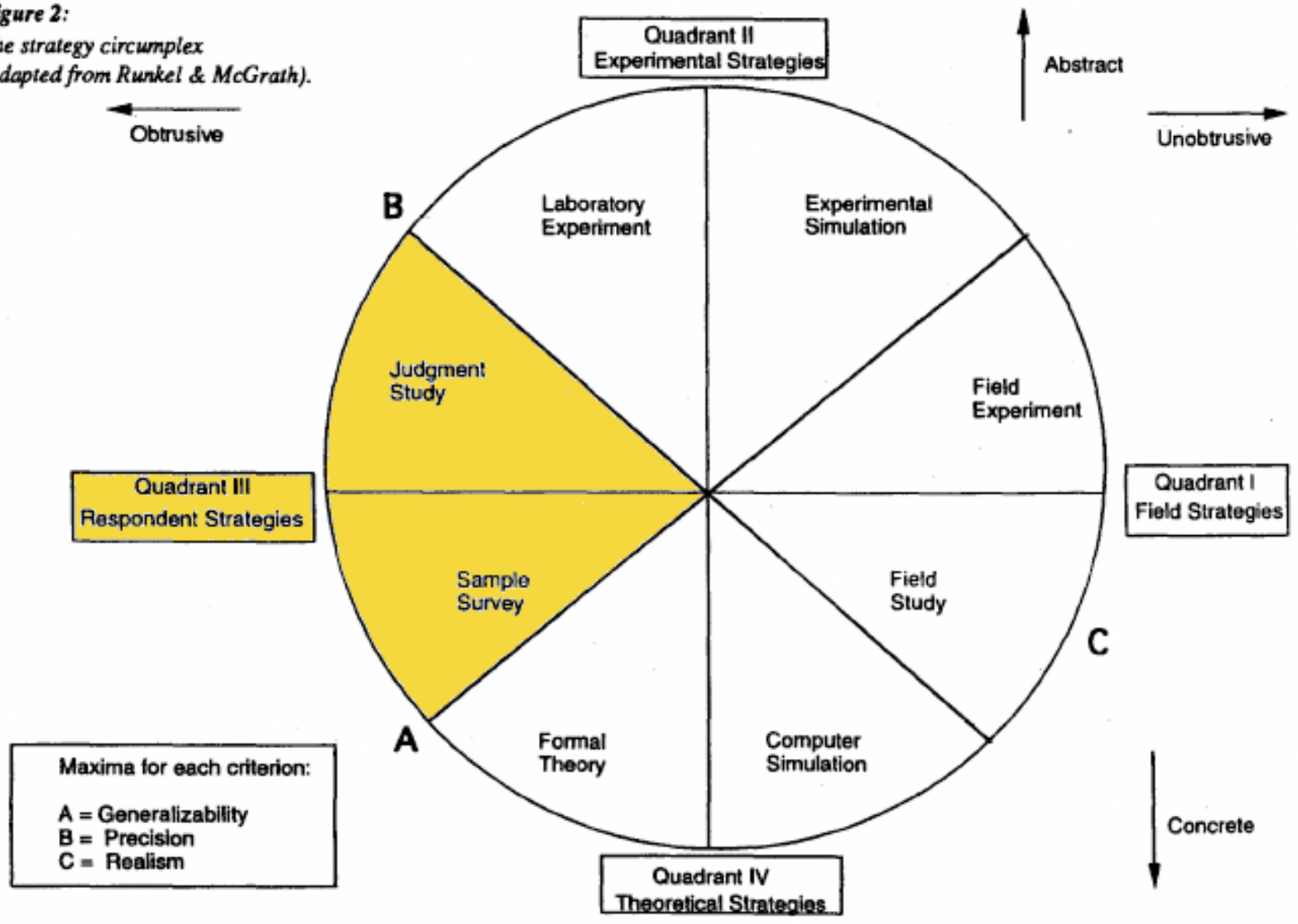
Taxonomy of Methods [McGrath et al. 1994]

Figure 2:
The strategy circumplex (adapted from Runkel & McGrath).



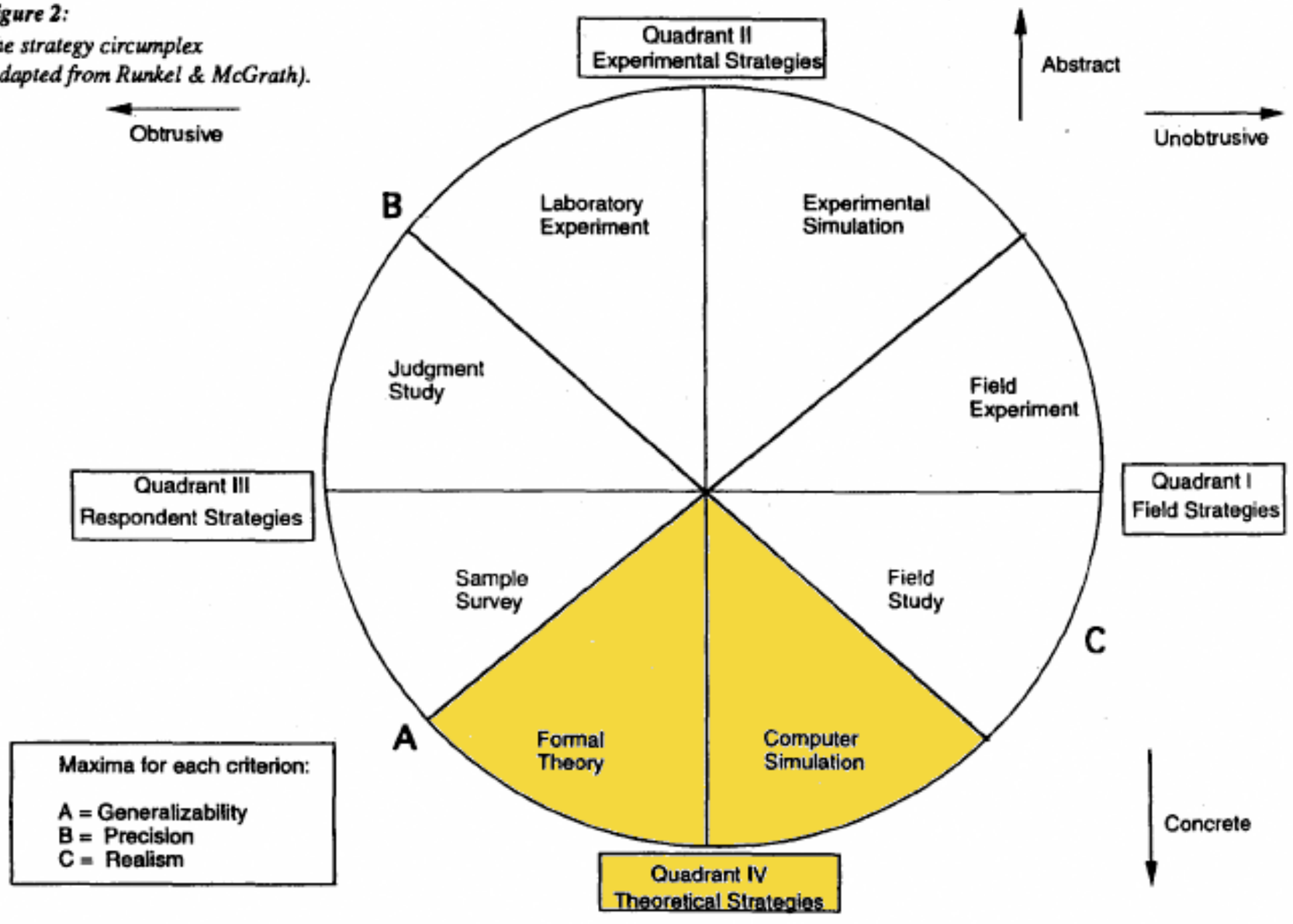
Taxonomy of Methods [McGrath et al. 1994]

Figure 2:
The strategy circumplex (adapted from Runkel & McGrath).



Taxonomy of Methods [McGrath et al. 1994]

Figure 2:
The strategy circumplex (adapted from Runkel & McGrath).



Évaluation et tests

- Introduction
- Approches d'évaluation
- **Méthodes analytiques**
- Méthodes empiriques
- Évaluation 2.0 : passer à l'échelle
- Design expérimental

Les grands types d'évaluation analytique

Basée sur des modèles

- Évaluation selon des modèles d'interaction

Basée sur l'inspection

- Review d'experts / critique du design
- Cognitive walkthrough
- Évaluation heuristique

Évaluation basée sur des modèles

GOMS (Goals, Operators, Methods, and Selection rules)

- Les objectifs (Goals) correspondent à ce que l'utilisateur essaie de faire
- Les opérateurs sont les actions faites pour atteindre cet objectif.
- Les méthodes sont les séquences d'opérateur qui permettent d'accomplir un objectif. Il peut y avoir plus d'une méthode pour un même objectif, dans ce cas :
- Les règles de sélection (Selection rules) sont utilisées pour décrire quand un utilisateur voudrait choisir tel méthode plutôt qu'une autre. On ignore souvent ces règles dans une analyse GOMS simple.

KLM

- On analyse une action et on la découpe en étapes atomiques
- On cherche la durée de chaque étape dans une table
- On prédit la durée de l'action complète
- *Permet la prédire avant d'implémenter !*

From http://www.cs.umd.edu/class/fall2002/cmsc838s/tichi/printer/goms.html

[John & Kieras, 1996]

GOMS analysis

```
GOAL: EDIT-MANUSCRIPT
.     GOAL: EDIT-UNIT-TASK ... repeat until no more unit tasks
.     .     GOAL: ACQUIRE UNIT-TASK
.     .      .      GOAL: GET-NEXT-PAGE ... if at end of manuscript page
.     .      .      GOAL: GET-FROM-MANUSCRIPT
.     .      GOAL: EXECUTE-UNIT-TASK ... if a unit task was found
.     .      .      GOAL: MODIFY-TEXT
.     .      .      .   [select: GOAL: MOVE-TEXT* ...if text is to be moved
.     .      .      .       GOAL: DELETE-PHRASE ...if a phrase is to be deleted
.     .      .      .       GOAL: INSERT-WORD] ... if a word is to be inserted
.     .      .      .     VERIFY-EDIT

*Expansion of MOVE-TEXT goal
GOAL: MOVE-TEXT
.     GOAL: CUT-TEXT
.     .     GOAL: HIGHLIGHT-TEXT
.     .     .     [select**: GOAL: HIGHLIGHT-WORD
.     .     .     .   MOVE-CURSOR-TO-WORD
.     .     .     .   DOUBLE-CLICK-MOUSE-BUTTON
.     .     .     .   VERIFY-HIGHLIGHT
.     .     .           GOAL: HIGHLIGHT-ARBITRARY-TEXT
.     .     .     .      MOVE-CURSOR-TO-BEGINNING    1.10
.     .     .     .      CLICK-MOUSE-BUTTON          0.20
.     .     .     .      MOVE-CURSOR-TO-END          1.10
.     .     .     .      SHIFT-CLICK-MOUSE-BUTTON    0.48
.     .     .     .      VERIFY-HIGHLIGHT]           1.35
.     .     GOAL: ISSUE-CUT-COMMAND
.     .     .     MOVE-CURSOR-TO-EDIT-MENU          1.10
.     .     .     PRESS-MOUSE-BUTTON                0.10
.     .     .     MOVE-CURSOR-TO-CUT-ITEM           1.10
.     .     .     VERIFY-HIGHLIGHT                  1.35
.     .     .     RELEASE-MOUSE-BUTTON              0.10
```

• • •

From http://www.cs.umd.edu/class/fall2

[John & Kieras, 1996]

GOMS analysis

```
*Expansion of MOVE-TEXT goal
GOAL: MOVE-TEXT
.     GOAL: CUT-TEXT
.     .     GOAL: HIGHLIGHT-TEXT

...

.     .   GOAL: ISSUE-CUT-COMMAND
.     .   .     MOVE-CURSOR-TO-EDIT-MENU          1.10
.     .   .     PRESS-MOUSE-BUTTON                0.10
.     .   .     MOVE-CURSOR-TO-CUT-ITEM           1.10
.     .   .     VERIFY-HIGHLIGHT                  1.35
.     .   .     RELEASE-MOUSE-BUTTON              0.10
.     GOAL: PASTE-TEXT
.     .    GOAL: POSITION-CURSOR-AT-INSERTION-POINT
.     .   MOVE-CURSOR-TO-INSERTION-POIONT         1.10
.     .   CLICK-MOUSE-BUTTON                      0.20
.     .   VERIFY-POSITION                         1.35
.     .   GOAL: ISSUE-PASTE-COMMAND
.     .   .     MOVE-CURSOR-TO-EDIT-MENU          1.10
.     .   .     PRESS-MOUSE-BUTTON                0.10
.     .   .     MOVE-MOUSE-TO-PASTE-ITEM          1.10
.     .   .     VERIFY-HIGHLIGHT                  1.35
.     .   .     RELEASE-MOUSE-BUTTON              0.10
TOTAL TIME PREDICTED (SEC)                        14.38

Based on the above GOMS analysis, it should take 14.38 seconds to move text.
```

From http://www.cs.umd.edu/class/fall2002/cmsc838s/tichi/printer/goms.html

KLM

Description	Operation	Time (sec)
Reach for mouse	H[mouse]	0.40
Move pointer to "Replace" button	P[menu item]	1.10
Click on "Replace" command	K[mouse]	0.20
Home on keyboard	H[keyboard]	0.40
Specify word to be replaced	M4K[word]	2.15
Reach for mouse	H[mouse]	0.40
Point to correct field	P[field]	1.10
Click on field	K[mouse]	0.20
Home on keyboard	H[keyboard]	0.40
Type new word	M4K[word]	2.15
Reach for mouse	H[mouse]	0.40
Move pointer on Replace-all	P[replace-all]	1.10
Click on field	K[mouse]	0.20
Total		10.2

Limites

- Des prédictions valides pour un utilisateur expert qui ne fait pas d'erreur
  - les experts font aussi des erreurs !
  - pas de prise en compte des utilisateurs novices ou intermédiaires qui font des erreurs occasionnelles.
  - il existe des extensions qui essaient de modéliser l'apprentissage
- Toutes les tâches ont un objectif clair
  - Beaucoup de tâches ne sont pas si dirigées spécialement en design UX.
- Ne prend pas en compte les différences individuelles entre utilisateurs
  - Basé sur des moyennes statistiques
- Ne prend pas en compte les aspect sociaux et organisationnels du produit
- Pas d'information sur la qualité d'utilisation et le plaisir provoqué par le produit.
- Pas représentatif des théories actuelles sur la cognition humaine
  - Supposition d'un model linéaire de la cognition avec une activité faite à la fois.

Inspections et critiques d’experts

- Tout au long du processus de développement
- Conduite par des développeurs et des experts (internes ou externes)
- Outil pour identifier des problèmes
- Peut aller d’une heure à une semaine de travail
- Préférer une approche structurée
  - Les reviewers doivent pouvoir communiquer sur tous les problèmes (sans fâcher l’équipe)
  - Les critiques ne doivent pas être aggressive envers les développeurs / designers
  - L’objectif principal est d’identifier les problèmes (pas leur source)
- Des solutions peuvent être suggérées a l’équipe

Méthodes d'inspection

Inspection des "Guidelines"

- Vérifier que l'interface respecte bien un ensemble de règles.

Inspection centrée consistance

- Vérifier que l'interface et consistante
  avec elle-même, avec les applications liées, avec l'OS
- Une vue générale peut aider, ex : impression des pages clés du site collées au mur.
- On peut forcer la consistance par les outils, ex : via les CSS pour les sites Web

Procédure d'inspection

- Trouver des experts/reviewers
- Définir un plan avec des limites temporelles
- Préparer le materiel pour les reviewers, y compris les critères d'intérêt
- Sur place ou sur un autre site
- Rédaction d'un rapport et définition des conséquences

+/- des évaluations par experts

- Results of informal reviews and inspections are often directly used to change the product
  - ... still state of the art in many companies!
  - The personal view of the CEO, or his partner ...

- Really helpful evaluation
  - Is explicit
  - Has clearly documented findings
  - Can increase the quality significantly

- Expert reviews and inspections are a starting point for change

plus ici : http://hcibib.org/tcuid/chap-4.html#4-1

Cognitive Walkthrough

One or more evaluators going through a set of tasks

- Evaluating understandability and ease of learning

Procedure:

- Defining the input:
  - Who will be the users of the system?
  - What task(s) will be analyzed?
  - What is the correct action sequence for each task?
  - How is the interface defined?
- During the walkthrough:
  - 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 to be achieved?
  - If the correct action is performed, will the user see that progress is being made toward solution of the task?

Les critères ergonomiques (usability guidelines)

- Don Norman’s principles:
  - visibility, affordances, natural mapping, and feedback
- Ben Shneiderman’s 8 Golden Rules of UI design
- Bruce Tognazzini’s 16 principles:
  - http://www.asktog.com/basics/firstPrinciples.html
- Christian Bastien’s Ergonomic Criteria
- Jakob Nielsen’s Heuristics

L’évaluation heuristique

Conçue comme une méthode d’évaluation “discount” basée sur l’inspection :

- Évaluation rapide, pas cher et facile d’interfaces
- http://www.useit.com/papers/heuristic/

Principes :

- Il y a une liste de propriétés désirable dans une interface : les “heuristiques”
- Ces heuristiques peuvent être vérifiées par des experts avec un résultat clair et précis

10 Usability Heuristics

http://www.useit.com/jakob/photos/

- Meet expectations
  1. Match the real world
  2. Consistency & standards
  3. Help & documentation
- User is boss
  4. User control & freedom
  5. Visibility of system status
  6. Flexibility & efficiency
- Errors
  7. Error prevention
  8. Recognition, not recall
  9. Error reporting, diagnosis, and recovery
- Keep it simple
  10. Aesthetic & minimalist design

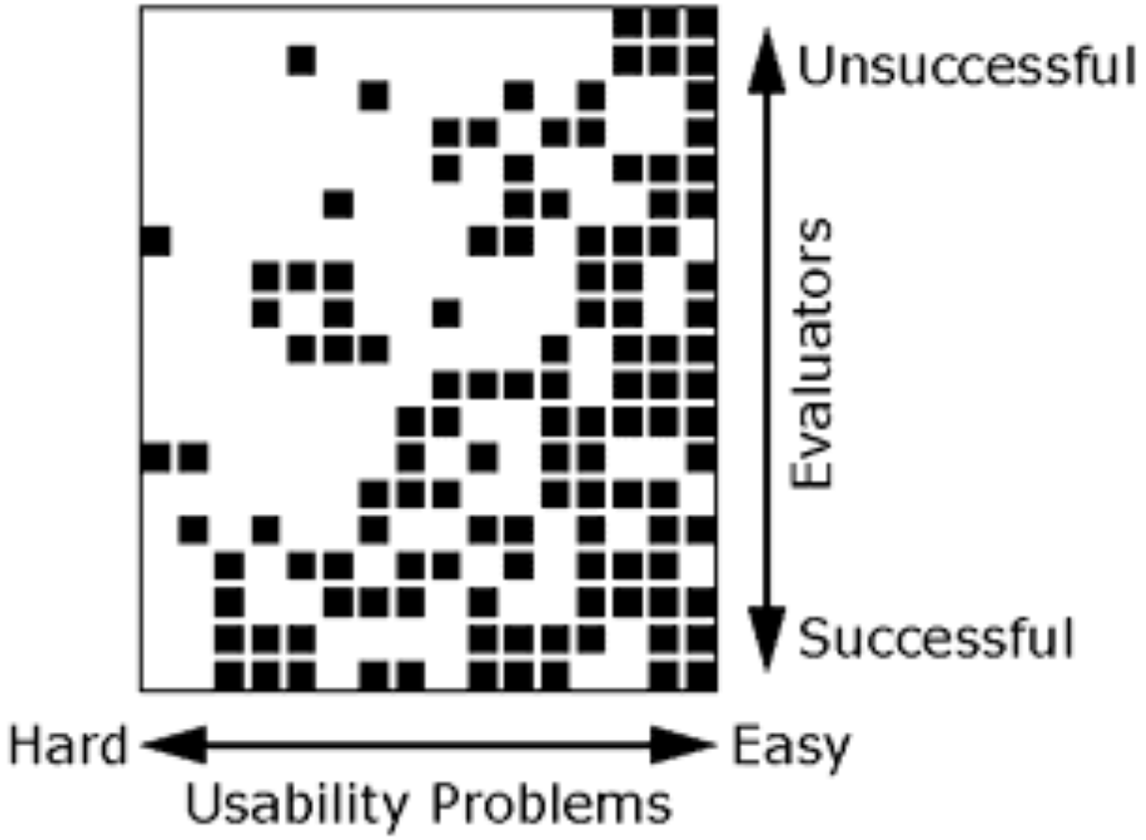
Procédure

- Un petit nombre d'évaluateurs examine l'interface en utilisant et juge de son adéquation aux critères d'utilisabilité (les "heuristiques").
- Soit par inspection, soit au travers d'un scénario
- Les problèmes sont listés et organisé par sévérité
- Les opinions des évaluateurs sont synthétisées dans un rapport.

Quantité d’évaluateurs

Chaque évaluateur ne trouve pas tous les problèmes

Les bons évaluateurs trouvent à la fois des problèmes faciles et d’autres plus difficiles

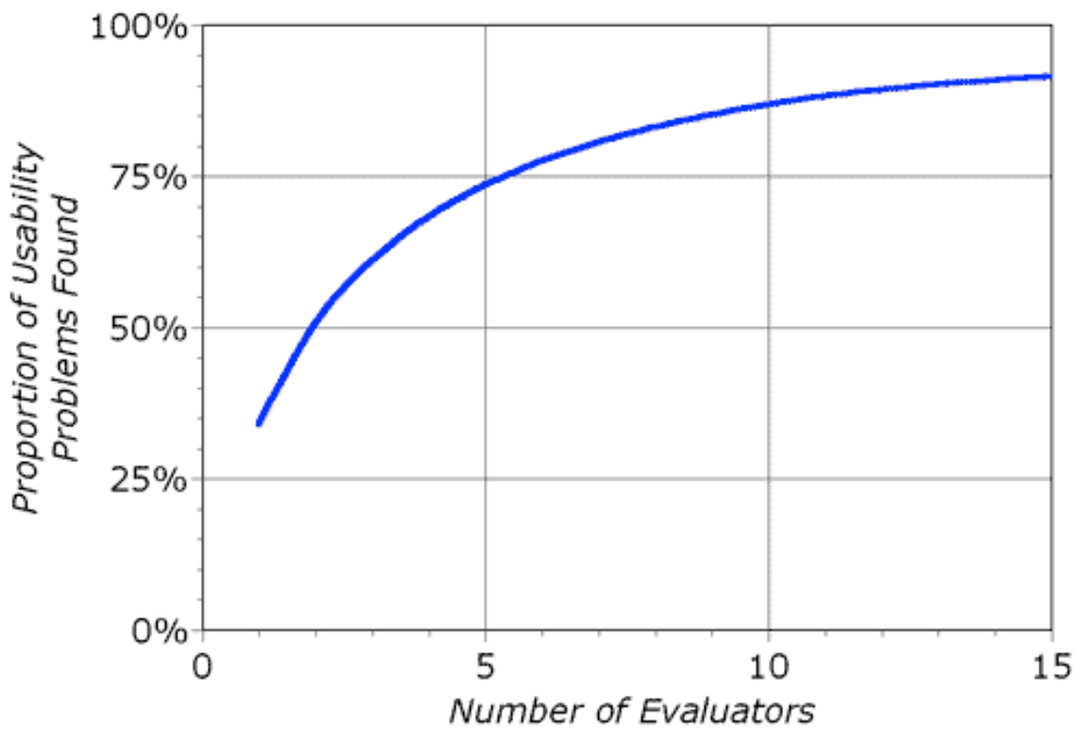


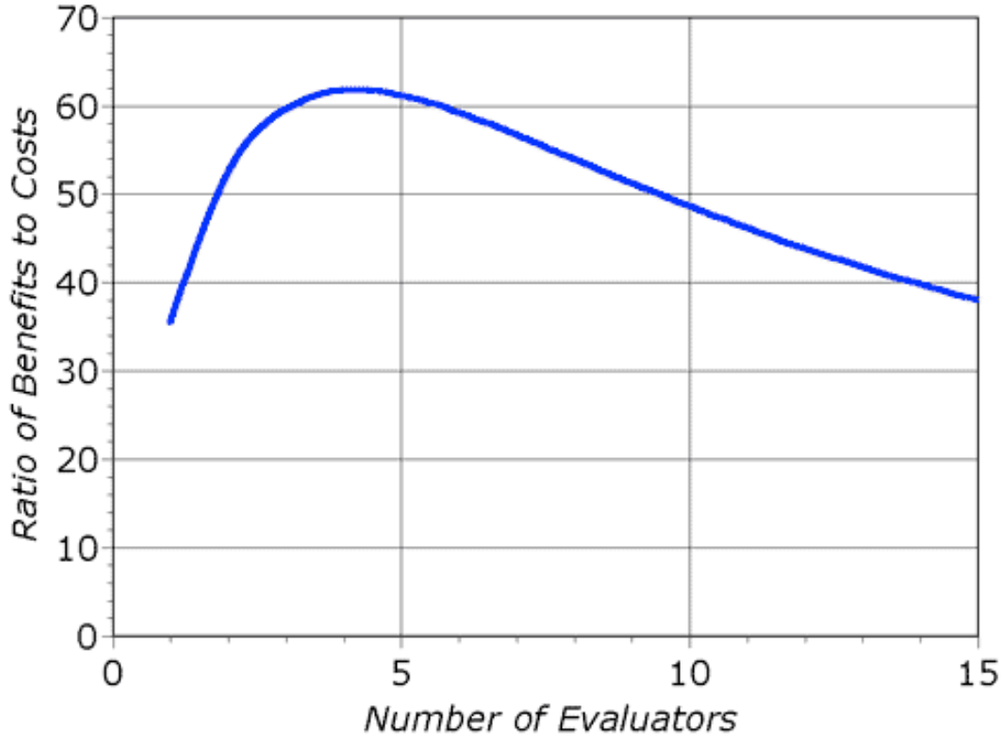
Quantité d’évaluateurs

Un évaluateur unique fournit de mauvais résultats

Trouve seulement 35% des problèmes d’utilisabilité

5 évaluateurs trouvent ~ 75% des problèmes d’utilisabilité



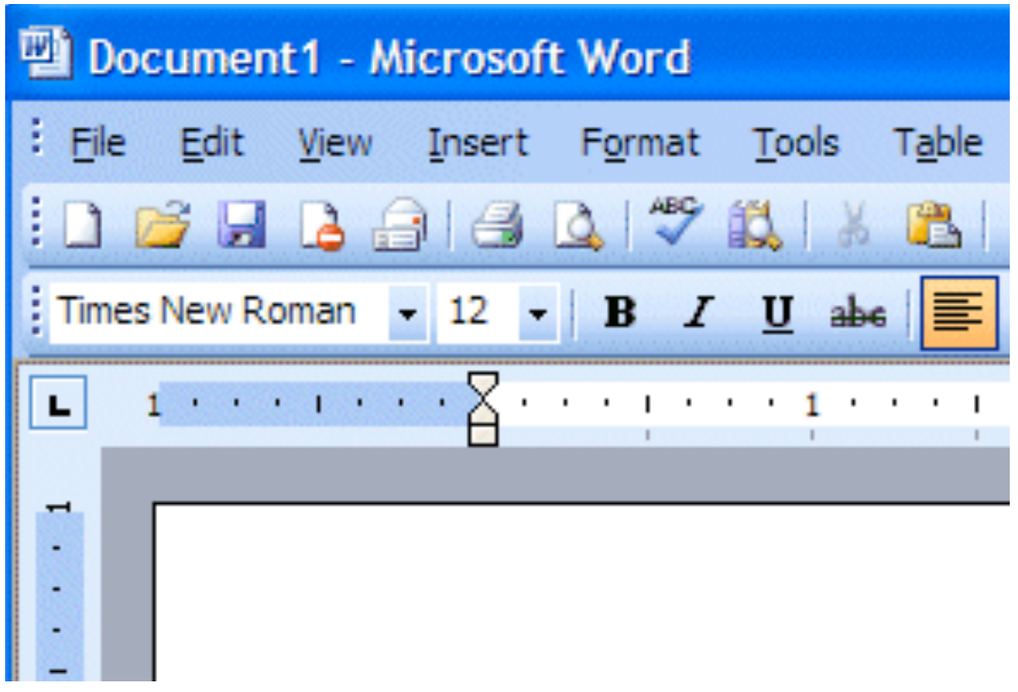


Heuristics

- 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 use
- Aesthetic and minimalist design
- Help users recognise, diagnose, and recover from errors
- Help and documentation

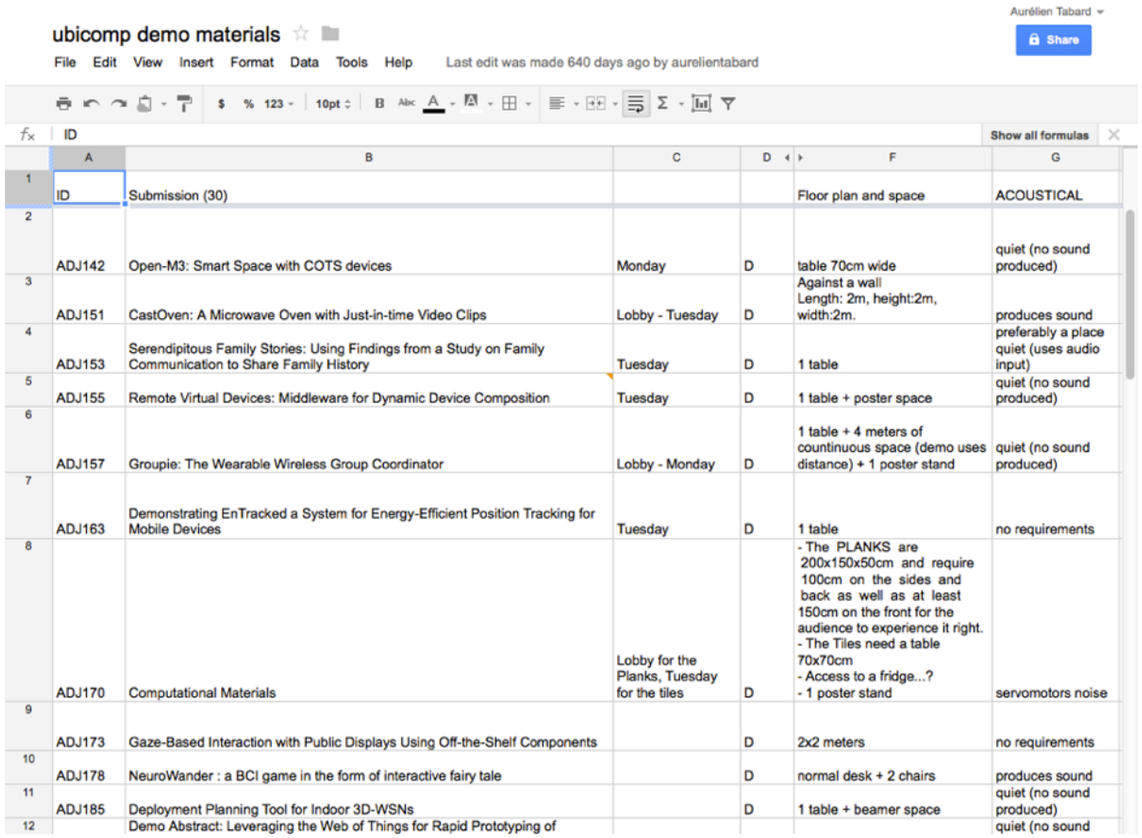
Heuristics

- 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 use
- Aesthetic and minimalist design
- Help users recognize, diagnose, and recover from errors
- Help and documentation



Heuristics

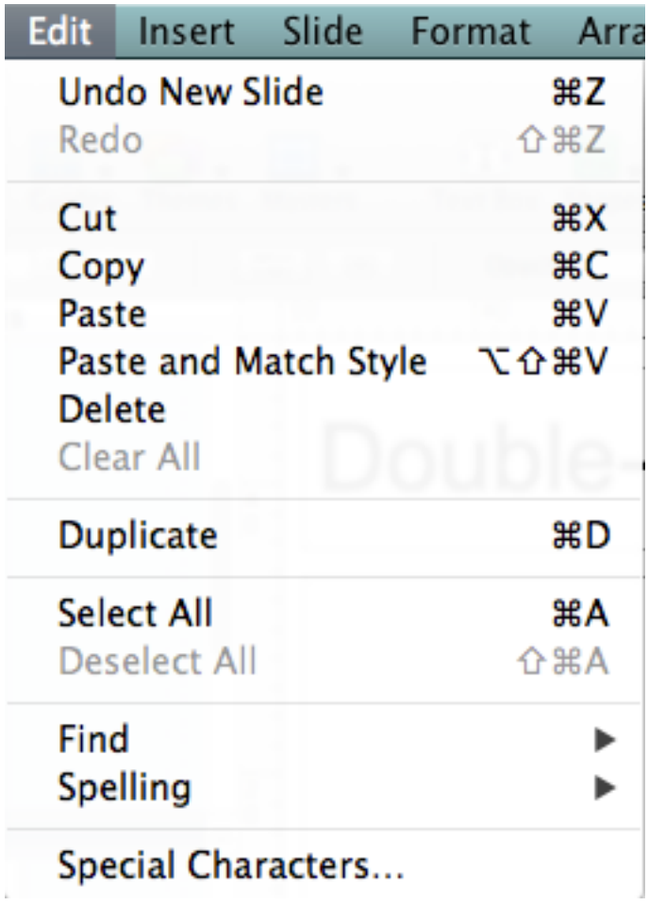
- Visibility of system status
- Match between system and the real world
- User control and freedom
- Consistency and standards
- Error prevention
- Recognition rather than recall
- Flexibility and efficiency of use
- Aesthetic and minimalist design
- Help users recognize, diagnose, and recover from errors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



	A	B	C	D	F	G
1	ID	Submission (30)			Floor plan and space	ACOUSTICAL
2	ADJ142	Open-M3: Smart Space with COTS devices	Monday	D	table 70cm wide	quiet (no sound produced)
3	ADJ151	CastOven: A Microwave Oven with Just-in-time Video Clips	Lobby - Tuesday	D	Against a wall Length: 2m, height:2m, width:2m.	produces sound
4	ADJ153	Serendipitous Family Stories: Using Findings from a Study on Family Communication to Share Family History	Tuesday	D	1 table	preferably a place quiet (uses audio input)
5	ADJ155	Remote Virtual Devices: Middleware for Dynamic Device Composition	Tuesday	D	1 table + poster space	quiet (no sound produced)
6	ADJ157	Groupie: The Wearable Wireless Group Coordinator	Lobby - Monday	D	1 table + 4 meters of countinuous space (demo uses distance) + 1 poster stand	quiet (no sound produced)
7	ADJ163	Demonstrating EnTracked a System for Energy-Efficient Position Tracking for Mobile Devices	Tuesday	D	1 table	no requirements
8	ADJ170	Computational Materials	Lobby for the Planks, Tuesday for the tiles	D	- The PLANKS are 200x150x50cm and require 100cm on the sides and back as well as at least 150cm on the front for the audience to experience it right. - The Tiles need a table 70x70cm - Access to a fridge...? - 1 poster stand	servomotors noise
9	ADJ173	Gaze-Based Interaction with Public Displays Using Off-the-Shelf Components		D	2x2 meters	no requirements
10	ADJ178	NeuroWander : a BCI game in the form of interactive fairy tale		D	normal desk + 2 chairs	produces sound
11	ADJ185	Deployment Planning Tool for Indoor 3D-WSNs		D	1 table + beamer space	quiet (no sound produced)
12		Demo Abstract: Leveraging the Web of Things for Rapid Prototyping of				quiet (no sound

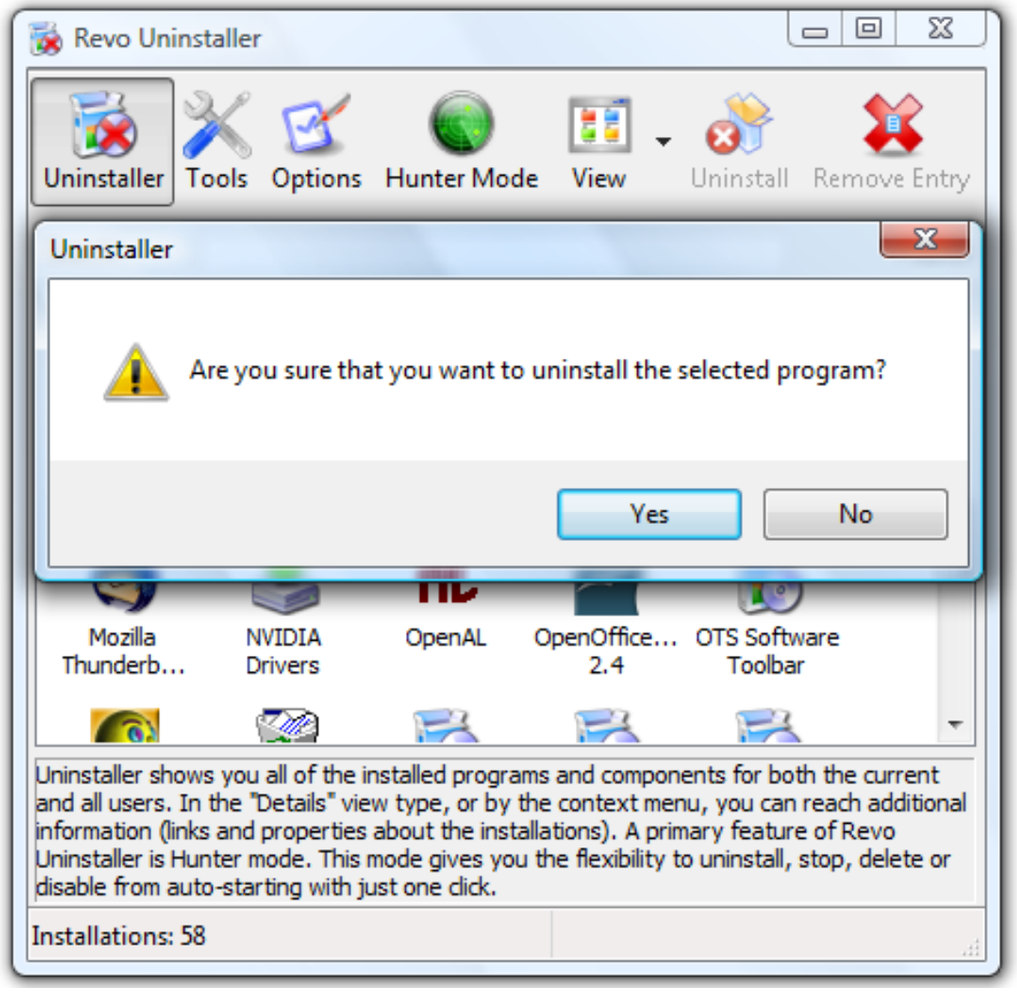
Heuristics

- 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 use
- Aesthetic and minimalist design
- Help users recognize, diagnose, and recover from errors
- Help and documentation



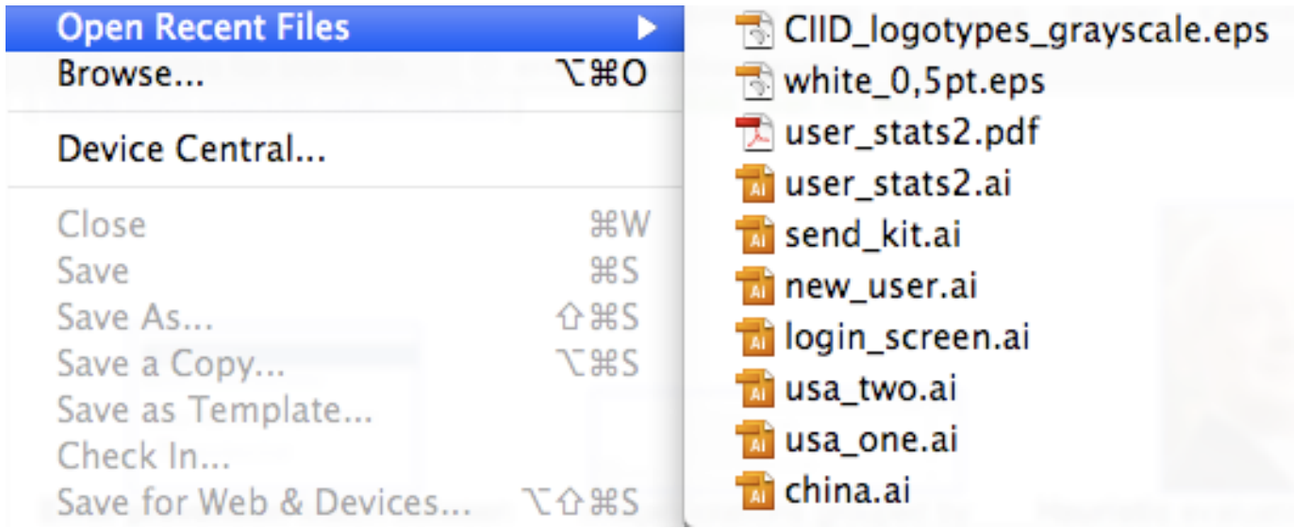
Heuristics

- 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 use
- Aesthetic and minimalist design
- Help users recognize, diagnose, and recover from errors
- Help and documentation



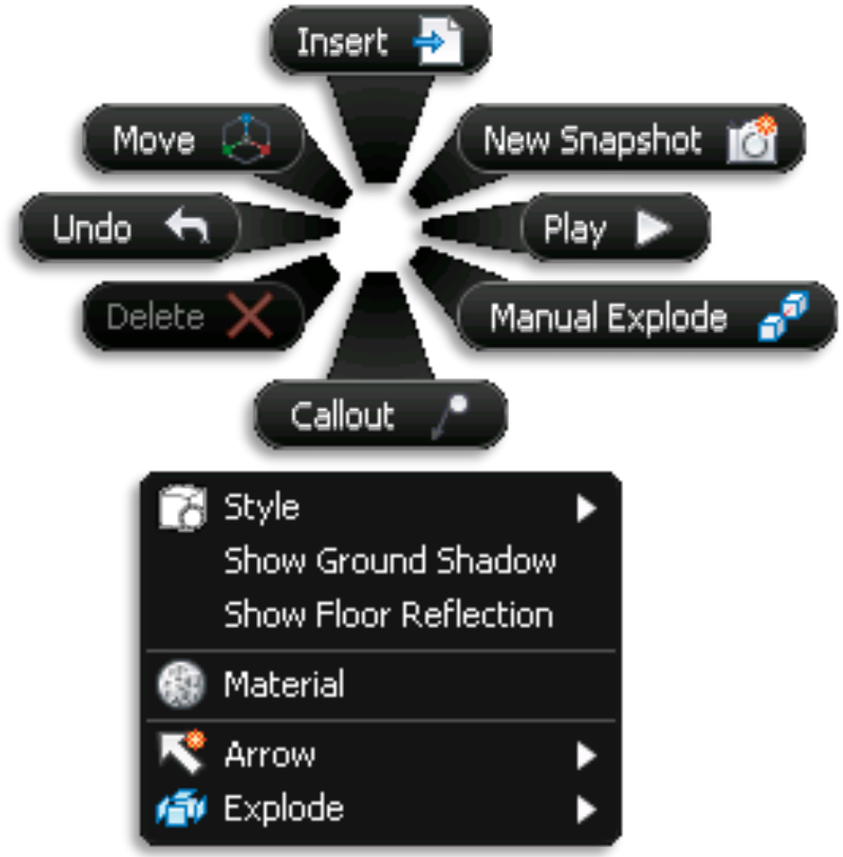
Heuristics

- 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 use
- Aesthetic and minimalist design
- Help users recognize, diagnose, and recover from errors
- Help and documentation



Heuristics

- 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 use
- Aesthetic and minimalist design
- Help users recognize, diagnose, and recover from errors
- Help and documentation

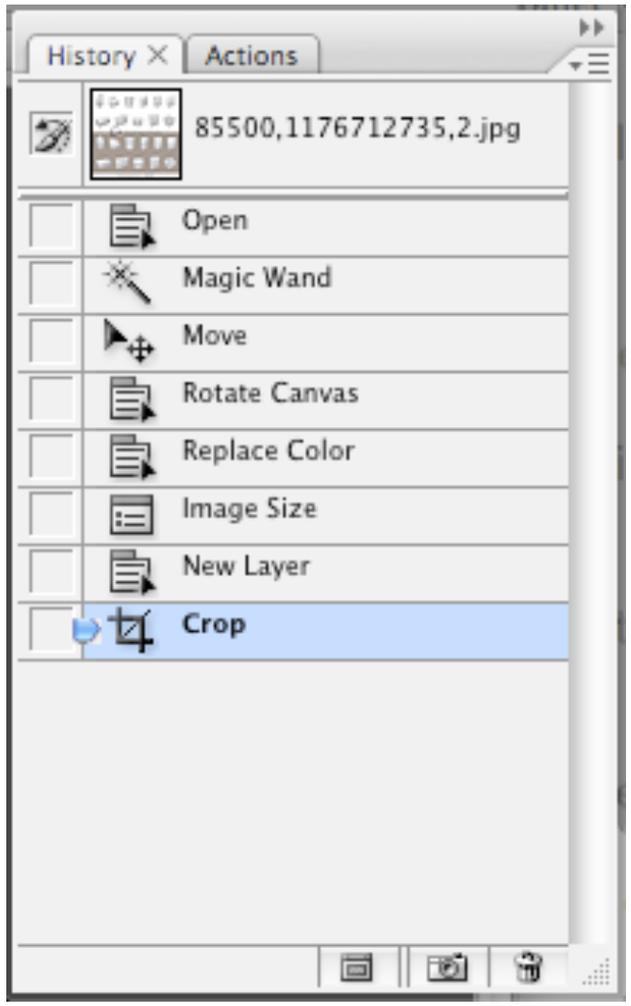


Heuristics

- 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 use
- Aesthetic and minimalist design
- Help users recognize, diagnose, and recover from errors
- Help and documentation

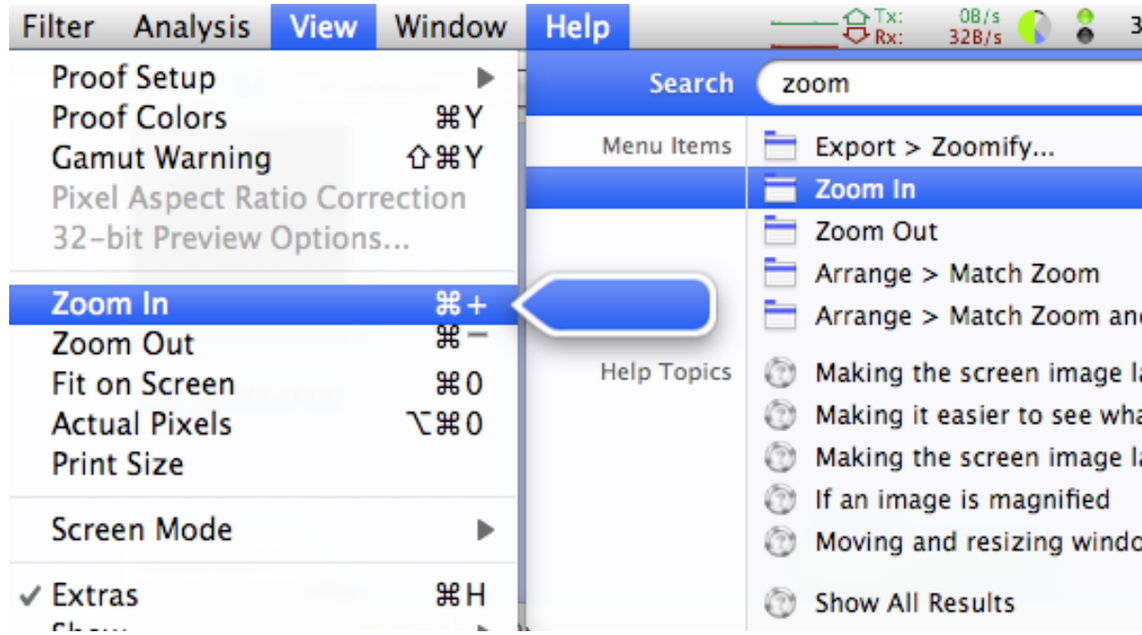
Heuristics

- 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 use
- Aesthetic and minimalist design
- Help users recognize, diagnose, and recover from errors
- Help and documentation



Heuristics

- 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 use
- Aesthetic and minimalist design
- Help users recognize, diagnose, and recover from errors
- Help and documentation



Échelle de sévérité

Facteurs

- Fréquence : est ce commun?
- Impact: quelle difficulté à résoudre le problème?
- Persistance: faut il affronter le problème dans la durée?

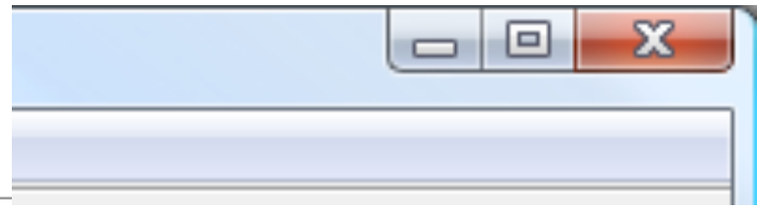
Échelle de sévérité

- Cosmétique : pas nécessaire de le résoudre
- Mineur: résolution nécessaire mais basse priorité
- Majeur: résolution nécessaire et haute priorité
- Catastrophique : résolution obligatoire

Rédiger de bonnes évaluations heuristiques

- Les évaluations heuristiques doivent être compréhensible des développeurs et des manageurs
- Inclure les commentaires positifs et négatifs
  - Ex : La barre d'icône est simple, avec un bon contraste et peu de couleur (minimalist design)
- Avoir du tact
  - Pas : l'organisation des menus est bordélique
  - Plutôt : les menus ne sont pas organisés par fonction
- Être spécifique
  - Pas : le texte est illisible
  - Plutôt : le text est trop petit avec un mauvais contraste (texte noir sur fond vert)

Exemple



Quoi inclure :

- Le problème
- L'heuristique
- La description
- La sévérité
- Des recommendations
- Un screenshot (si ça aide)

Severe: User may close window without saving data (error prevention)

If the user has made changes without saving,
and then closes the window using the Close
button, rather than File >> Exit, no confirmation
dialog appears.

Recommendation: show a confirmation dialog or save automatically

Résumé

- L’évaluation heuristique est une méthode “low-cost”
- Il est mieux que les évaluateurs passe deux fois sur l’interface
  - Leur demander si elle respecte les heuristiques
  - Noter quand ce n’est pas le cas et pourquoi
- Les évaluateurs notent la sévérité indépendamment
- Combiner les découvertes de 3 à 5 evaluateurs
- Discuter les problemes avec l’équipe de design
- Alternative (moins chère) au test utilisateur
- Identifie souvent des problèmes différents alors bonne alternative.

Évaluation et tests

- Introduction
- Approches d’évaluation
- Méthodes analytiques

Prochaine séance :

- Méthodes empiriques
- Évaluation 2.0 : passer à l’échelle
- Design expérimental

Activité : Design Walkthrough

Évaluation analytique formative informelle

Évaluer un aspect précis pas à pas

- l'interface
- un scénario
- un prototype
- ...

Comme pour le brainstorming :

- implication de divers participants (designers, développeurs, marketing...)
- temps limité
- recueillir le maximum de commentaires
- les critiques viennent avec des suggestions (= être constructif)

Field Studies

Activities studied in situation (where they usually happen).

Advantages:

- Can reveal results on user acceptance
- Allows longitudinal studies, including learning, collaboration and adaptation

Problems:

- In general very expensive
- Reliable product (or prototype) needed
- How to get observations?
  - Collecting usage data
  - Direct observation, regular interviews

Focus Groups

- Informal, qualitative group discussion of specific topic
  - Get indication of how people think and feel
  - Collecting opinions, attitudes, feelings, needs and ideas
  - Understand why people act or react in a certain way
- Early in the design process, before UI design or implementation
- Complementary to more elaborate, quantitative studies
- Setup:
  - Groups of 6 to 8 participants
  - Conducted by a moderator
  - Duration 1,5 to 2 hours
- Analysis of script, video recording => simple report with quotes

Focus groups pro&cons

Advantages

- Fast, easy, cheap
- In depth information about users' opinions, motives, motivations
- Flexible, exploration of different topics and materials

Disadvantages

- Not representative, hard to generalize
- What users think vs. what users actually do
- Analysis can be laborious
- Can be biased by moderator or people with strong opinions

Getting Participants

Representative of target users

- job-specific vocabulary / knowledge
- tasks

Approximate if needed

- system intended for doctors
  - get medical students
- system intended for engineers
  - get engineering students

Use incentives to get participants

Ethics!

Pressures on users:

- Performance anxiety
- Feels like an intelligence test
- Comparing self with other subjects
- Feeling stupid in front of observers
- Competing with other subjects

Respect and control

Time

- Don’t waste it

Comfort

- Make the user comfortable

Informed consent

- Inform the user as fully as possible

Privacy

- Preserve the users privacy

Control

- The user can stop at any time

Before a test

Time

- Pilot-test all materials and tasks

Comfort

- We’re testing the system; were not testing you.
- Any difficulties you encounter are the systems fault. We need your help to find these problems.

Privacy

- Your test results will be completely confidential.

Information

- Brief about purpose of study
- Inform about audio-taping, video-taping, other observers, make sure it is ok or disable the ones the subject is not comfortable with

During the test

Time

- Eliminate unnecessary tasks

Comfort

- Calm,relaxed atmosphere
- Take breaks in long session
- Never act disappointed
- Give tasks one at a time
- First task should be easy, for an early success experience

Privacy

After the test

Comfort

- Say what they've helped you do

Information

- Answer questions that you had to defer to avoid biasing the experiment

Privacy

- Don't publish user-identifying information
- Don't show video or audio without users permission

Metrics

Ease of learning

.:learning time, …

Ease of use

.:performance time, error rates…

User satisfaction

.:surveys…

Not “natural”!

Not “intuitive”!

Not “user friendly”!

What data to gather

observations of what users are doing & thinking

summary of what happened (time, errors, success)

	Process	*Bottom-line*
Qualitative		
Quantitative		

What you gather (quantitative)

Quantitative data, which might include:

.:Success rates

.:Accuracy / Error rates : How many mistakes did people make? And were they fatal or recoverable with the right information?

.:Time on Task: How long does it take people to complete basic tasks? (For example, find something to buy, create a new account, and order the item.)

.:Pages visited, number of steps to reach goal...

.:Recall: How much does the person remember afterwards or after periods of non-use?

.:Emotional Response: Ratings on a satisfaction questionnaire, How does the person felt about the tasks completed? (Confident? Stressed? Would the user recommend this system to a friend?)

What you gather (qualitative)

Qualitative data, which might include notes on:

.:How people reacted to the system.

.:How participants understood it.

.:Which the pathways participants took.

.:Which problems participants had (critical incidents).

.:What participants said as they worked.

.:Participants’ answers to open-ended questions.

You need a plan!

A good plan for usability testing gives the participants:

.:a goal/task (what to do or what question to find the answer for)

.:data, if needed, that a real user would have when going to the site to do that task

You can give the scenario as just the statement of the goal/task or you can elaborate it a little with a very short story that adds motivation to get to the goal.

Participants

The participants must be like the people who will use your product.

Be ready to screen participants (do not grab the first person in the corridor)

Plan on a cost associated with finding the people

.:you may still need to plan on incentives to get participants to participate ...

Test!

Make sure you have everything you need

.:the prototype you are going to test

.:the computer set up for the participant with the monitor, resolution, and connection speed that you indicated in the test plan

.:note-taking forms on paper or set up on a computer

.:consent forms for participants to sign and a pen in case the participant does not bring one

.:questionnaires, if you are using any

.:the participant's copy of the scenarios

.:cameras, microphones, or other recording equipment if you are using any

.:folders to keep each person's paperwork in if you are using paper

Before starting

.:You should know, and have written down

.:objective

.:description of system being testing

.:task environment & materials

.:participants

.:methodology

.:tasks

.:test measures

.:Will help you design a good usability test

.:Will help you figure out how to analyze your data

Usability laboratory

- .:Specifically constructed testing
  - .:Instrumented with data collection
  - .:devices (e.g. microphones, cameras)
- .:Separate observation room
  - .:Usually connected to testing room
  - .:by one-way mirror and audio system
  - .:Data recording and analysis
- .:Test users perform prepared sce
  - .:“Think aloud” technique
  - .:Decide whether to interrupt or not
  - .:Keep variances among tests low

From C|Net “How Google tested Google Instant”
http://news.cnet.com/8301-30684_3-20019652-265.html

Think aloud

.:Need to know what users are thinking, not just what they do

.:Ask users to talk while performing tasks

- .:tell us what they are thinking
- .:tell us what they are trying to do
- .:tell us questions that arise as they work
- .:tell us things they read

.:Make a recording or take good notes

- .:make sure you can tell what they were doing
- .:use a digital watch/clock
- .:take notes, plus if possible record audio & video (or even event logs)

.:Prompt the user to keep talking

- .:“tell me what you are thinking”
- .:Only help on things you have pre-decided

Usability testing analysis and limitations

.:Summarize the data

.:make a list of all critical incidents

.:positive & negative

.:include references back to original data

.:try to judge why each difficulty occurred

.:What does data tell you?

.:UI work the way you thought it would? users take approaches you expected?

.:something missing?

.:Update task analysis & rethink design

.:rate severity & ease of fixing CIs

.:fix both severe problems & make the easy fixes

.:Will thinking aloud give the right answers?

Usability measures

Situations in which numbers are useful

.:time requirements for task completion

.:successful task completion

.:compare two designs on speed or # of errors

Ease of measurement

.:time is easy to record

.:error or successful completion is harder

.:define in advance what these mean

Do not combine efficiency measures with thinking-aloud.

.:talking can affect speed & accuracy

Physiological measurements

Eye tracking

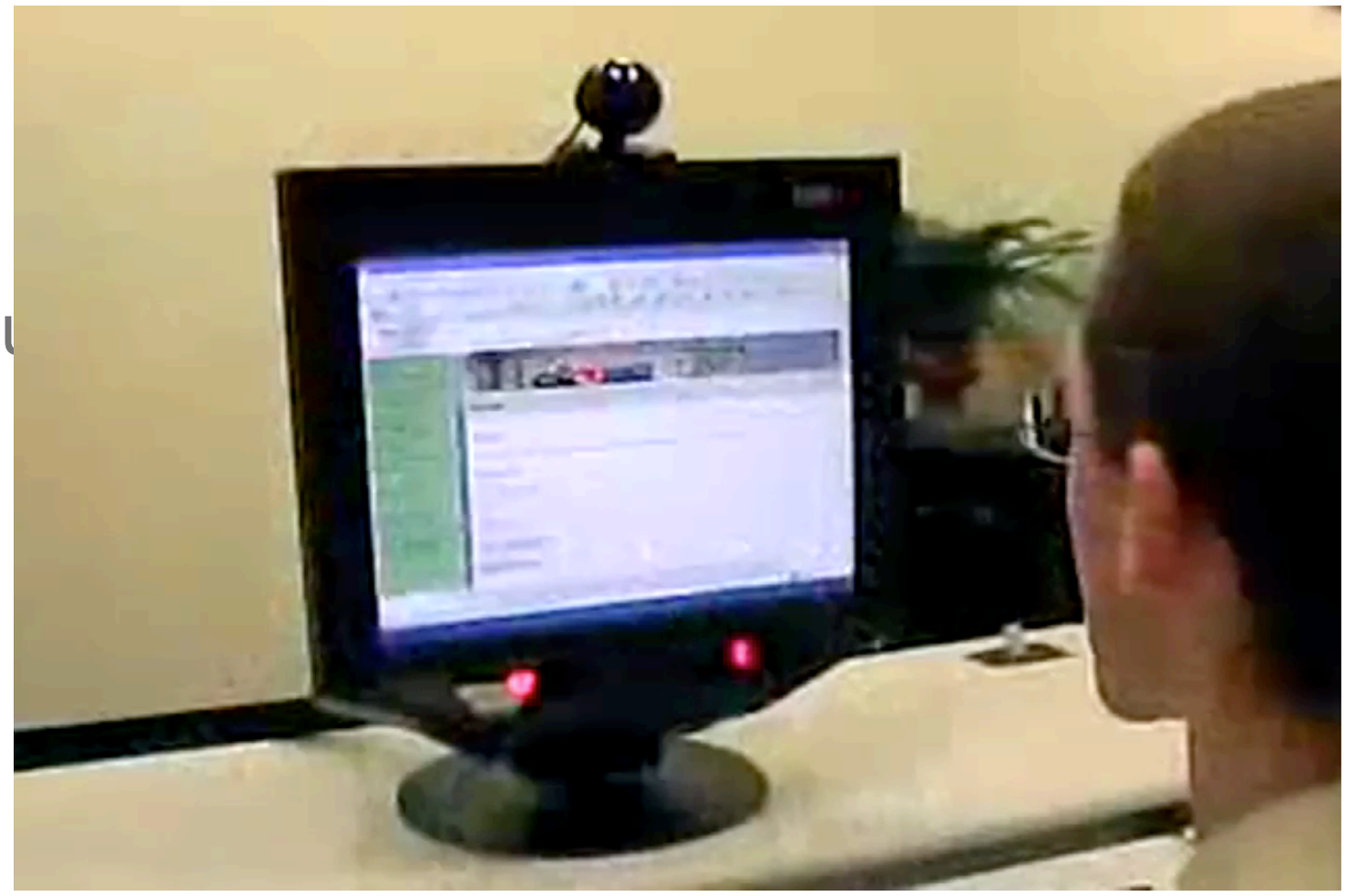
.:well developed and robu

Stress

.:e.g. skin conductivity

Brain activity

.:experimental



Eye-tracker - © Kent State University (US)

Usability lab on the cheap

Goal: Integrate multiple views

.:Capture screen with pointer

.: View of the person interacting with the system

.: View of the environment

Setup:

.:Computer for the test user

.:Application to test

.:Capture tool

.:Computer for the observer

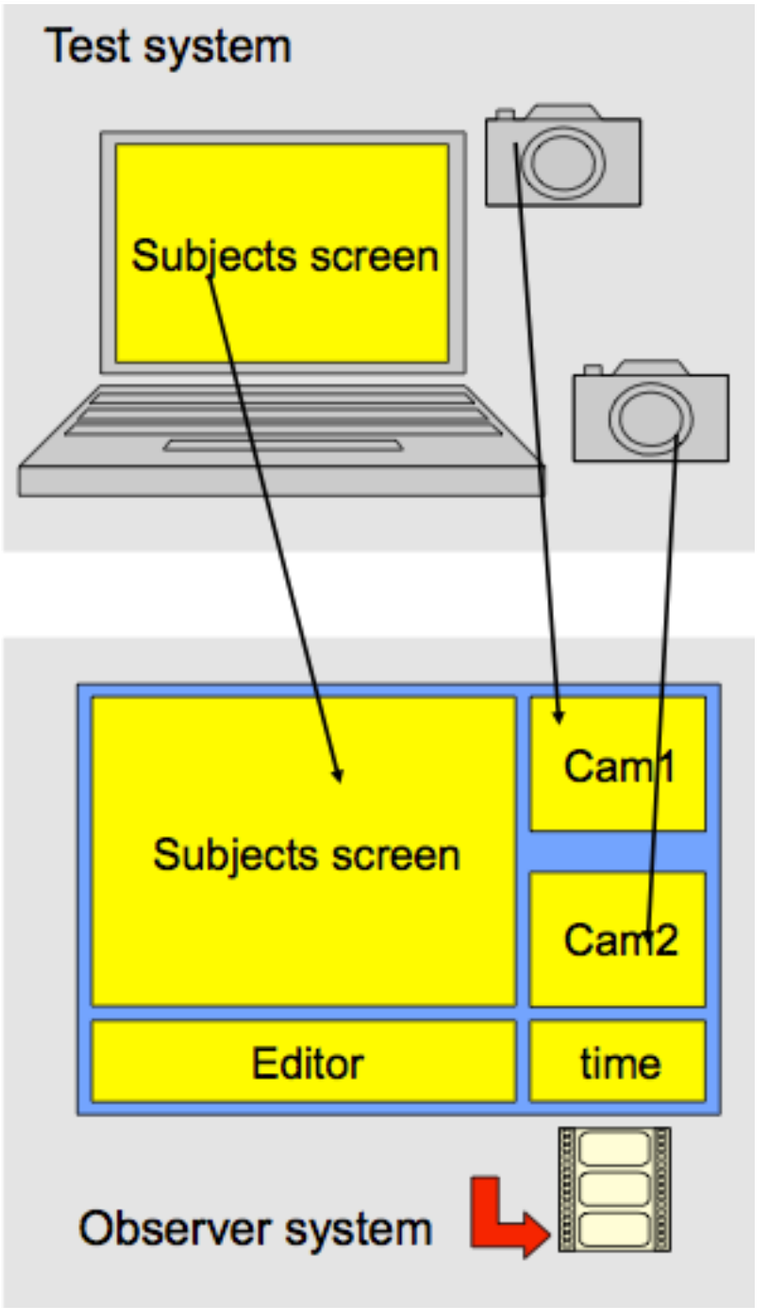
.:See the screen of the subject

.:Attach 2 web cams (face and entire user)

.:Display them on the observer‘s screen

.:Have an editor for the observer‘s notes

.:Capture this screen



Really cheap

Existing tools

Morae

.:http://www.techsmith.com/m

Ovo studio (free for students)

.:http://www.ovostudios.com

Silverback

.:http://silverbackapp.com/