

Development Cycles for Usability Testing in Augmented Reality

Marcus Tönnis

<toennis@in.tum.de>

with very special thanks to Björn Schwerdtfeger (TUM) and Margarita Anastassova (LIST)

Fachgebiet Augmented Reality
Technische Universität München, Germany

Decision Helper Silde – 3 hours in 2 minutes

This tutorial covers 3 selected topics

- Session 1: **Utility and Usability**
 - Definitions
 - Current Methods
 - Issues & Challenges
- Session 2: **Usability, Utility Engineering**
 - Formative, summative and psychological experiments
 - Some experiences with the ‘general’ process
 - Recommendation for an adopted process
- Session 3: **Confounding Factors**
 - Presentation Principles
 - Combinations

What it does not cover

- How descriptive statistics is calculated

Reality

- “Each time a new technology comes along, new designers make the same horrible mistakes as their predecessors. Technologists are not noted for learning from the errors of the past. They look forward, not behind, so they repeat the same problems over and over again”
- “The most egregious failures always come from the developers of the most recent technologies.”

DONALD A. NORMAN,
THE DESIGN OF EVERYDAY THINGS



Motivation I



The Past

We work at the **limit of what is known or common practice** in the design of user interfaces

- Huge and mostly unclassified design space
- Technologies in search for applications
- Scarce user experience and knowledge about ‘our’ technologies
- Limited number of applications used in real industrial settings
- Limitation of HCI knowledge and guidelines for design and evaluation

Motivation II

- This tutorial presents **our experiences** in **designing** and **evaluating** industrial AR applications and gives insights into methods to improve:
 - **Utility**
 - **Usability**
 - **The design process itself**



The Future?

Outline

- ➔ • **Session 1: Current Methods, Issues & Challenges**
 - Utility
 - Methods for elicitation of user needs
 - Advantages & limitations
 - Usability
- **Session 2: Usability, Utility Engineering**
 - Formative, summative and psychological experiments
 - Some experiences with processes
- **Session 3: Confounding Factors**
 - Presentation Principles
 - Combinations

Utility



Utility - Definition

- Utility refers to the functions of a system:
Does it do what users need? (Nielsen, 1994)
- AR
 - Technology searching for applications, but
 - We do not always know what the user needs
- **Utility: Key feature for AR applications**



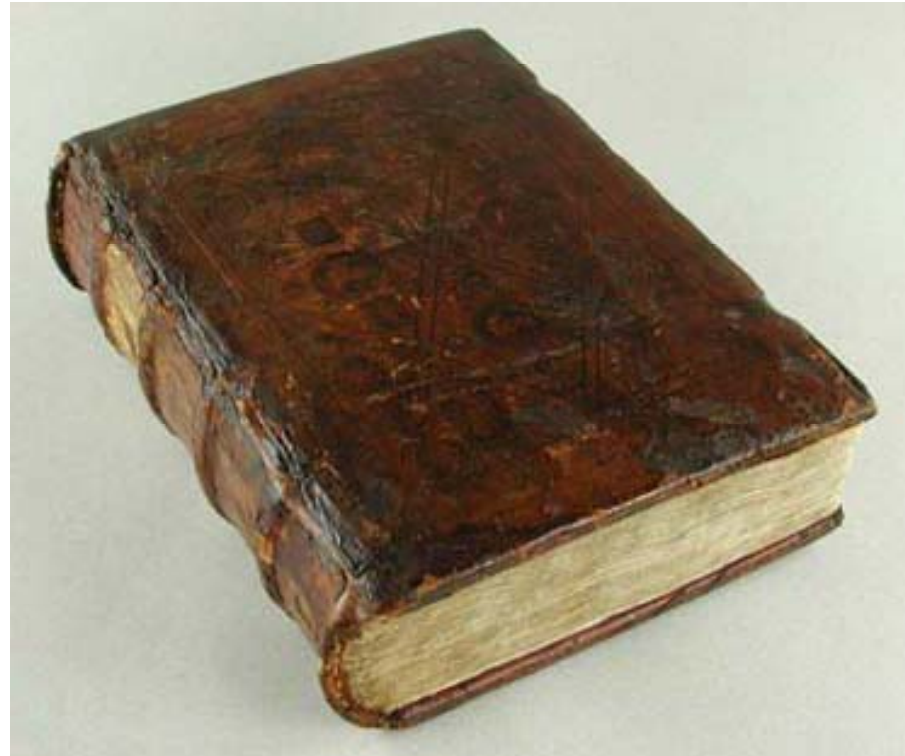
Useful: Provides a lot of useful features – Usable?

User Needs - Requirements Elicitation

- **Analysis of user characteristics and the tasks in order to increase utility**

- But how?

Option: State-of-the-Art Reviews and Surveys



Option: State-of-the-Art Reviews and Surveys

- Information on **current** and **potential AR applications**
- Extraction of **potentially useful functions** from existing systems
- **Feedback** about **user satisfaction** from existing systems
- **Identification** of application **pitfalls** and **errors**
- **Few existing reviews** (design criteria, guidelines and heuristics)
- **Knowledge scattered** in different sources
- **Limited empirical validation** of assumptions (few real AR applications)
- **Rather general** findings and conclusions
- Paradox of emerging technologies: great **need** for **HCI expertise** vs. **limited** available **HCI knowledge**

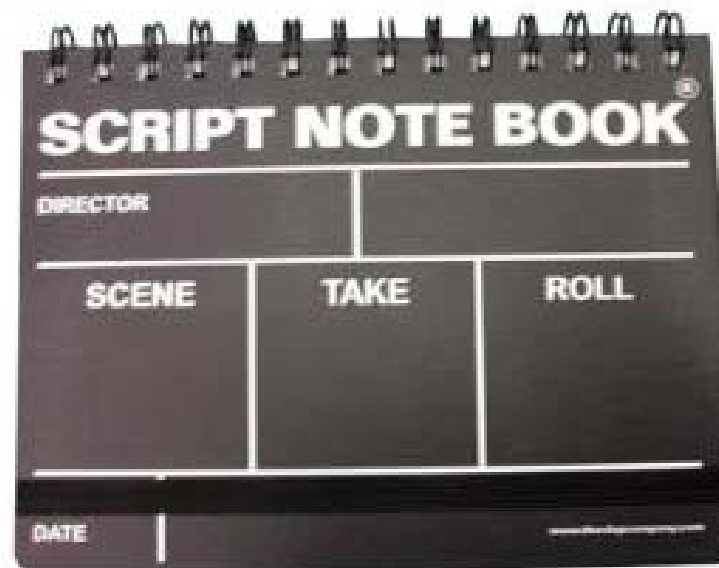
Option: Focus groups and design workshops



Option: Focus groups and design workshops

- **Rapid and inexpensive** way to collect ideas for **possible** applications
- Often invokes **creativity**
- Facilitates **communication** between **professionals** and **users**
- **Limited or no user knowledge, or no experience with AR interfaces**
- Discussions may be **unstructured** and **important topics** may be **missed**
- **People** with **strong opinions** may **dominate** and may lead the discussion
- **Not all participants feel comfortable** in the group or during a design workshop

Option: Scenarios, Storytelling and Narratives



Option: Scenarios, Storytelling and Narratives

- **Engaging way of focusing on user goals, desires and limitations**
- **Give stakeholders a specific idea of AR and intended application**
- **Facilitate**
 - **discussions about functions of the system**
 - **generation of „undreamed-of“ requirements (Robertson, 2001)**
- **May provide incomplete information about user tasks because AR is in search of applications**
- **May be too synthetic because future users and their characteristics are barely known**
- **May inhibit stakeholders' imagination**
- **May predominantly express author's point of view**

Option: Interviews and Field Observations



Option: Interviews and Field Observations

- The **most traditional technique** to trawl for **user needs**
- May **provide** very useful **information** about user **wishes, difficulties, context**
- Facilitate the **inclusion** of **future users** in design
- **Particularly useful** in the **beginning** of the **design process**
- Interviews usually **result** in **conscious requirements** only
- **Field studies** in **context** implying **complex activities** can be **very difficult** in the beginning
- The **amount** of **raw data** can be **overwhelming**
- The **time** necessary for its **analysis can be long**, but is **important** (especially for data from field studies)

Option: User Evaluations of Mock-Ups and Prototypes



Option: User Evaluations of Mock-Ups and Prototypes

- **Physical representation** of the **future technology**
- **Functions** and **different interactional levels** can be **tested**
- **Facilitate** the **identification** of design **errors** and usability **pitfalls**
- **Heterogenous, unstable** and **immature** prototypes
- User **habits** of using **traditional systems** may **negatively influence user acceptance** and the **interaction** with the prototype
- **Risk** for the user to **concentrate** on **usability problems** instead on the generation of **requirements**
- **Risk** of **frustration** (user; stakeholders)
- May inhibit stakeholders' **imagination**
- **Use** of evaluation **guidelines** for more **traditional** technologies

Usability



Usability - Definition

- A measurable characteristic of a product user interface that is present to a greater or lesser degree...(Mayhew, 1999; Nielsen, 1994)
 - **Learnability:** easiness to accomplish basic tasks (novice users)
 - **Efficiency:** task completion time
 - **Memorability**
 - **Errors:** number, impact on the result, easy recovery
 - **Satisfaction & Hedonic Quality**
 - **Physical, Psychological Strain**
- „If your computer was a person, how long ‘till you punch it in the face’?“

Usability - Testing

- ... is often the only time developers see users
 - **Using** their interface
 - **Fail** with their interface
- Methods used
 - Heuristic evaluation
 - Thinking-aloud testing
 - Usability and cognitive walkthroughs
 - User interface inspections
 - User Tests

Usability - Testing – Problems I

- Only HCI Experts **no AR experts** -> no complete and reliable heuristic evaluation possible
- **Laboratory conditions** may be **distort** the **situation** so much that the conclusions may have no application \Leftrightarrow **Field studies** expensive
- Typically **short term usage** and artificial tasks
- **Usability** is the **dominant research** orientation in AR, while **utility** is often **overlooked!**
- **Anecdotal evidence** or individual insights may be given **too little emphasis** because of the **authoritative influence** of **inferential statistics**

Usability - Testing – Problems II

- Very few experienced **AR Users....**
- **Technical newbies need ages** and have a **long learning curve**
 - They really need long introduction phases (sometimes longer as the actual test is)
 - People with prejudices need even longer
- **Difficult**, expensive to find **representative future users**
- **People being afraid of new technologies**, often give a **negative feedback**, being afraid of having to work with such new stuff in the future

Usability - Testing – Problems III

- People with **visual problems** (without knowing it and not wearing glasses)
 - have naturally problems with HMDs and **are slower**
- OST-HMDs produce **varying brightness** with the same system
 - depending on people
 - varying sun light
- **Not everybody can see stereo**
- At **what distance** do you place the focus of the HMD?
 - Does **not really affect 3D** visualization
 - Problem if you have to **compare 2D texts** (real/virtual)
 - Problem if you have to **walk in a factory hall**

End of Session 1

- Questions?

Outline

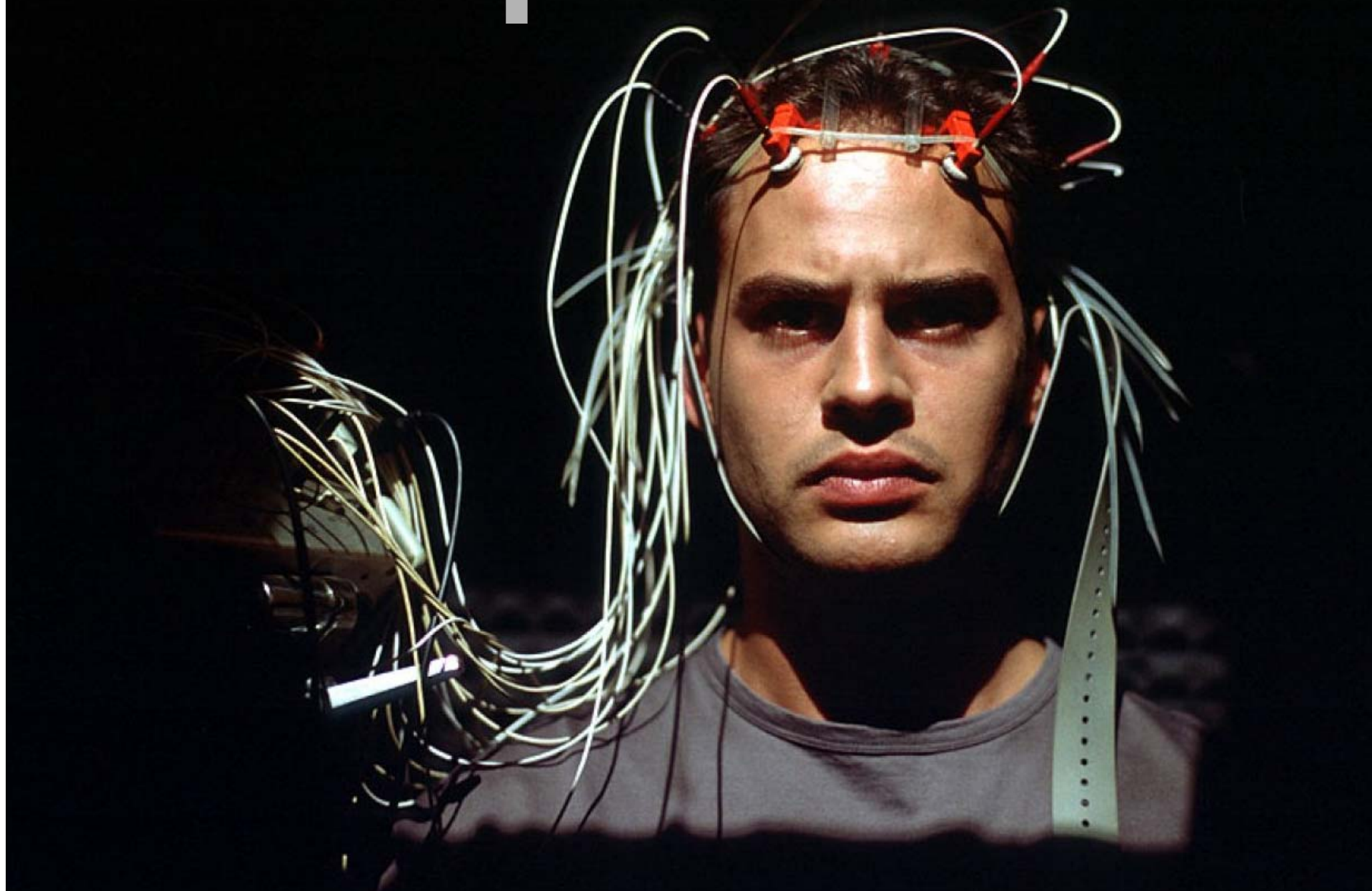
- **Session 1: Current Methods, Issues & Challenges**
 - Utility
 - Methods for elicitation of user needs
 - Advantages & limitations
 - Usability
- ➔ • **Session 2: Usability, Utility Engineering**
 - Formative, summative and psychological experiments
 - Some experiences with processes
- **Session 3: Confounding Factors**
 - Presentation Principles
 - Combinations

Motivation

Why and how do we evaluate our applications in **research**?

- We develop something (leisure, PhD topic, ...)
 - We need to publish about it, and user studies are requested
 - ... so we develop an experiment to test our system against something else in an experiment...
-
- But be aware: the AR Prototype will not outperform an established mature technology

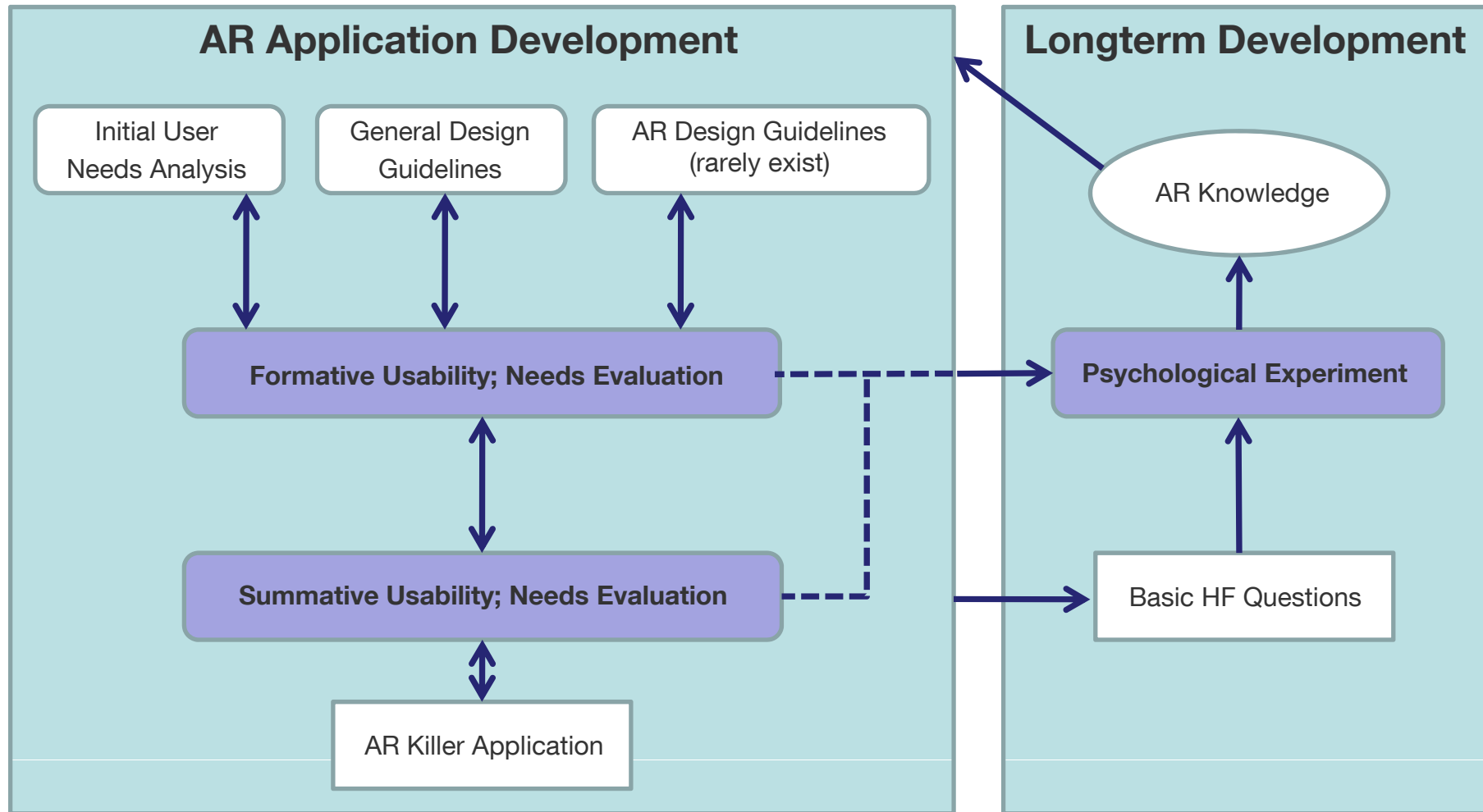
The Experiments



Types of Experiments

- Formative Experiments
- Summative Experiments
- Psychological Experiments (often called Usability Studies)

General Standard Process



Formative Usability Evaluation

- Objective
 - Iteratively „shape‘ the interface -> impact on design
 - Refine the specifications (if they exist...)
 - Eliminate usability problem
- A few users (3-6), **no** statistical significance
- Generates
 - Qualitative Data for the Redesign
 - Rarely: Quantitative Data for benchmarking against usability criteria
- not necessarily informal...



Summative Usability; Needs Evaluation

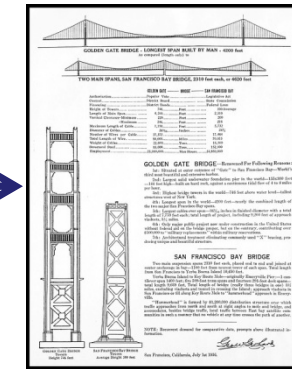
- Is a **comparative** study
- Is usually done at the end of development – **for the stakeholder**
- The objective is to judge the design against quantitative goals or competitive products
- Larger user samples
- Results are quite formal
- **Metrics:** Qualitative and quantitative



Competitor



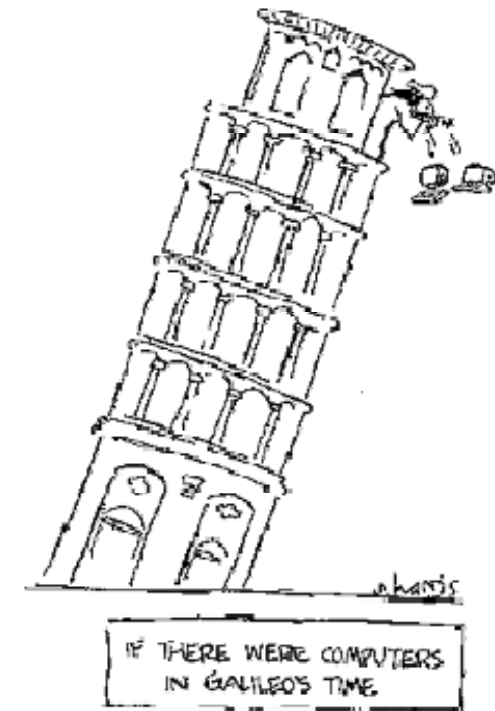
Your product



Usability specification

Psychological Experiment

- Find out about the basics of human factors in HCI
- **Strictly controlled experiment**
 - **One ,thing‘** to be tested, relating to the **human**
 - Executed in a **controlled environment** (exactly one independent variable)
 - **Many users** (really many users – valid results)
 - Intensive **Data Analysis**
 - Intensive **Hypothesis-Testing**
- Leads to **narrow**, but **reliable** results
- **Knowledge** of the **AR design space** is **far from being rigorous** - > a lot of unknown **confounding** factors

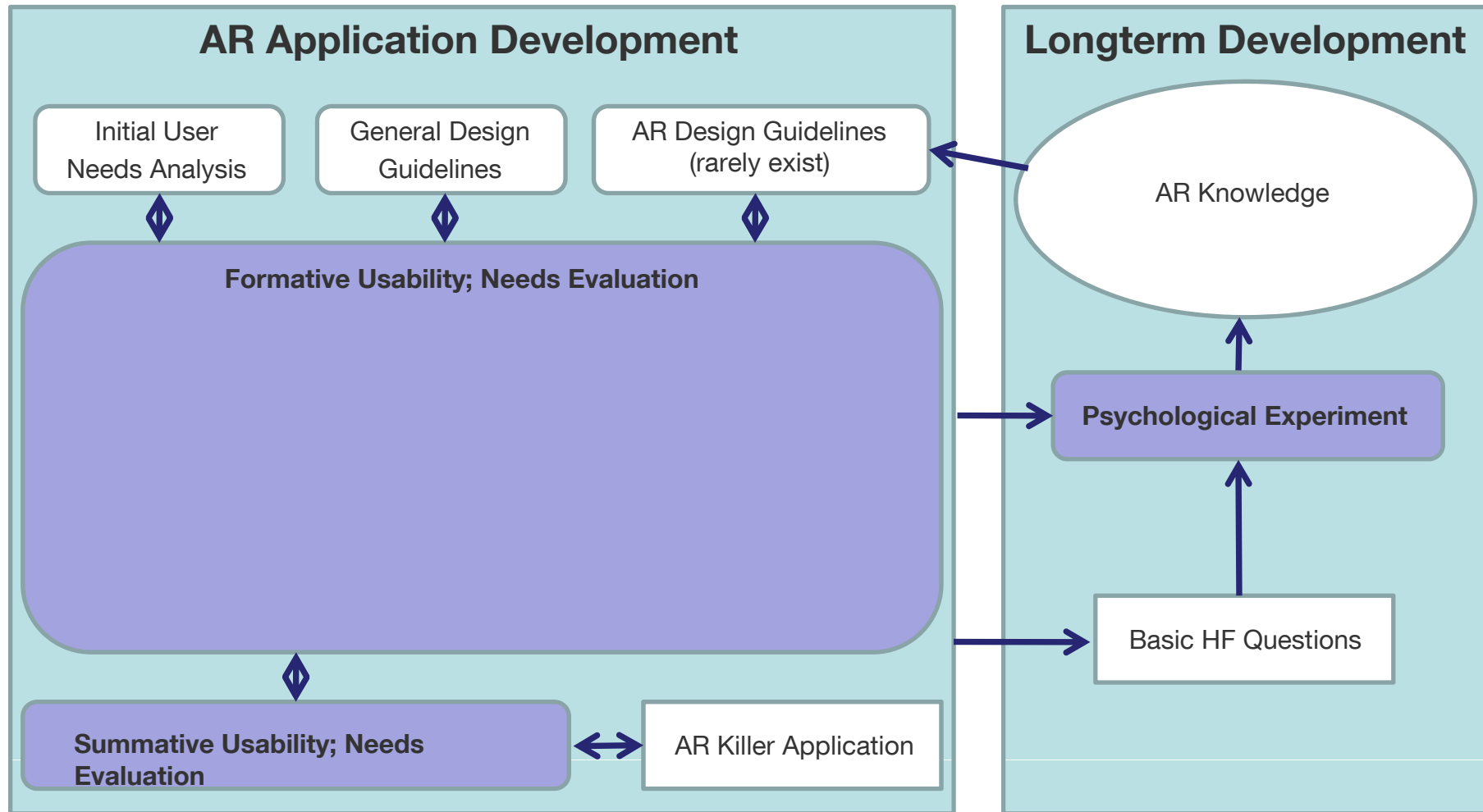


MARS 500

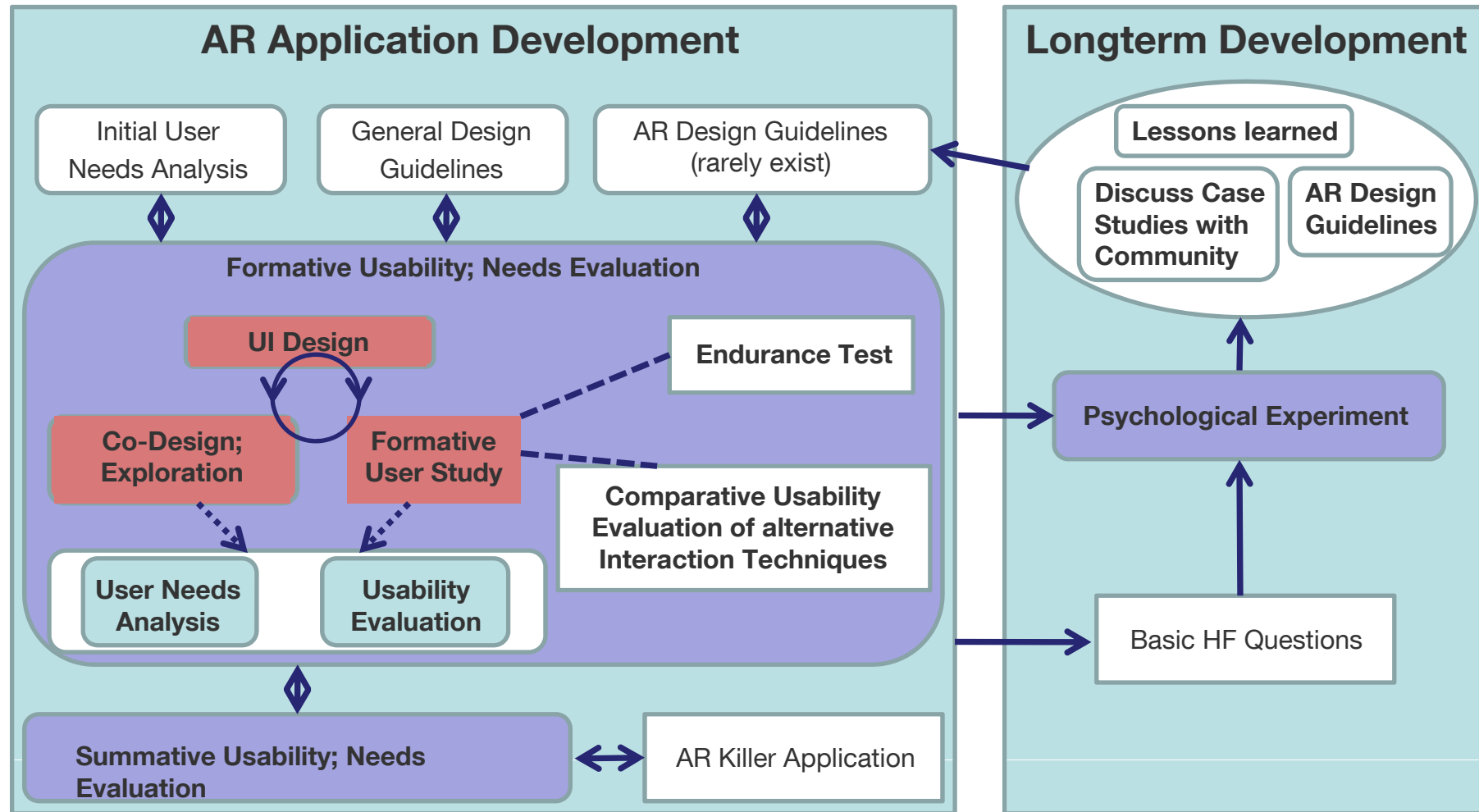
Process



General Standard Process



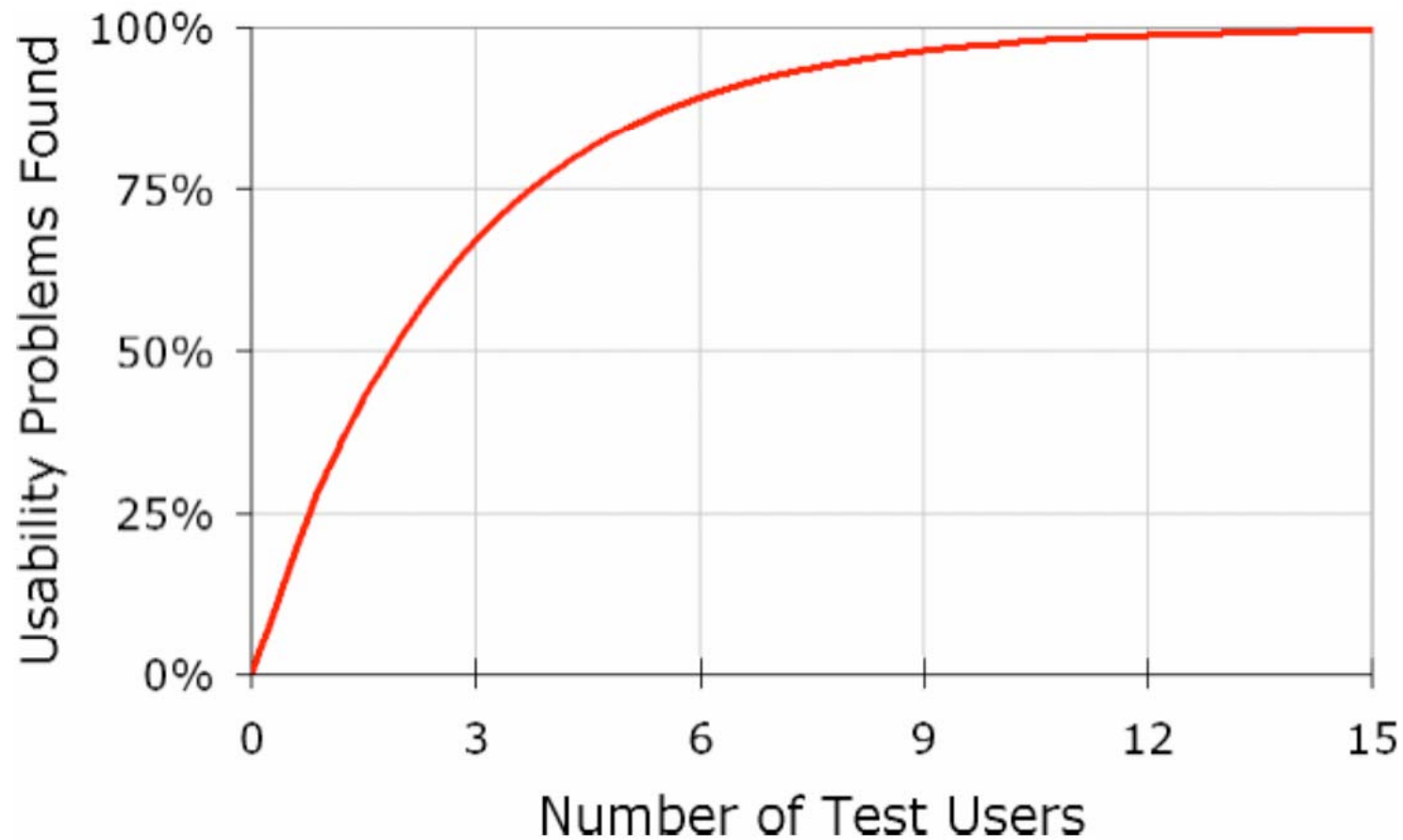
Extended Process for AR Applications



References for Previous Slide

- The Loop
 - J. L. Gabbard, D. Hix, and J. E. Swan. User Centered Design and Evaluation of Virtual Environments. IEEE Computer Graphics and Applications 19(6), 1999
 - J. L. Gabbard and J. E. Swan. Usability Engineering for Augmented Reality : Employing User-based Studies to Inform Design. IEEE Transactions on Visualisation and Computer Graphics, 14, 2008
- Low level testbed evaluation
 - D. Bowman and L. Hodges. Formalizing the Design, Evaluation, and Application of Interaction Techniques for Immersive Virtual Environments. The Journal of Visual Languages and Computing, 1999
- Comparing "good apples" to "bad oranges"
 - D. Hix, J. E. Swan, J. L. Gabbard, M. McGee, J. Durbin, and T. King. Usercentered design and evaluation of a real-time battleeld visualization virtual environment. In Proceedings of the IEEE VR'99 Conference, 1999
- The refactored development cycle
 - B. Schwerdtfeger, Pick-by-Vision: Bringing HMD-based Augmented Reality into the Warehouse, Dissertation, Technische Universität München, 2010: <http://mediatum2.ub.tum.de/node?id=992985>

Formative Evaluation - Number of Test Participants?



J. Nielsen, 1993

Kinds of feedback you get...

Depends on the subject



- **Expert/Scientist:**
 - has a lot of really **good feedback**
 - take care, he **conquers/ ignores** a lot of **basic usability problems** – observe **how** he does it
 - tells you **countless (mostly useless) features**, you should implement



"I know nothing about the subject, but I'm happy to give you my expert opinion."

- **Domain Expert**
 - gives you **domain specific input** of high value
 - good for **user needs elicitation**
 - ask him **which** of the **features**, the scientist told you, are **worth** to be **implemented**
 - can be quite creative, what else you could do with your technology -> **killer apps**



- **Housekeeper** (someone inexperienced)
 - - **falls into traps** and you have to figure out when and where it happens
 - - helps you to **find basic usability problems**

Co-Design; Exploration

- Objective: **Cooperative-Exploration to reduce the design space**
- Have an **assistant who can change design parameters quickly, while you are** discussing with the user
- **1st Phase: you are a partner of the test user** (under different conditions)
- **2nd Phase: take a step back, let people do real task and observe**

Formative User Study I

- **Objective: Get deeper insides into the results from**
- **1. phase: Let people play and ask**
- **2. phase**
 - **Case 1: Stress test: 2x3 Subjects over a longer period**
 - **Case 2: Comparative experiment**
 - Few more subjects (10-20)
 - Ordered experimental design; with-in subject, permuted order; informal hypothesis

Formative User Study II

- Observation Strategy: The **Muppet Show Balcony Experience** (ca. 30 min)
 - Focus on **User Strategies**
 - Try to figure out all **sub steps**
 - What the user is **currently doing or thinking**
 - What of this seems **strange** or to be done in an **inefficient way**
 - **Interrupt** them and **ask**, why they **behave in a different way**
- Interview them specifically



Summary

- Just try it out and observe people
- Start small
- Let people do real tasks!!!
- Have different people participate!

- Test and shape prototypes iteratively using formative evaluations during open sessions (e.g. design workshops).
- Most iterations are for you: to get more knowledge (AR, user studies, the domain, utility and usability)
- AR Prototypes will not ad-hoc outperform established mature systems (take into account limitations of prototype)

- Tell us about your experiences!

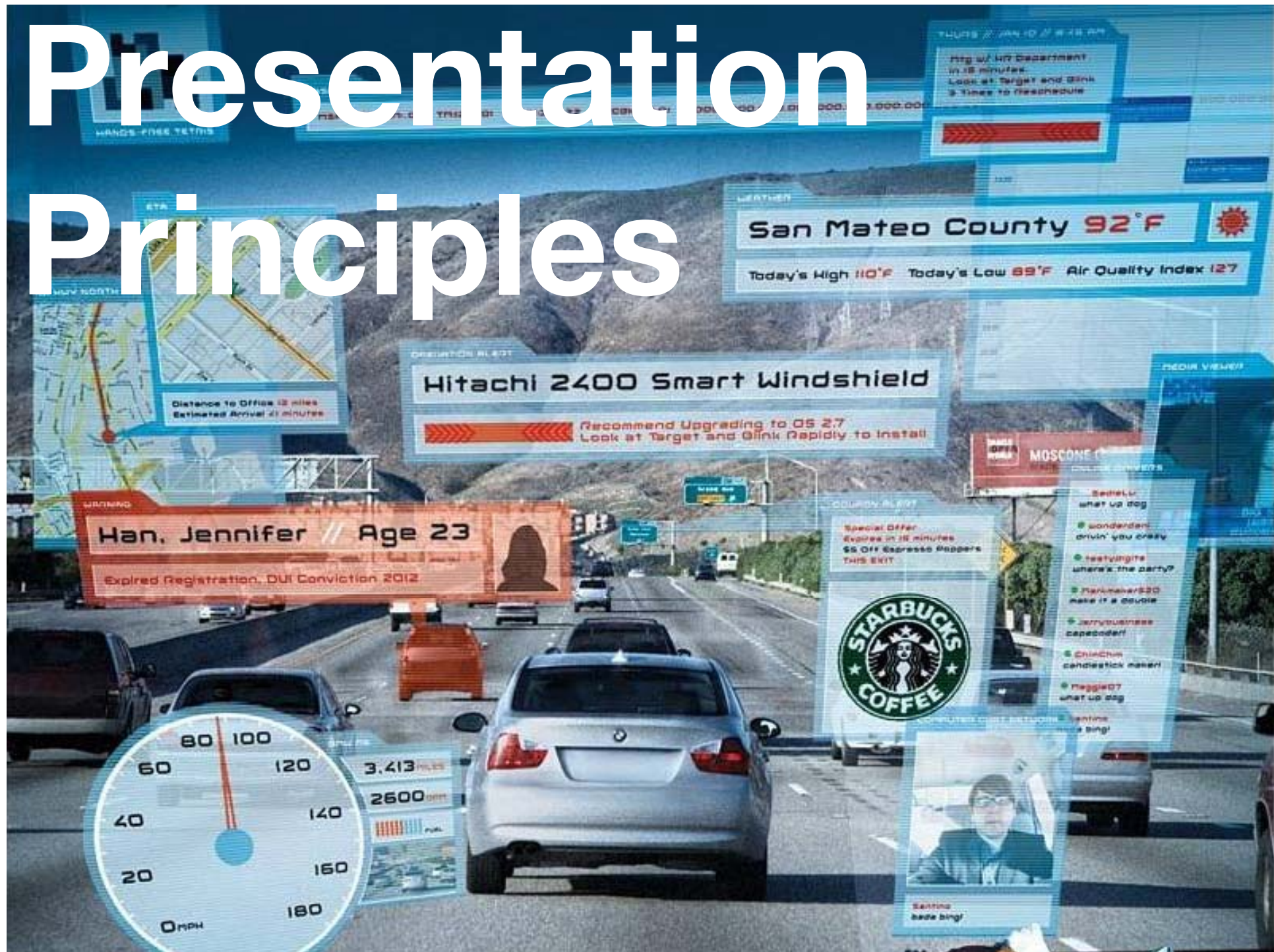
End of Session 2

- Questions?

Outline

- **Session 1: Current Methods, Issues & Challenges**
 - Utility
 - Methods for elicitation of user needs
 - Advantages & limitations
 - Usability
- **Session 2: Usability, Utility Engineering**
 - Formative, summative and psychological experiments
 - Some experiences with processes
- ➔ • **Session 3: Confounding Factors**
 - Presentation Principles
 - Combinations

Presentation Principles



Motivation

- Various AR systems are already under development and run through user studies
- Problem: *Independent variables*
- Reason: AR visualizations use *multiple principles* of presentation. To clearly attribute measured effects to a specific independent variable only one principle at a time may be changed between two variants
- Issue: Different system variants often have multiple parameters affected
- Awareness: Know about different principles of presentation before you start system and test design

Overview

- 3D space for information presentation
- Classes of dimensions for information presentation
- Design examples and potential cross-relationships of designs
- Conclusion

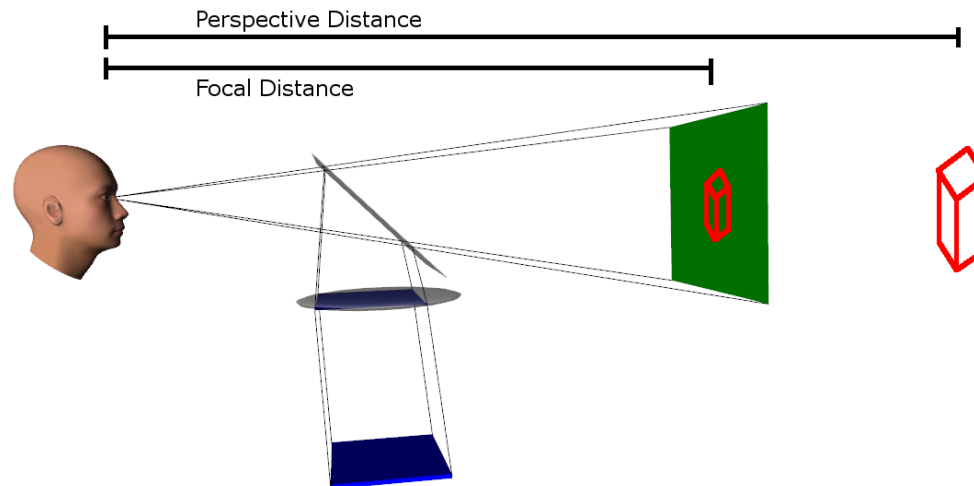
Much content of the subsequent slides is taken from this paper: M. Tönnis, M. Plavšić, G. Klinker, Survey and Classification of Head-Up Display Presentation Principles, The 17th World Congress on Ergonomics (International Ergonomics Association, IEA), Beijing, China, Aug. 9 - 14, 2009.
<http://www.navab.in.tum.de/Chair/PublicationDetail?pub=toennis2009ieaDimensions>

3D Space for Information Presentation

- With AR, information no longer requires stationary displays as carrier – information moves into the surrounding world
- With the paradigm of AR, information has the potential to be presented at the direct place where the origin for the need of information presentation is located
- Instead of 2D on conventional displays, AR extends to 3D

Issues of the 3D Presentation Space

- Information **locally fixed** to the environment **moves** over the display
- Dynamic layouting for avoidance of occlusion of relevant objects
- Focal accommodation – depth queues



Classes of Dimensions for Information Presentation I

- **Continuous vs. Discrete Information Presentation**
 - Continuous information must not be immersive information
 - Discrete information (e.g. warning events) cause driver to leave control circuit of driving task
- **2D Symbolic vs. 3D Information Presentation**
 - 2D symbolic information can use flat icons
 - 3D information renders virtual 3D objects
- **Contact-analog vs. Unregistered Presentation**
 - Information may be registered with the environment (contact-analog)
 - Information may be placed independently of a location in the surrounding

Classes of Dimensions for Information Presentation II

- Presentation in Different Frames of Reference
 - Virtual information can be presented from the driver's point of view, embedding in the perceived scenery
 - Virtual information can also use another frame of reference – e.g. a bird's eye map
- Direct vs. Indirect Referencing of Objects or Situations
 - Direct referencing refers to objects that reside in the drivers field of view
 - Indirect referencing refers to objects that lie occluded in the drivers field of view
 - Pure referencing intends to guide the attention of the driver to a direction outside the field of view

Classes of Dimensions for Information Presentation III

- Location of Presentation in Relation to Glance Direction
 - With glance tracking systems, information can be placed w.r.t. the glance direction of the driver
 - Issues are not to obstruct the view but to keep the information perceivable

Design Examples and potential Cross-relationships of Designs

- Paper illustrates and discusses pair-wise combinations of dimensions

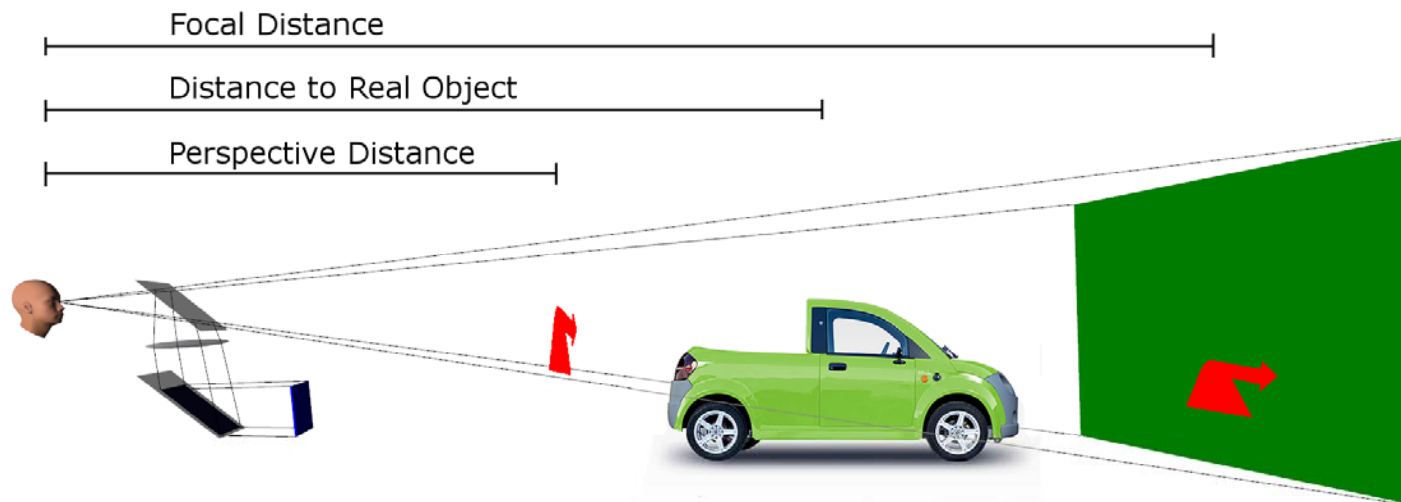
Continuity (1)	-					
Representation (2)	X	-				
Registration in Space (3)	X	X	X			
Frame of Reference (4)	X	X	X	-		
Type of Reference (5)	X	X	X	X	-	
Glance Relation (6)	X	X	X	X	X	-
	1	2	3	4	5	6

- Only marked (⊗) will be illustrated in subsequence – see paper for full survey

• M. Tönnis, M. Plavšić, G. Klinker: Survey and Classification of Head-Up Display Presentation Principles, The 17th World Congress on Ergonomics (International Ergonomics Association, IEA), Beijing, China, Aug. 9 - 14, 2009

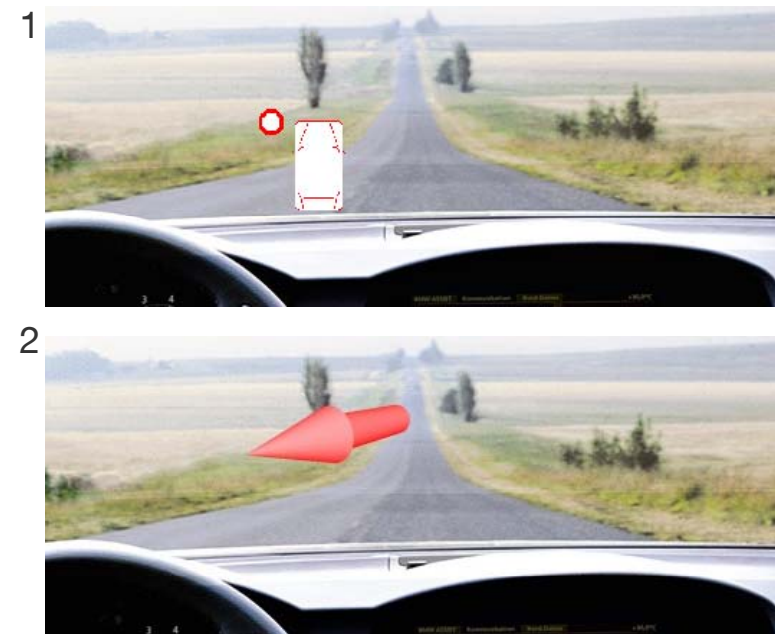
Constraints of Display Technology (3)

- Human eye focuses to the **focal distance** to perceive the image
- Image is rendered in a **perspective distance** shorter than a **real object** (green car).
- => Reverted Depth Cue



Registration in Space vs. Type of Referencing (3 vs 5)

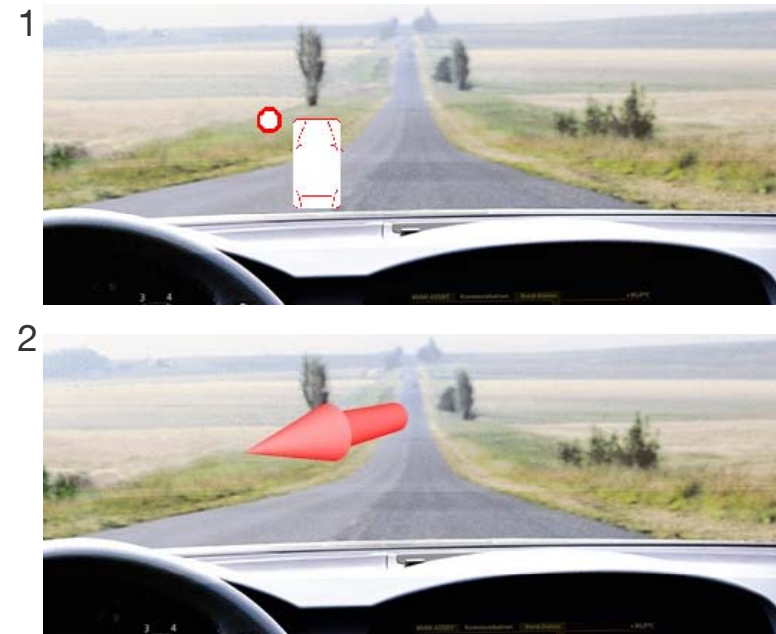
- Example: system for guidance of a car driver's attention
- Different registration in space
 - Bird's eye scheme is unregistered (1)
 - 3D arrow is contact-analog (2)
- Different types of referencing
 - Bird's eye scheme shows location (1)
 - 3D arrow shows direction (2)
- Issues when testing
 - Benefit for pointing to location instead of pointing to a direction? (1)
 - Benefit for information embedded into the world (less need for transformation between frames of reference)? (2)



- Tönnis, M., Sandor, C., Lange, C., Klinker, G., & Bubb, H. (2005, October). Experimental Evaluation of an Augmented Reality Visualization for Directing a Car driver's Attention. In Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR)
- Tönnis, M., & Klinker, G. (2006, October). Effective Control of a Car Drivers Attention for Visual and Acoustic Guidance towards the Direction of Imminent Dangers. In Proc. of International Symposium on Mixed and Augmented Reality (ISMAR)

Registration vs. Frames of Reference (3 vs 4)

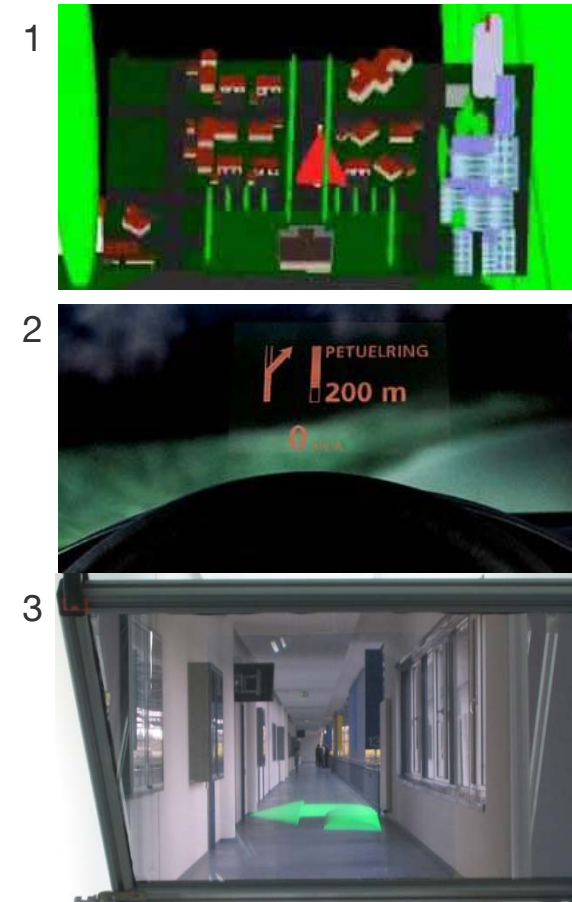
- Example: system for guidance of a car driver's attention
- Different registration in space
 - Bird's eye scheme is unregistered (1)
 - 3D arrow is contact-analog (2)
- Different frames of reference
 - Bird's eye: Transform to coordinate system presentation - gather information - transform back to real world coordinate system – interpret (1)
 - 3D arrow: Embedded as object floating in the world coordinate system (2)



- Tönnis, M., Sandor, C., Lange, C., Klinker, G., & Bubb, H. (2005, October). Experimental Evaluation of an Augmented Reality Visualization for Directing a Car driver's Attention. In Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR)
- Tönnis, M., & Klinker, G. (2006, October). Effective Control of a Car Drivers Attention for Visual and Acoustic Guidance towards the Direction of Imminent Dangers. In Proc. of International Symposium on Mixed and Augmented Reality (ISMAR)

Representation vs. Frame of Reference (2 vs 4)

- Example: Navigation systems
- Different frames of reference
 - North Up: Exocentric (1)
 - Face Up: Exocentric, but motion compensated to egomotion (2)
 - AR presentation: Fully egocentric (in perspective and in motion behavior) (3)
- Varying Representation
 - 2D: Available HUD (2)
 - 3D: In embedded visualization (1) and AR (3)
- To which variation do results of studies attribute to?



• Colquhoun, H., & Milgram, P. (2000). Dynamic Tethering for Enhanced Remote Control and Navigation. In Proceedings of the International Ergonomics Association (IEA), Human Factors Ergonomic Society (HFES) (pp. 146–149)

• Lamb, M., & Hollands, J. G. (2005). Viewpoint Tethering in Complex Terrain Navigation and Awareness. In 49th Annual Meeting of the Human Factors and Ergonomics Society

Registration vs. Glance Behavior (3 vs 6)

- Virtual objects can/could be registered to the glance behavior of the user
- Upcoming issues
 - Direct registration to the line of sight (foveal area of retina) occludes the whole surrounding
 - Adding a static offset to the virtual object disables looking at the virtual object – it always keeps its offset to the line of sight
- Floating algorithms are necessary to establish a relation between an object of concern, its associated information and the dynamic placement if this information

Conclusion

- Spatial AR displays are not yet explored and standardized as conventional 2D displays are
- System development must carefully focus on even small changes to a presentation strategy
- Even minor changes may change the test outcome of a system in comparison to another
- Knowing about presentation principles and possible cross-relationships can avoid misleading results of user studies
- Future work has to investigate these dimensions to reveal foundations for presentation concepts

In other words...

- Next time you are developing some systems and try to put them through a user study, for example
 - **Continuous, 3D** presentation with **contact-analog registration** in space showing **egocentric** information and referring **directly** to the object of concern and in **not glance mounted** manner
 - **Discrete, 2D** presentation **without spatial registration** showing its information in an **exocentric** manner but **indirectly** refers to the object of concern thereby being **glance mounted**
- **Think** if you really want to treat all these principles as **individual** independent variables!

End of Session 3

- Questions?

That's it!

- Thanks for
 - Your patience!
 - Listening!