

Methoden zur Normenkonformitätsprüfung im Rahmen von ISO 9241

Matthias Rauterberg

Forschungsgruppe "Mensch-Maschine Interaktion"
Institut für Hygiene und Arbeitsphysiologie
Eidgenössische Technische Hochschule Zürich

Anschrift:

Dr. M. Rauterberg

IHA-ETHZ

Clausiusstrasse 25

CH-8092 Zürich

Schweiz

Tel: 0041-1-632 7082

Fax: 0041-1-632 1173

Email: rauterberg@iha.beprethz.ch

Internet: <http://www.iha.beprethz.ch/pages/forschung/MMI/MMI.HTM>

Usability Methods

Usability inspection is the name of a set of highly cost-effective methods for finding usability problems and improving the usability of a user interface design by inspection.

Topics to be covered include...

- **Definition of usability inspection,**
- **the heuristic evaluation method,**
- **other inspection methods.**
- **Relation between usability inspection methods and user testing.**
- **Severity of usability problems found by usability inspection.**
- **Cost-benefit characteristics of usability inspection methods.**
- **Positioning inspection in the usability engineering lifecycle.**

Evaluation

Assessing the usability of an existing design

- **finding usability problems (to fix them)**
- **formative evaluation: improve interface, find good/bad parts**
- **summative evaluation: are goals met?**

Only one part of the usability engineering lifecycle
(task analysis, goal setting, design, prototyping, iteration, field studies, etc.)

Inspection methods

- **pluralistic walkthrough [Bias 1991]**

- define a scenario (linear path through interface)
- get users, designers/developers, usability specialists in one room
- show user interface one screen at a time (e.g., overheads)
- have participants write down problems before discussion
- discuss the screen (let users speak first)
 - { may use designer/developer as 'living manual' for early help }

- **standards inspection [Wixon, Jones, Tse & Casaday 1994]**

- have a standard expert inspect interface for compliance
 - { may cover most of standards without much task knowledge }

- **consistency inspection [Wixon, Jones, Tse & Casaday 1994]**

- team of designers/developers (one from each project) inspects a set of interfaces
 - for more than one system/application, one at a time

- **feature inspection [Bell 1992]**

- imagine typical user task
- list sequence of features used to accomplish the task
- check for long sequences, cumbersome steps, additional knowledge, etc.

- **cognitive walkthrough [Polson, Lewis, Rieman & Wharton 1992]**

- imagine typical user task
- use the system to perform the task, 'defining' the correct solution sequence
- hand-simulate user's problem solving process at each step
- check if user's goal/memory leads to the defined solution sequence

- **quantitative metrics [Rauterberg 1994]**

Evaluation methods

- **highly informal evaluation: heuristic evaluation**

Look at interface and make lists of its problems [Nielsen and Molich 1990]:

- according to checklist of established usability heuristics
- may also apply any additional usability knowledge

Two or more passes through interface:

- inspect flow of interface
- inspect each screen (dialog box, system message, etc.), one at a time

Typical session length: 1–2 hours.

May use observer to help evaluator and note problems mentioned.

Afterwards: aggregate lists of problems from multiple evaluators

- **informal evaluation: usability inspection**

Goals to be met in a somewhat systematic way:

- generate list of usability problems (main goal)
- contribute to building design rationale (artifact inspection)
- provide feedback in design courses [Nielsen et al. 1992]
- evolve a parallel design [Nielsen 1993]

Tools support for inspection:

- mostly none
- online forms for cognitive walkthroughs [Lewis et al. 1992]
- online/hypertext guidelines/standards documents
- CSCW tools for team heuristic evaluations

(show panel to be discussed for annotation/drawing/pointing)

- **structured evaluation: usability tests**

Frage:

Gibt es Methoden zur Normenkonformitätsprüfung im Rahmen der ISO 9241 ?

Antwort:

Nein !

Begründung:

Die ISO 9241 enthält keine ge-"normten" Metriken mit zugehörigen Grenzwerten.

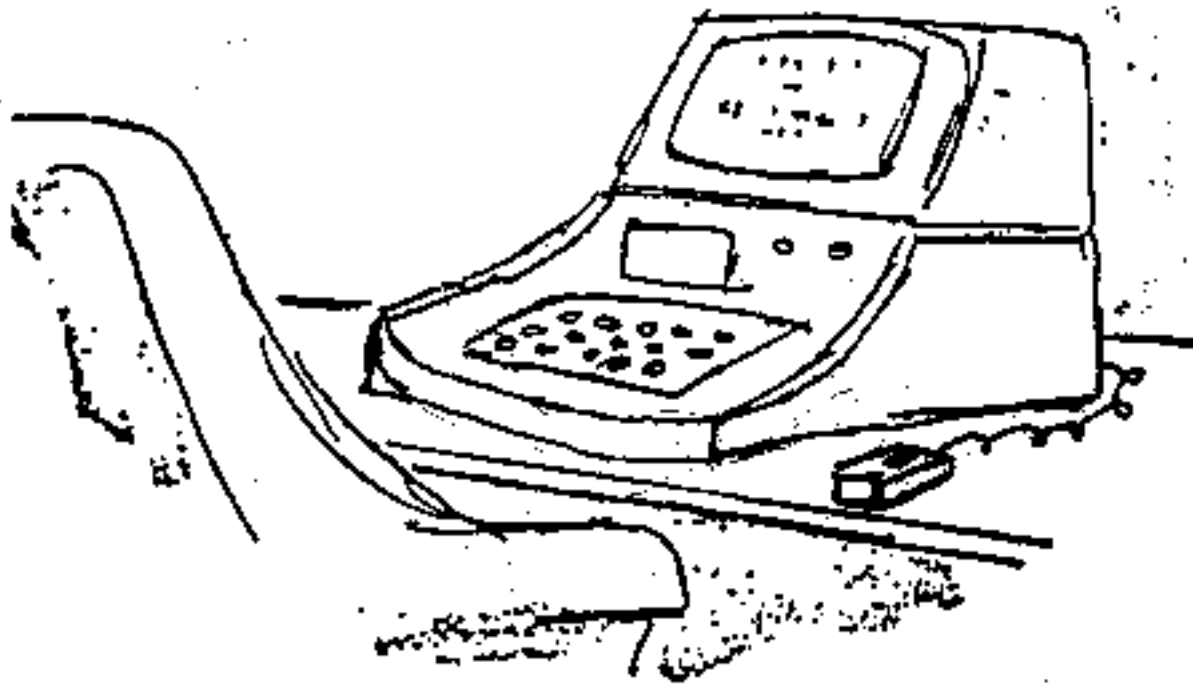
Was nun?

Methoden zur Qualitätssicherung

Benutzer
vorhanden...

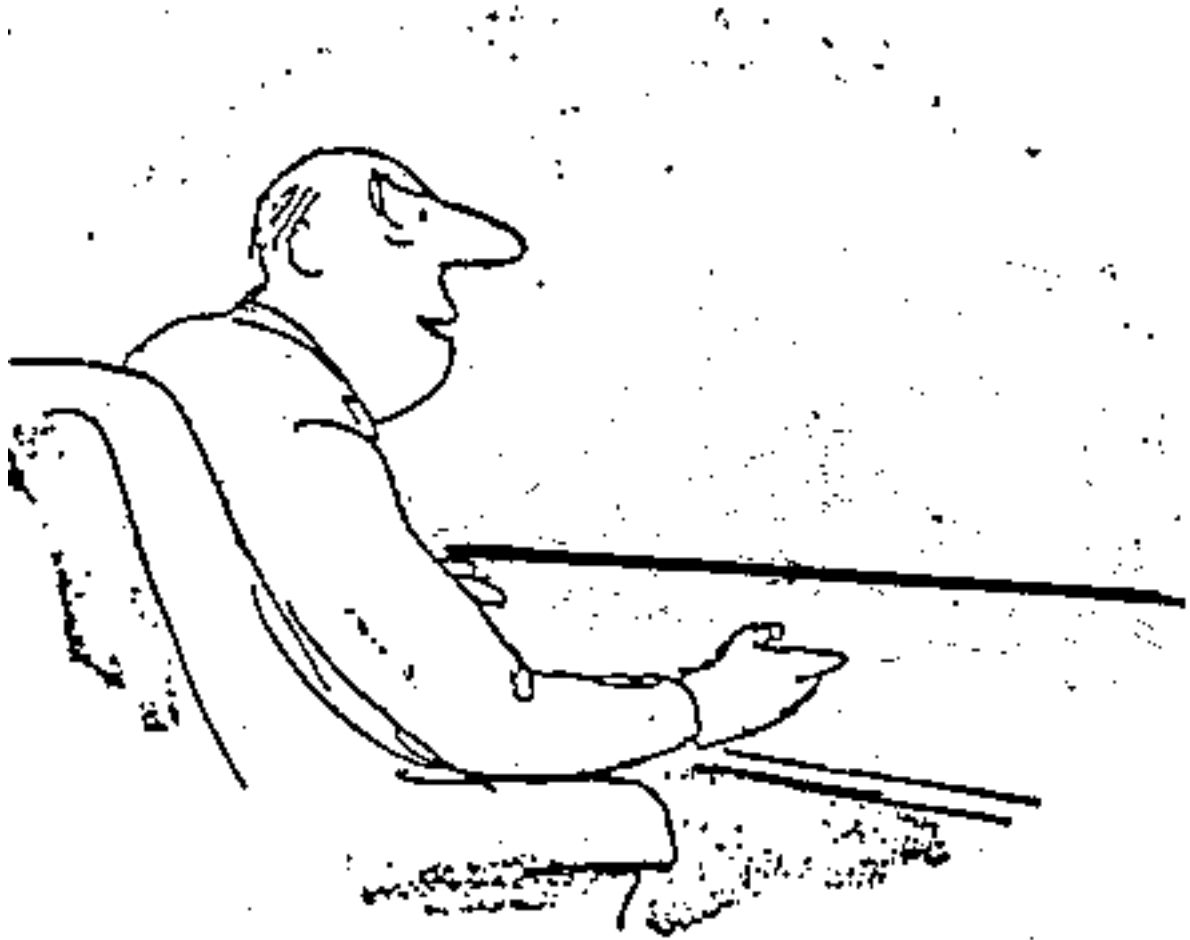
| | | | | |
|--------------------------|----------|---|---|-----------------------|
| | | virtuell | real | |
| Computer vorhanden... | virtuell | formaler Ansatz: formale Theorie | benutzer-zentriert: Fragebogen, Interview, Mock-ups | ökologische Validität |
| | real | produkt-zentriert: Experten Evaluation | interaktions-zentriert: Usability-Test | |
| | | Aufwand und Kosten | | |

der produkt-zentrierte Meß-Ansatz



- Checklisten
- Experten-Evaluation

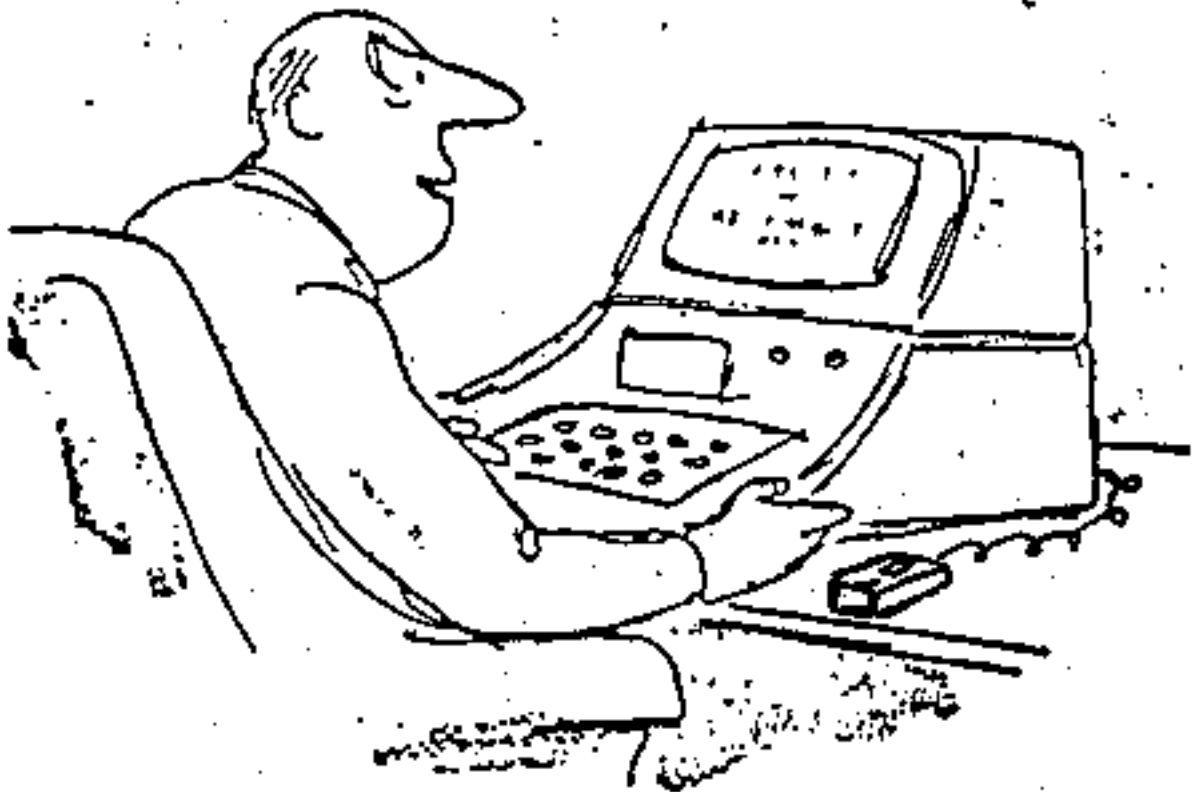
der benutzer-zentrierte Meß-Ansatz



- **mündliche Befragung (Interview)**
- **schriftliche Befragung (Umfragen)**
- **Diskussionen (zB. in Workshops)**

interaktions-zentrierter Meß-Ansatz

**He !
Ich Chef - du Werkzeug !
Begreifen ?**



- **aufgaben-orientierte Usability-Tests**
- **induktive Usability-Tests (formative evaluation)**
- **deduktive Usability-Tests (summative evaluation)**

Modell-1

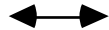
Auftraggeber



Software-Entwickler



BenutzerIn



Modell-2

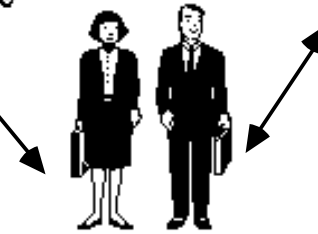
Auftraggeber



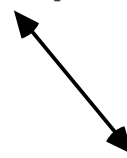
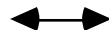
Software-Entwickler



BenutzerIn



Usability-Experten



Modell-3

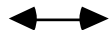
Auftraggeber



Software-Entwickler



BenutzerIn

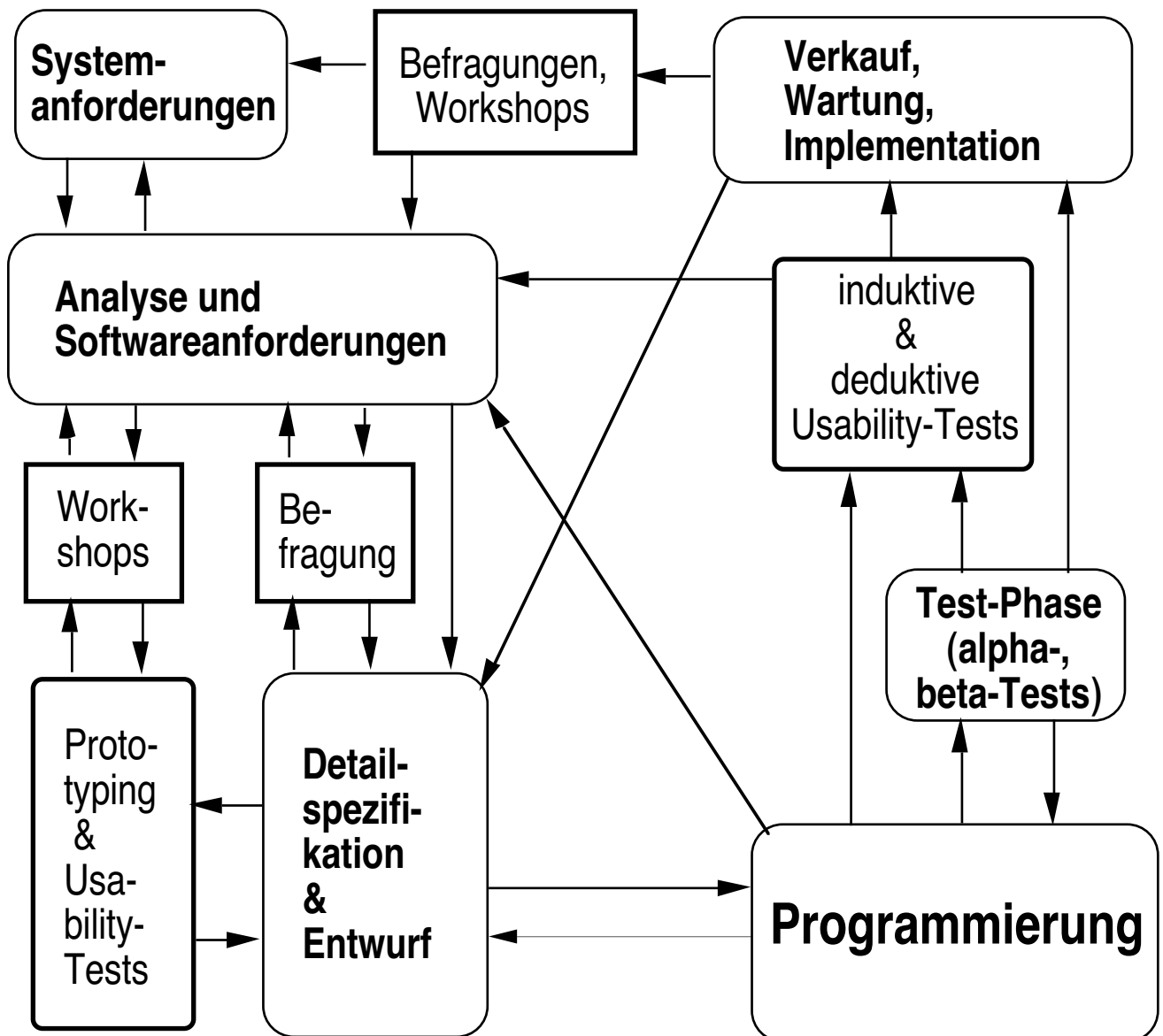


Das Quadranten-Modell

[BOSS-Projekt, Rauterberg 1991]

Quadrant-I: Analyse

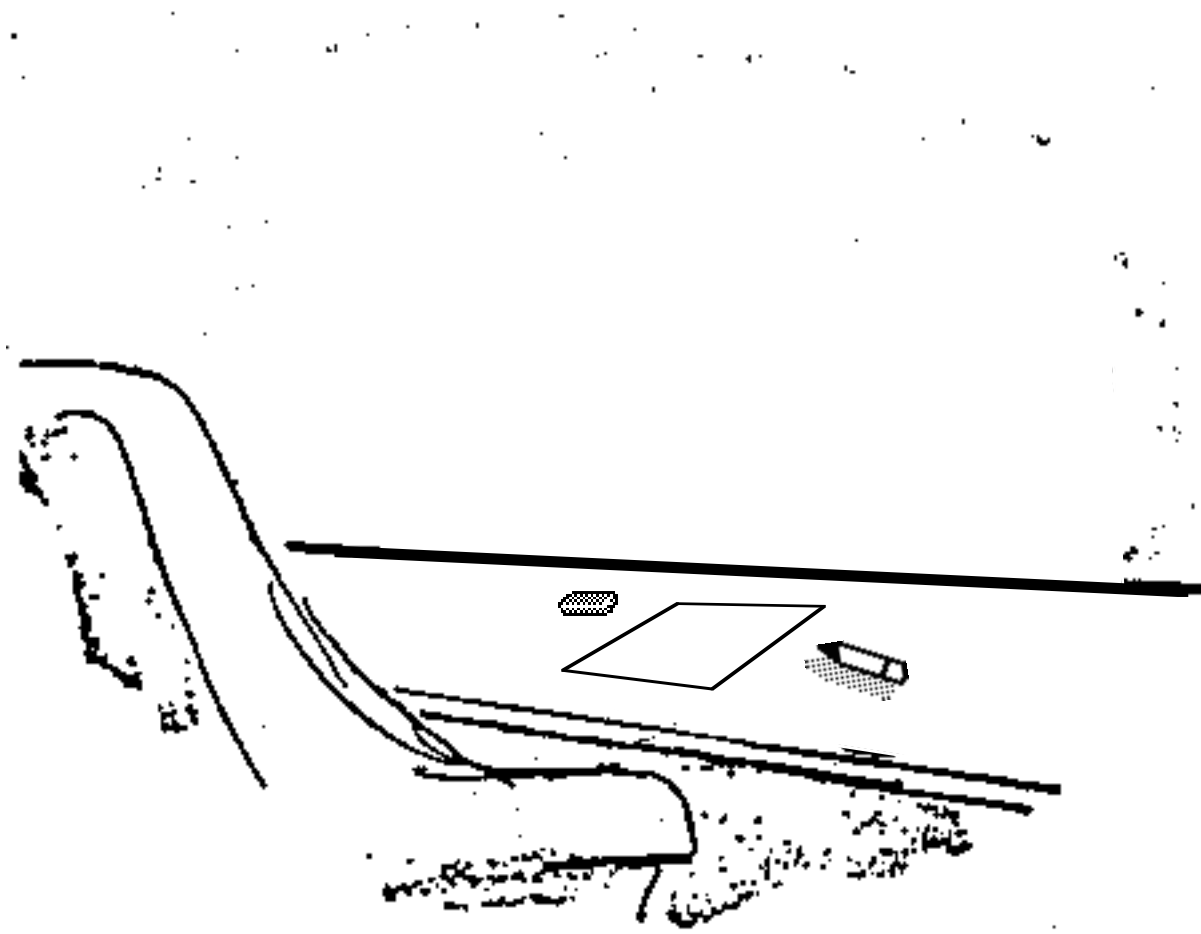
Quadrant-IV: Benutzung



Quadrant-II: Entwurf

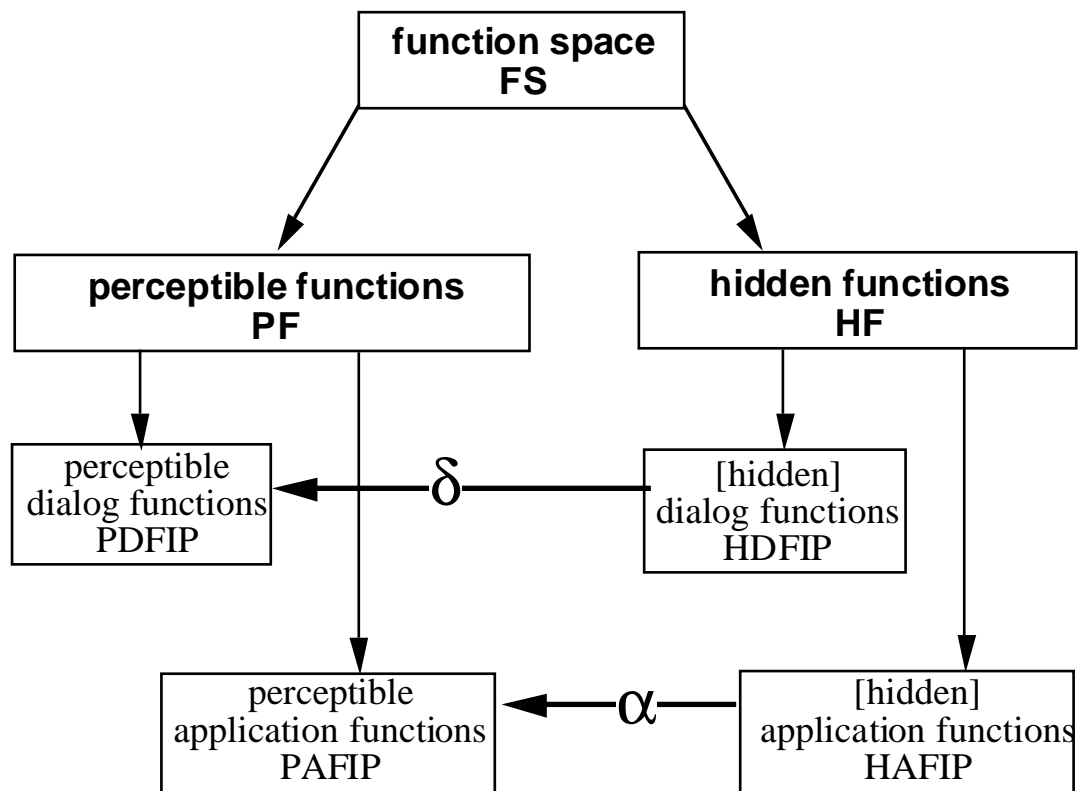
Quadrant-III: Realisierung

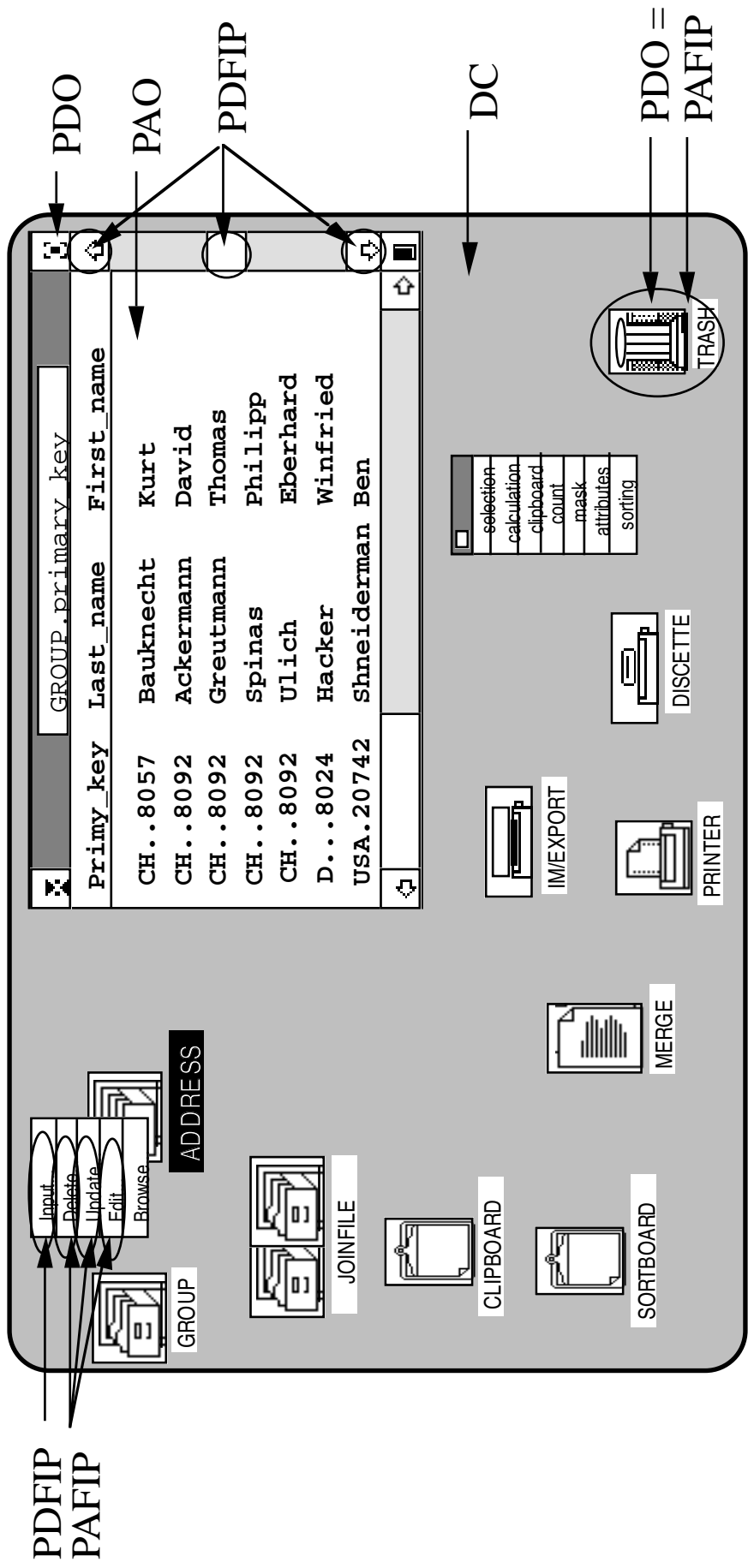
ein formale Gestaltungs-Theorie



- Metriken

An abstract concept to describe usability aspects





quantitative measure of "feedback":

$$FB = 1/D \sum_{d=1}^D (\#PF_d / \#HF_d) * 100\%$$

quantitative measure of "interactive directness":

$$ID = \{1/P \sum_{p=1}^P \min[\ln g(PATH_p)]\}^{-1} * 100\%$$

[visual] feedback (FB)

| | | low | high |
|-----------------------------|------|-------------------------------|---|
| interactive directness (ID) | low | batch | menu interface MI |
| | high | command language CI | desktop style direct manipulation DI |

The outcomes of nine (9) different comparison studies between command (CI) and menu (MI) interfaces.

"CI < MI" means that the average usage/preference with/for MI is better than with/for CI;

"CI = MI" means that there are no published data to decide;

"CI > MI" means that the average usage/preference with/for CI is better than with/for MI;

"sig." means that $p \leq 0.05$; "not sig." means that $p > 0.05$

| Reference | interface | skill level | usability metric | outcome | result |
|--------------------------|-------------|-------------|----------------------|---------|----------|
| Streitz et al. (1987) | CI, MI | beginner | task solving time | CI < MI | sig. |
| Chin et al. (1988) | CI, MI | beginner | subjective rating | CI < MI | sig. |
| Ogden & Boyle (1982) | CI, MI, HY | beginner | preferences | CI < MI | sig. |
| Roy (1992) | CI, MI | advanced | error rate | CI < MI | sig. |
| Roberts & Moran (1983) | CI, MI, DI | experts | task solving time | CI < MI | sig. |
| Chin et al. (1988) | CI, MI | experts | subjective rating | CI < MI | sig. |
| Peters et al. (1990) | CI, MI, DI | experts | slips | CI < MI | sig. |
| Peters et al. (1990) | CI, MI, DI | experts | recognition errors | CI < MI | sig. |
| Peters et al. (1990) | CI, MI, DI | experts | efficiency | CI < MI | sig. |
| Ogden & Boyle (1982) | CI, MI, HY | beginner | task time | CI < MI | not sig. |
| Roy (1992) | CI, MI | advanced | task solving time | CI < MI | not sig. |
| Antin (1988) | CI, MI, KMI | advanced | subjective rating | CI < MI | not sig. |
| Hauptmann & Green (1983) | CI, MI, NO | beginner | task solving time | CI = MI | not sig. |
| Hauptmann & Green (1983) | CI, MI, NO | beginner | number of errors | CI = MI | not sig. |
| Hauptmann & Green (1983) | CI, MI, NO | beginner | subjective rating | CI = MI | not sig. |
| Whiteside et al. (1985) | CI, MI, IO | beginner | task completion rate | CI > MI | not sig. |
| Antin (1988) | CI, MI, KMI | advanced | preferences | CI > MI | not sig. |
| Roberts & Moran (1983) | CI, MI, DI | experts | error-free task time | CI > MI | not sig. |
| Whiteside et al. (1985) | CI, MI, IO | advanced | task completion rate | CI > MI | sig. |
| Streitz et al. (1987) | CI, MI | advanced | task solving time | CI > MI | sig. |
| Antin (1988) | CI, MI, KMI | advanced | task completion rate | CI > MI | sig. |
| Whiteside et al. (1985) | CI, MI, IO | experts | task completion rate | CI > MI | sig. |

The outcomes of twelve (12) different comparison studies between command (CI) and direct manipulative (DI) interfaces.

"CI < DI" means that the average usage/preference with/for DI is better than with/for CI;

"CI = DI" means that there are no published data to decide;

"CI > DI" means that the average usage/preference with/for CI is better than with/for DI;

"sig." means that $p \leq 0.05$; "not sig." means that $p > 0.05$

| Reference | interface | skill level | usability metric | outcome | result |
|------------------------------|------------|-------------|-----------------------|---------|----------|
| Altmann (1987) | CI, DI | beginner | task solving time | CI < DI | sig. |
| Karat et al. (1987) | CI, DI | beginner | task solving time | CI < DI | sig. |
| Streitz et al. (1989) | CI, DI | beginner | task solving time | CI < DI | sig. |
| Sengupta & Te'eni (1991) | CI, DI | beginner | task solving time | CI < DI | sig. |
| Margono et al. (1987) | CI, DI | beginner | number of errors | CI < DI | sig. |
| Morgan et al. (1991) | CI, DI | beginner | number of errors | CI < DI | sig. |
| Morgan et al. (1991) | CI, DI | beginner | time between errors | CI < DI | sig. |
| Karat et al. (1987) | CI, DI | beginner | error correction time | CI < DI | sig. |
| Morgan et al. (1991) | CI, DI | beginner | error-free time | CI < DI | sig. |
| Margono et al. (1987) | CI, DI | beginner | subjective rating | CI < DI | sig. |
| Morgan et al. (1991) | CI, DI | beginner | subjective rating | CI < DI | sig. |
| Torres-Chazaro et al. (1992) | CI, DI | beginner | subjective rating | CI < DI | sig. |
| Sengupta & Te'eni (1991) | CI, DI | beginner | efficient usage | CI < DI | sig. |
| Tombaugh et al. (1989) | CI, DI | advanced | subjective rating | CI < DI | sig. |
| Torres-Chazaro et al. (1992) | CI, DI | advanced | subjective rating | CI < DI | sig. |
| Roberts & Moran (1983) | CI, MI, DI | experts | task solving time | CI < DI | sig. |
| Peters et al. (1990) | CI, MI, DI | experts | oblivion's errors | CI < DI | sig. |
| Peters et al. (1990) | CI, MI, DI | experts | recognition error | CI < DI | sig. |
| Peters et al. (1990) | CI, MI, DI | experts | efficiency | CI < DI | sig. |
| Margono et al. (1987) | CI, DI | beginner | task solving time | CI < DI | not sig. |
| Morgan et al. (1991) | CI, DI | beginner | task solving time | CI < DI | not sig. |
| Tombaugh et al. (1989) | CI, DI | advanced | task solving time | CI < DI | not sig. |
| Roberts & Moran (1983) | CI, MI, DI | experts | error correction time | CI < DI | not sig. |
| Altmann (1987) | CI, DI | beginner | subjective rating | CI > DI | not sig. |
| Masson et al. (1988) | CI, DI | advanced | task solving time | CI > DI | sig. |

Contingency tables of a meta-analysis for all data

[Cell Content: observed frequency (expected frequency)]

| | MI | DI | outcome of this meta-analysis |
|--------------|-----------|-----------|----------------------------------|
| CI better as | 7 (3.9) | 2 (5.1) | Chi** = 5.52, df = 1 p ≤ .019 |
| CI worse as | 12 (15.1) | 23 (19.9) | |

| | beginner | advanced+ | outcome of this meta-analysis |
|--------------------|-----------|-----------|----------------------------------|
| CI better as MI,DI | 2 (4.3) | 7 (4.7) | Chi** = 2.95, df = 1 p ≤ .086 |
| CI worse as MI,DI | 19 (16.7) | 16 (18.3) | |

Contingency tables only for significant differences

(SELECTION for "result" = "sig.").

[Cell Content: observed frequency (expected frequency)]

| | MI | DI | outcome of this meta-analysis |
|--------------|----------|-----------|----------------------------------|
| CI better as | 4 (2.0) | 1 (3.0) | Chi** = 4.07, df = 1 p ≤ .044 |
| CI worse as | 9 (11.0) | 19 (17.0) | |

| | beginner | advanced+ | outcome of this meta-analysis |
|--------------------|-----------|-----------|----------------------------------|
| CI better as MI,DI | 0 (2.4) | 5 (2.6) | Chi** = 5.55, df = 1 p ≤ .018 |
| CI worse as MI,DI | 16 (13.6) | 12 (14.4) | |

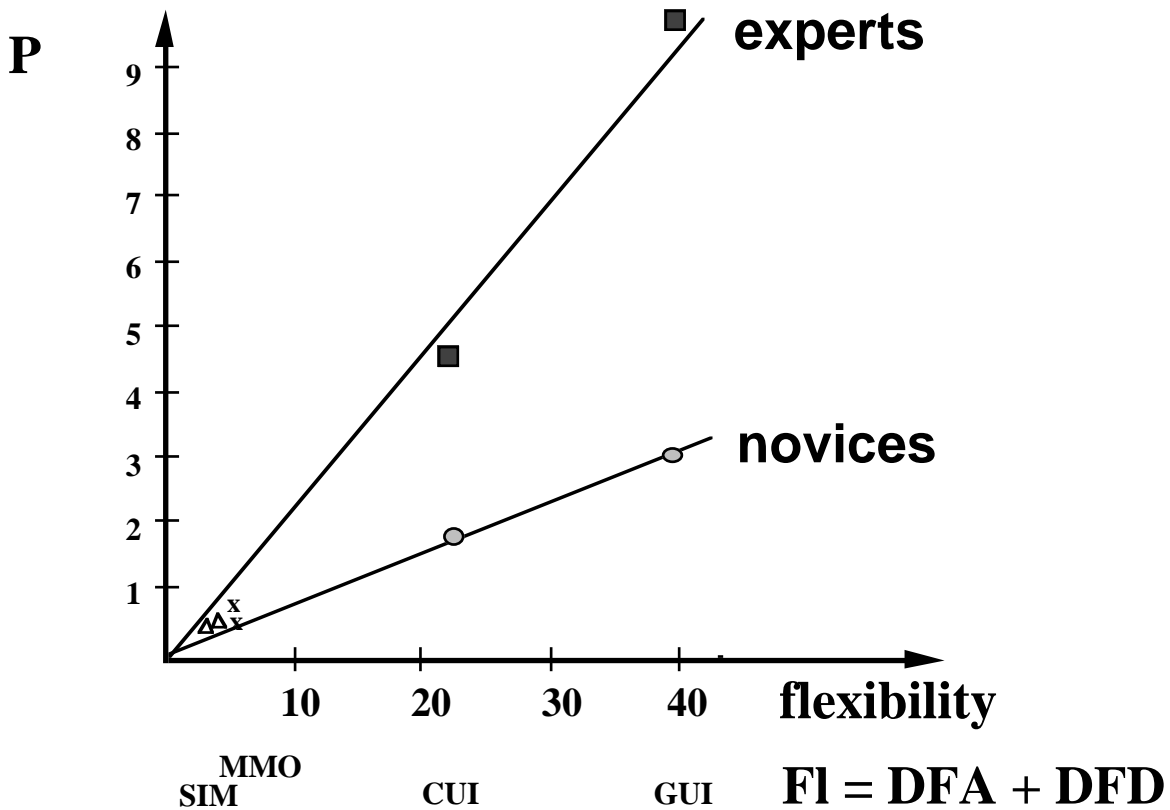
quantitative measure of "application flexibility":

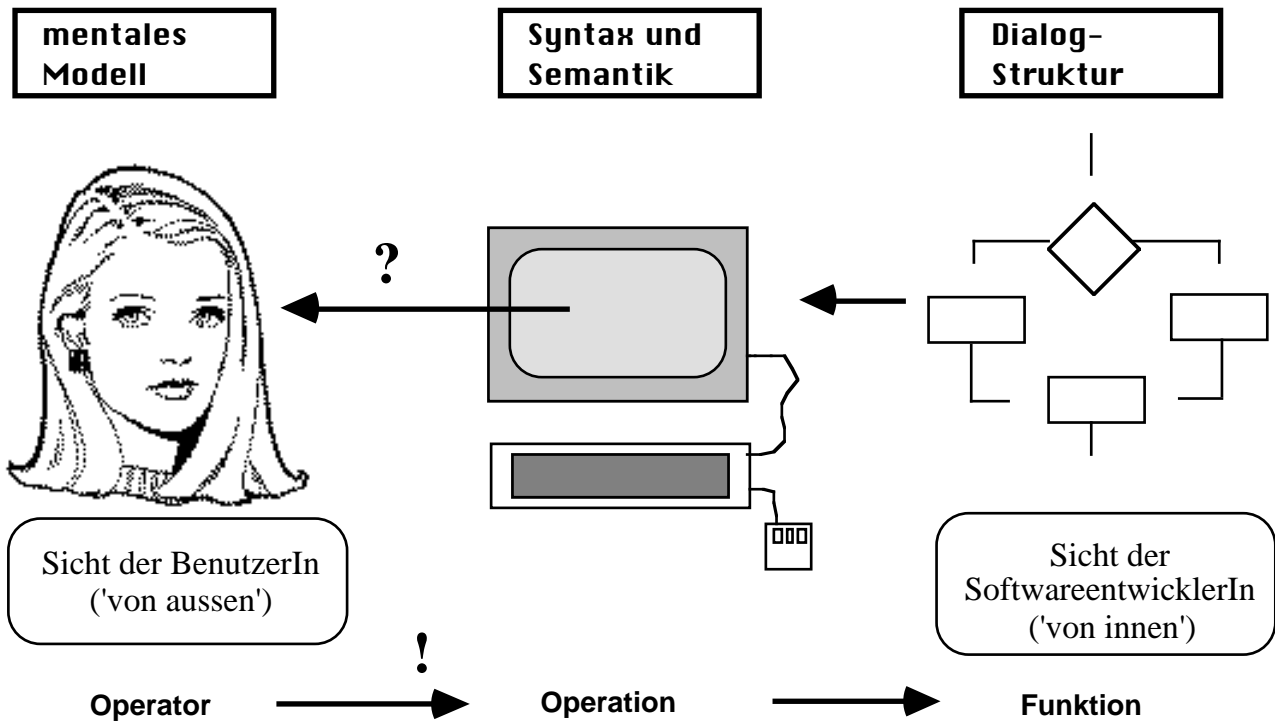
$$DFA = 1/K \sum_{d=1}^K \#HAFIP_d$$

quantitative measure of "dialog flexibility":

$$DFD = 1/K \sum_{d=1}^K \#HDFIP_d$$

performance







bmb+f
Bundesministerium für
Bildung, Wissenschaft,
Forschung und Technologie

**Bundesministerium für Bildung,
Wissenschaft,
Forschung und Technologie**

Zukunftsorientiert handeln!

Die EU-Bildschirmrichtlinie in der Praxis

Ergebnisse aus dem SANUS-Projekt



2. SANUS - Kongreß in Bad Honnef
23. und 24. Oktober 1997

- Tagungsunterlagen -

Herausgeber: Prof. Dr.Ing. Peter Kern
© by Institut für Arbeitswissenschaft und Technologiemanagement IAT,
Universität Stuttgart 1997, ISBN 3-8167-5143-1
