Methods in AI Research 9
Experimentation:
User studies & system evaluation

Chris Janssen

c.p.janssen@uu.nl
www.cpjanssen.nl

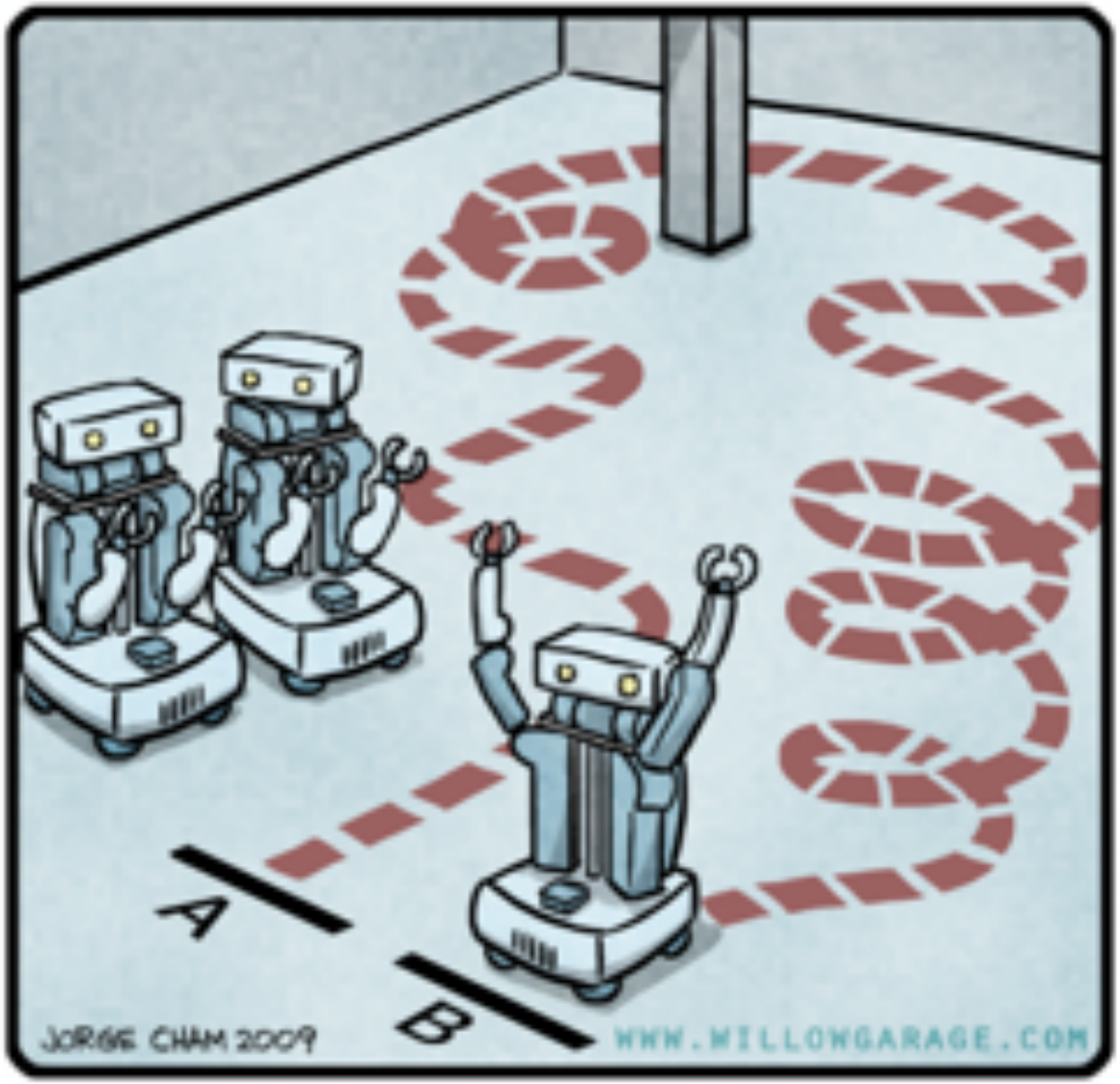
Exam - Of this lecture know:

- All the material that was discussed in class
- Article
  - Cairns, P. (2016) Experimental Methods in Human-Computer Interaction. In Soedergaard, Dam (Eds.) *The encyclopedia of Human-Computer Interaction* (2nd edition). Online available at: https://www.interaction-design.org/literature/
  - If you replace "HCI" with "AI systems" throughout the paper, most points still hold!
  - Other articles that I refer to are examples and NOT part of exam
- Slide appendix:
  - Study questions to give rough guideline of most important aspects to note when reading the article. For exam: know these, so you can *apply* knowledge to case studies
  - Example exam questions
  - Additional material to help in studies
- ***These topics are new to many of you:*** Start reading and studying ***early***

So far…

- Argumentation and dialogue systems
- Models:
  Cognitive, Machine learning, Language, Logic
- Assignment: intelligent system development

R.O.B.O.T. Comics



"HIS PATH-PLANNING MAY BE SUB-OPTIMAL, BUT IT'S GOT FLAIR."

How to….

… generate a correct (research) question?

… collect data?

… analyze data?

… draw correct conclusions?

… formalize this and make predictions?

… communicate this to others?

Cairns (2016): science is about

- **Theories**

- **Experiments** (incl testing / trying things out)

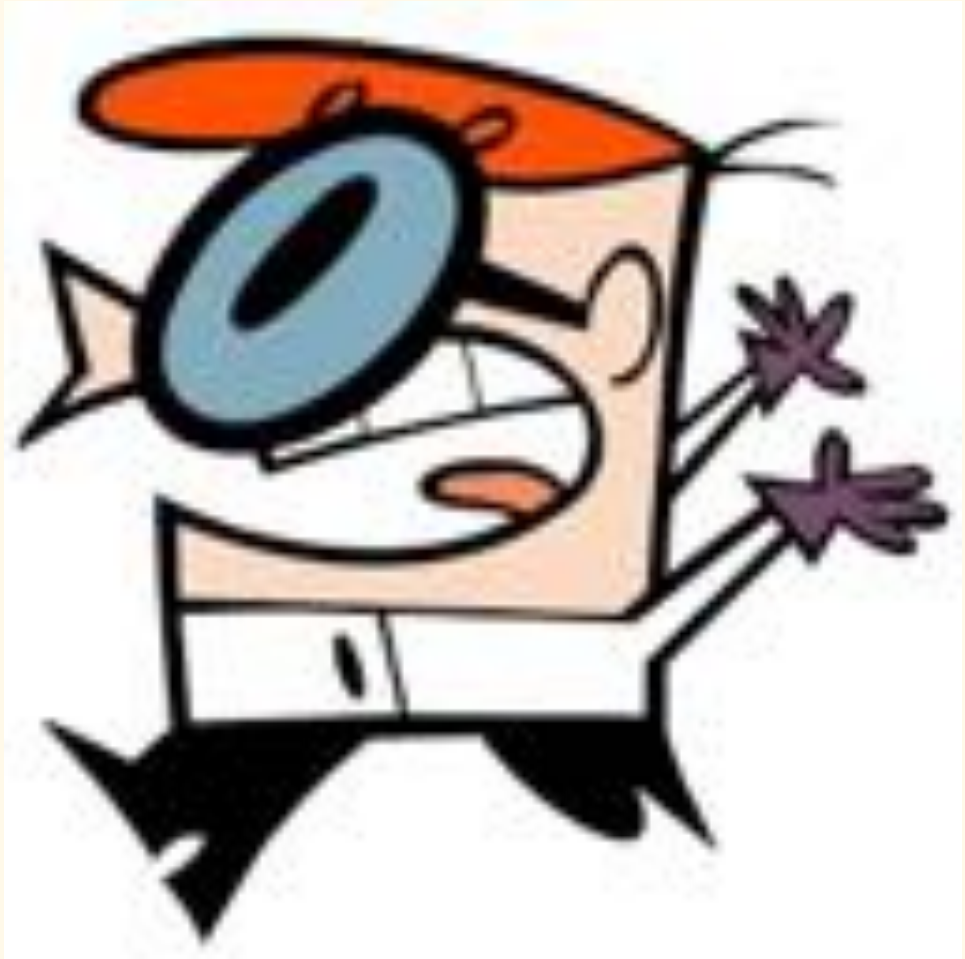
So far…

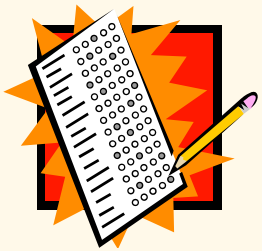
- Mostly about the theories
  - Argumentation and dialogue systems
  - Models: Cognitive, Machine learning, Language, Logic
  - Assignment: intelligent system development
- Today:
  - experiments, active testing
- Lab starting next week: run your own experiment
  “getting your hands dirty”

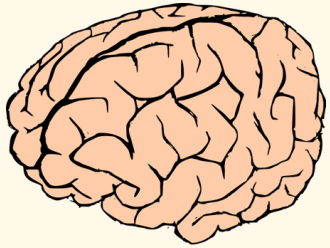
Today's topics

- **Why is experimentation useful?**
  - **In science**
  - **In practice / industry**
- **What constitutes experimentation? (see also: Cairns, 2016)**
  - **Empirical cycle**
  - **Experimental design**
  - **Statistical analysis**
  - **Experimental write-up**
- **Discussed using examples**
- **Topics we discuss along the way: manipulation, causality, validity (4 types), confounds & control, (in-)dependent variable, factor, condition, level, within- and between-subjects, counterbalancing….**
- **If you want to know more..**
- ***Appendix: more info that is useful for exam!***

Why care? Your (thesis) research….

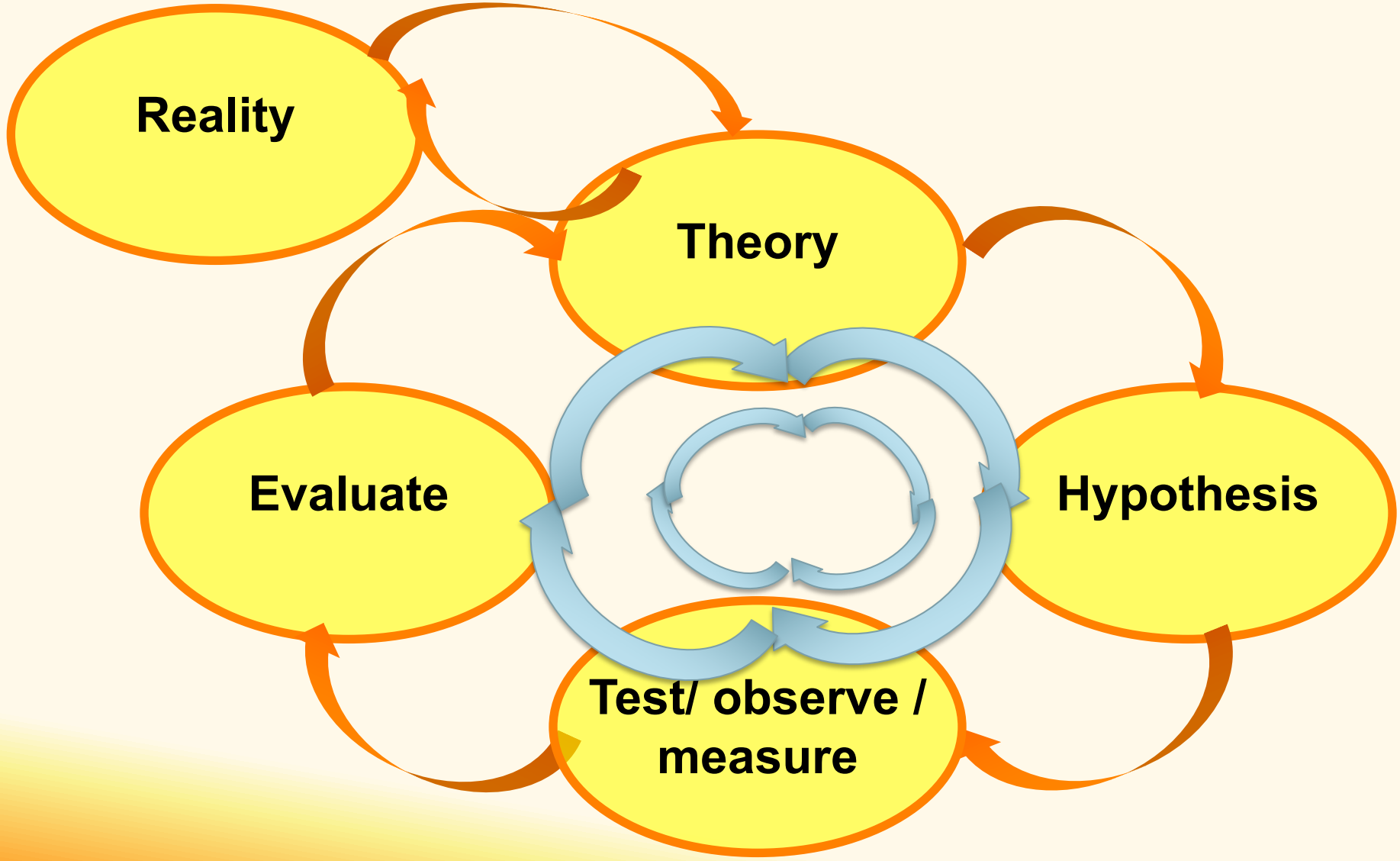






“I collected my data.
How do I now analyze these?”

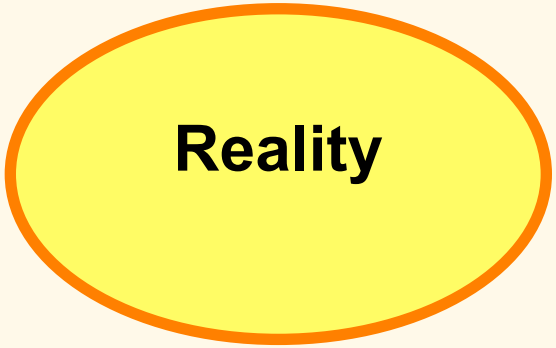
Empirical Cycle



Example 1: Human-Car Interaction

van der Heiden RMA, Janssen CP, Donker SF, Hardeman LES, Mans K, et al. (2018) Susceptibility to audio signals during autonomous driving. PLOS ONE 13(8): e0201963. https://doi.org/10.1371/journal.pone.0201963

Step 1: Reality



Driver crash risk factors and prevalence evaluation using naturalistic driving data

Thomas A. Dingus[a,1], Feng Guo[a,b], Suzie Lee[a], Jonathan F. Antin[a], Miguel Perez[a], Mindy Buchanan-King[a], and Jonathan Hankey[a]

[a]Virginia Tech Transportation Institute, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061; and [b]Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061

Edited by William J. Horrey, Liberty Mutual Research Institute for Safety, Hopkinton, MA, and accepted by the Editorial Board January 26, 2016 (received for review July 6, 2015)

The accurate evaluation of crash causal factors can provide fundamental information for effective transportation policy, vehicle design, and driver education. Naturalistic driving (ND) data collected with multiple onboard video cameras and sensors provide a unique opportunity to evaluate risk factors during the seconds leading up to a crash. This paper uses a National Academy of Sciences-sponsored ND dataset comprising 905 injurious and property damage crash events, the magnitude of which allows the first direct analysis (to our

driver-related factors (i.e., error, impairment, fatigue, and distraction) present in almost 90% of crashes. The results also definitively show that distraction is detrimental to driver safety, with handheld electronic devices having high use rates and risk.

naturalistic driving | crash risk | driver distraction | driver impairment | driver error

During recent years, the percentage of crashes involving some type of driver error or impairment before the crash was thought to be as high as 94% (1). Factors such as vehicle failures, roadway design or condition, or environment composed lower crash percentages. Naturalistic driving studies (NDSs) offer a unique opportunity to study driver performance and behavior

allows the sole use of crash events to determine the safety outcome for risk factor evaluation.

Using the SHRP 2 NDS crash database, this paper focuses on and addresses the following categories of driver performance and behavior that contribute to crash events: (*i*) observable impairment, which was determined from a 20-s precrash video segment; observable impairment includes apparent drug/alcohol influence, drowsiness/fatigue, or emotion (i.e., anger, sadness, crying, and/or early impacted driver performance; including a variety of vehicle operag., failing to yield properly to other urn); (*ii*) momentary driver judgment as aggressive driving and speeding; tion determined from a 6-s precrash se of in-vehicle and handheld devices, active interaction with passengers, and outside distractions.

Materials and Methods

Database and Instrumentation. The SHRP 2 NDS dataset comprises more than 2 PB of continuous naturalistic driving data collected during a 3-y period from more than 3,500 participants, aged 16–98, who resided near the following six site centers: Buffalo, NY; Tampa, FL; Seattle; Durham, NC; Bloomington, IN; and State College, PA. The naturalistic driving data were collected automatically from key-on to key-off for every trip taken in one of the volunteer


Case closed?

SAE level	Name
Human **monitors driving environment**	
0	No Automation
1	Driver Assistance
2	Partial Automation
Vehicle **monitors driving environment**	
3	Conditional Automation
4	High Automation
5	Full Automation

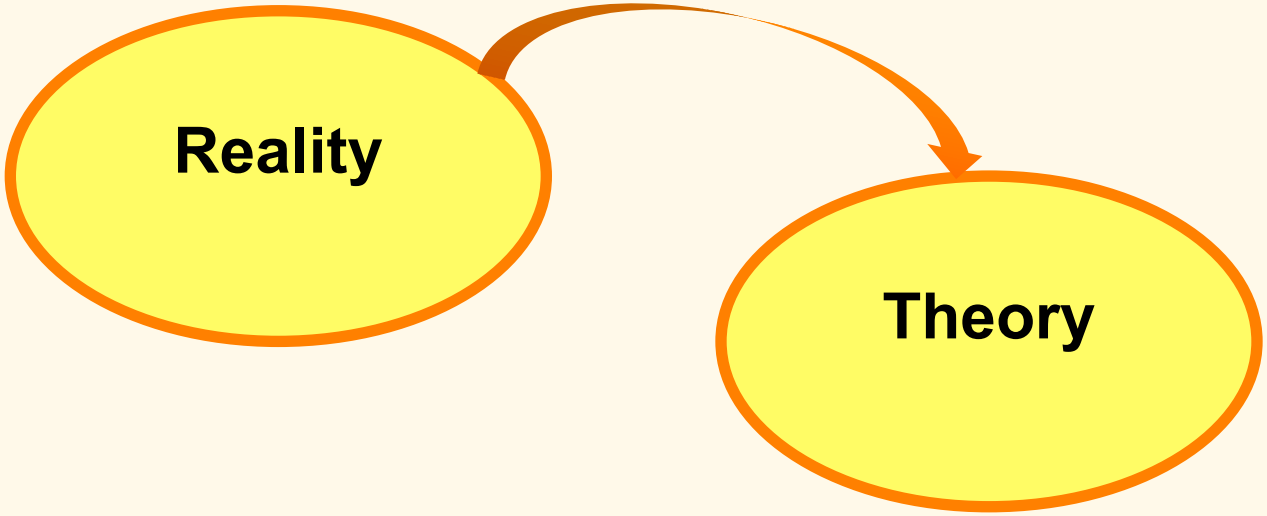
SAE International (2014)

http://jalopnik.com/teslas-autopilot-system-is-awesome-and-creepy-and-a-sig-1736573089

Research question

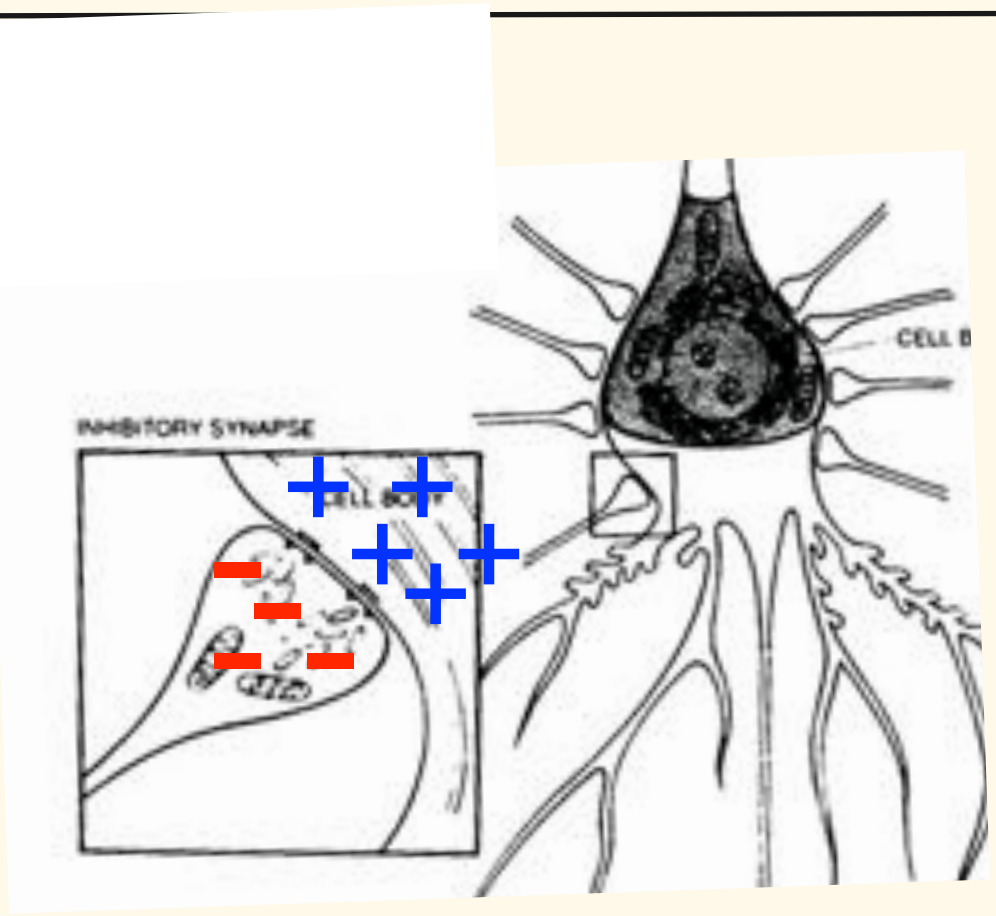
- **How susceptible are people to auditory alerts under (autonomous) driving conditions?**

Step 2: Theory

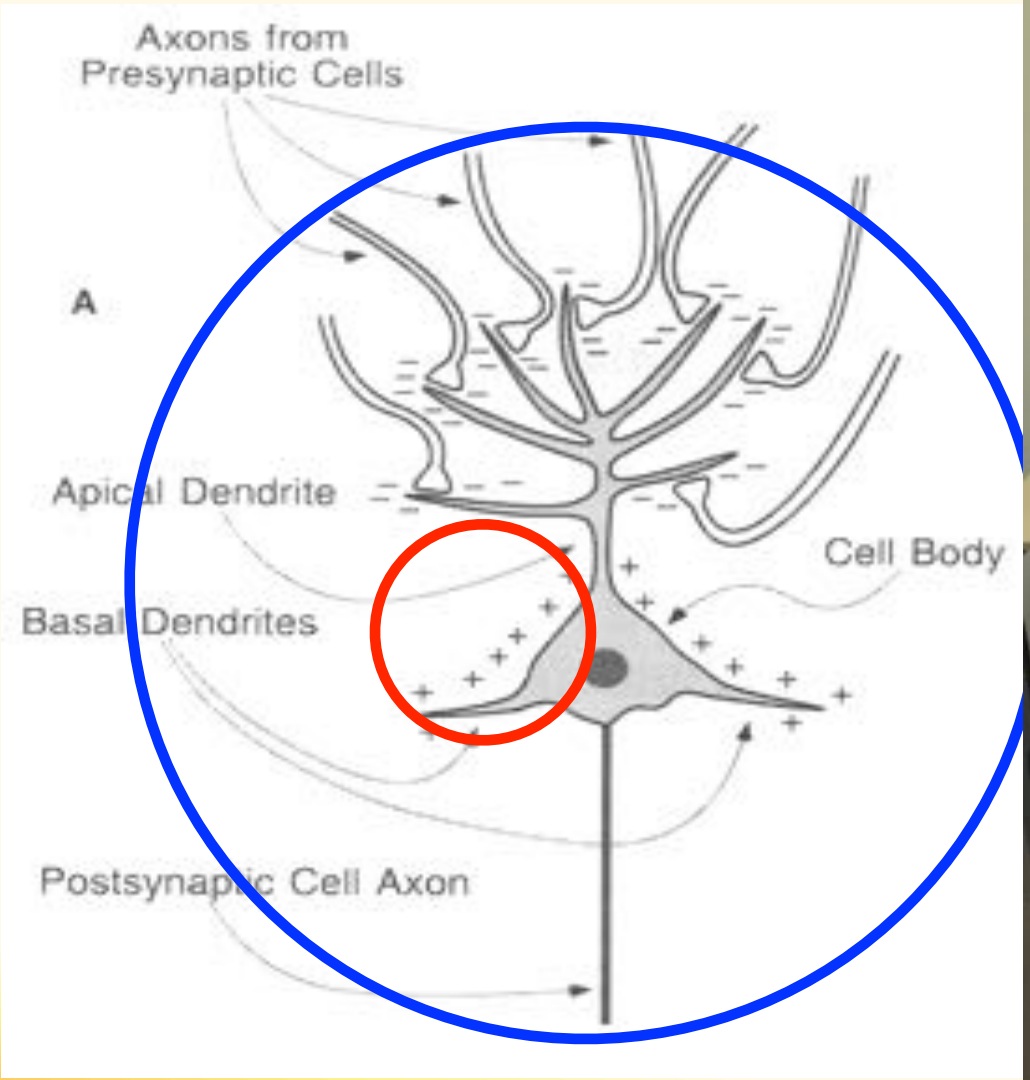


Sounds from Fabiani & Friedman (1995)

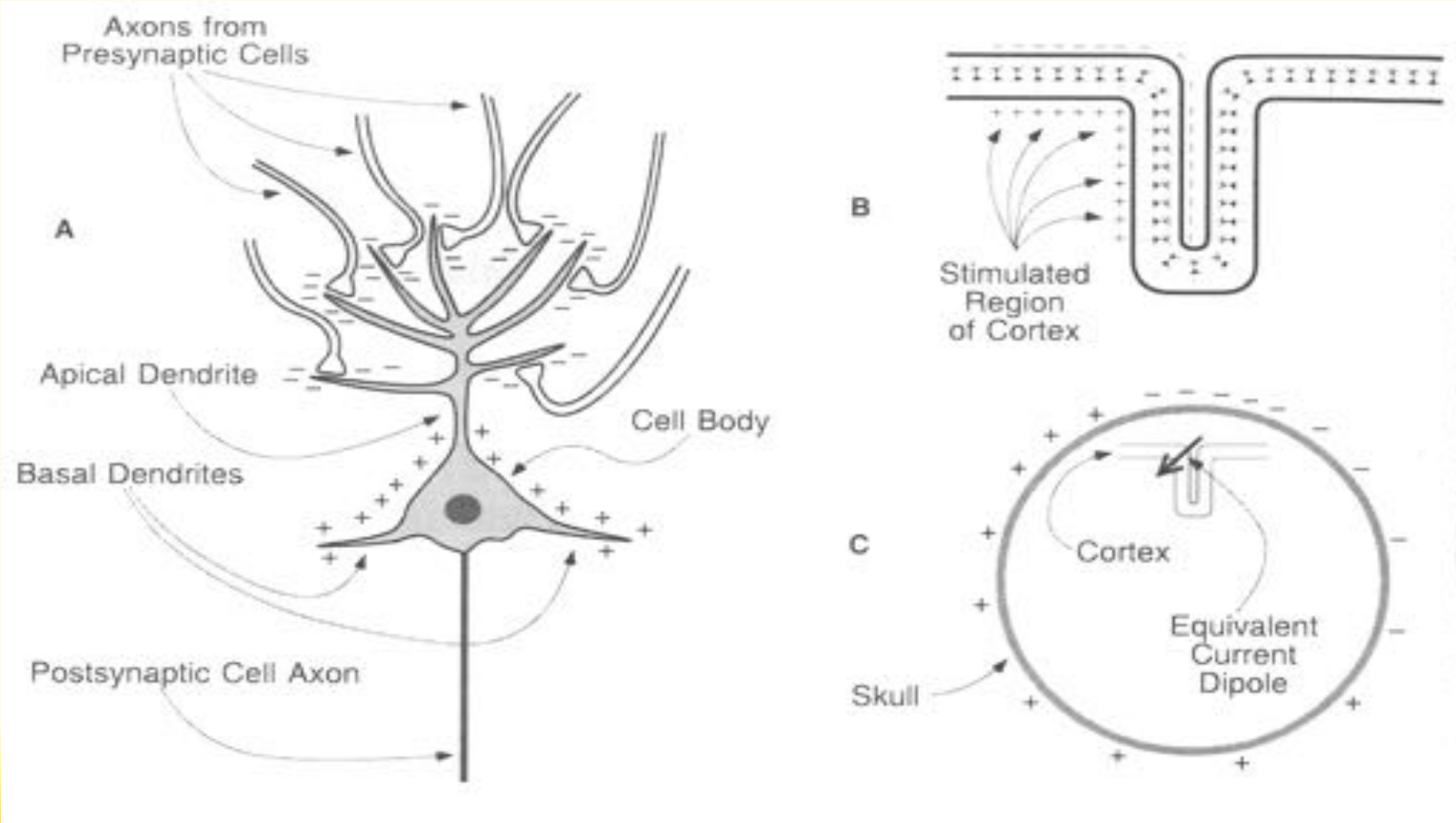
EEG



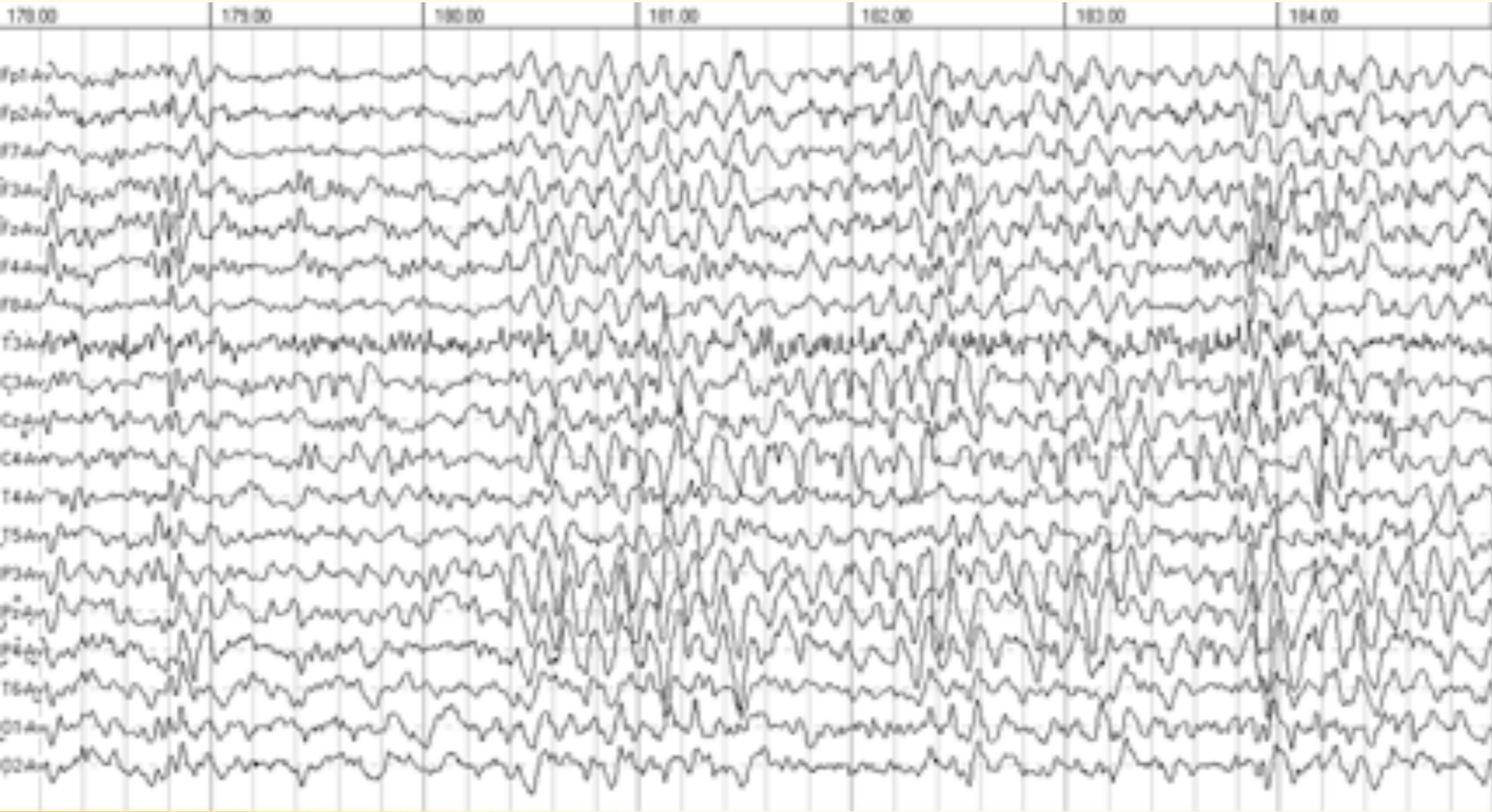
EEG

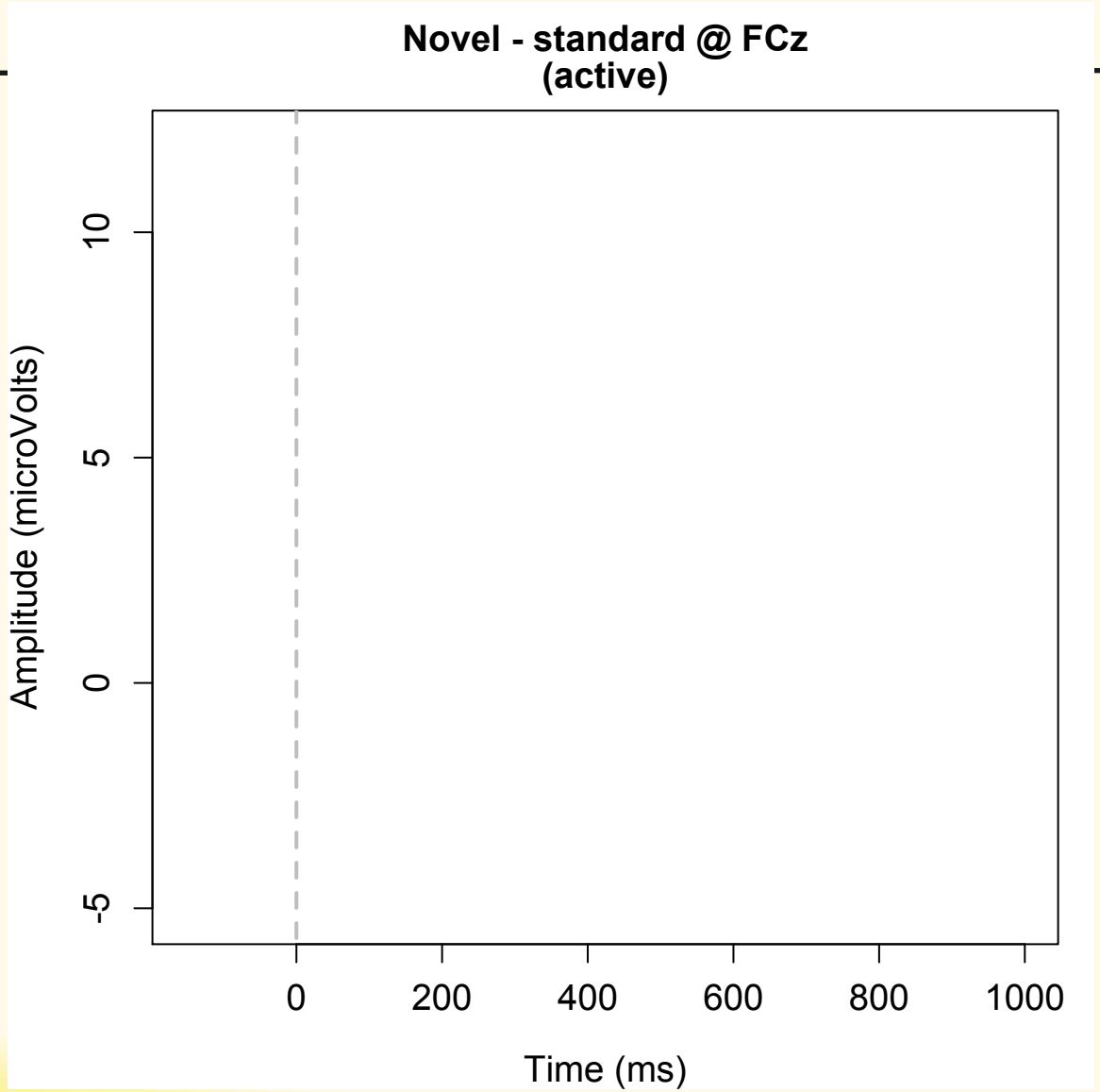

Axons from
Presynaptic Cells
A
Apical Dendrite
Cell Body
Basal Dendrites
Postsynaptic Cell Axon

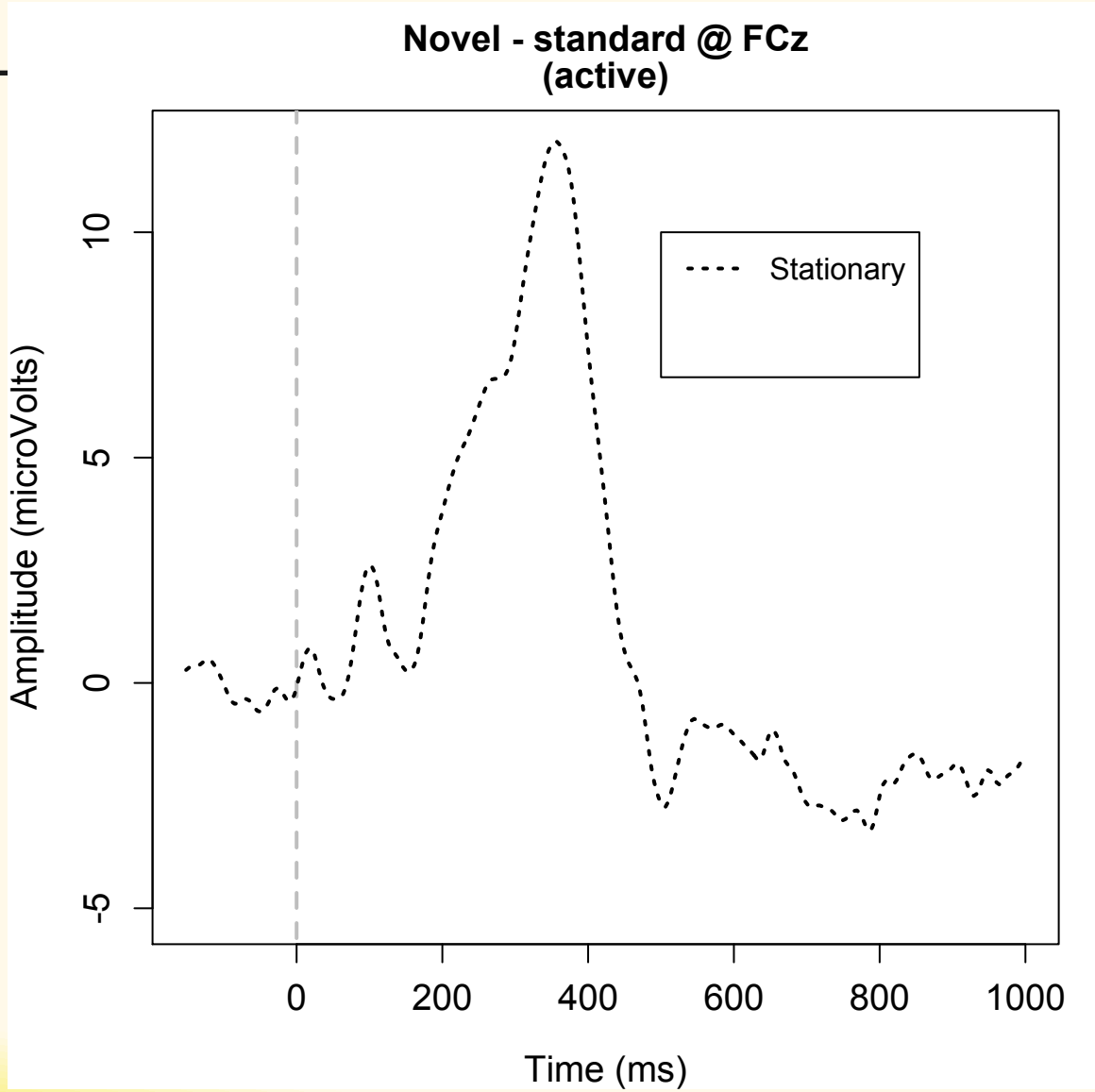
EEG


Axons from
Presynaptic Cells
A
Apical Dendrite
Cell Body
Basal Dendrites
Postsynaptic Cell Axon
B
Stimulated
Region
of Cortex
C
Cortex
Equivalent
Current
Dipole
Skull

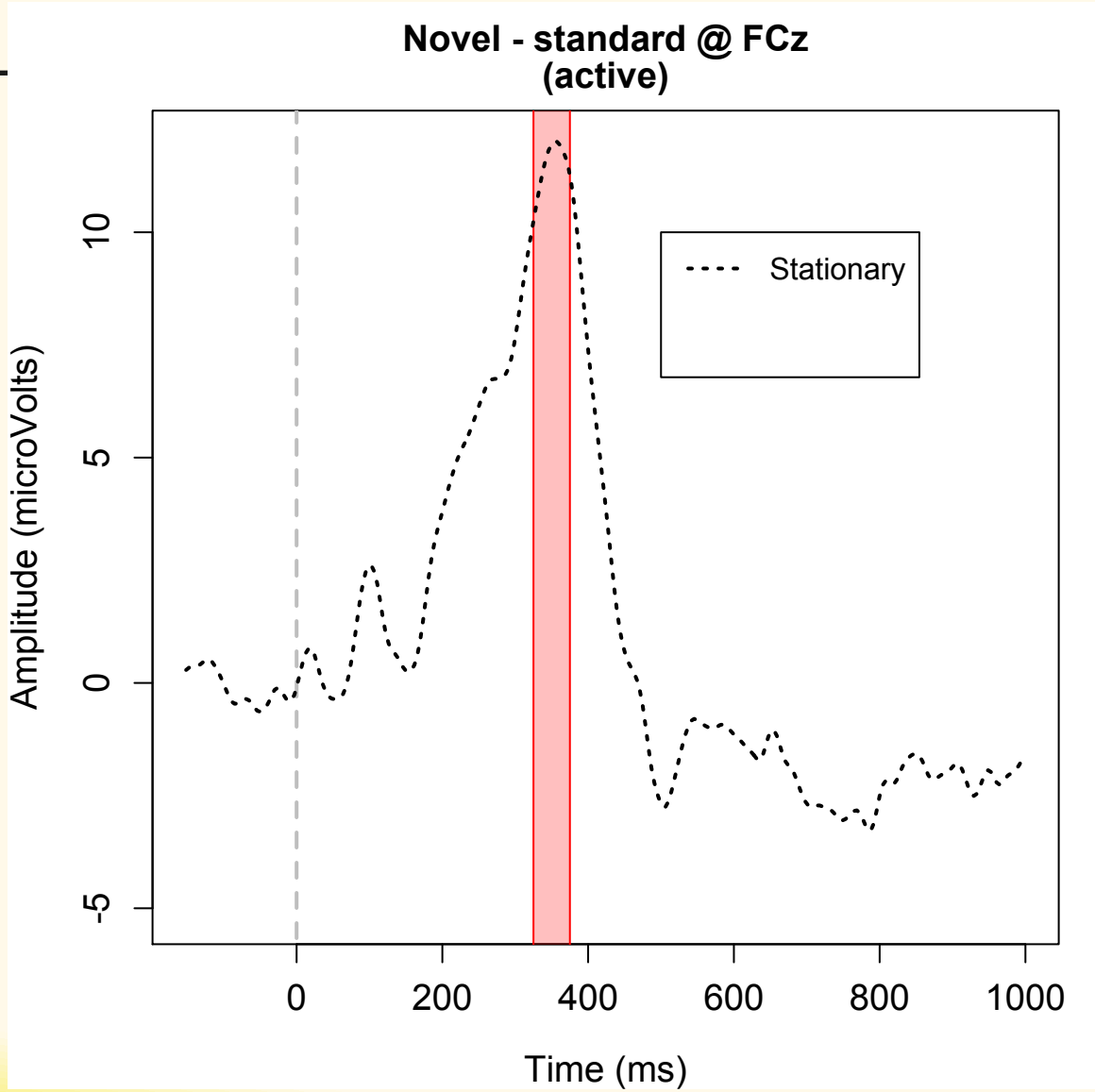
EEG output



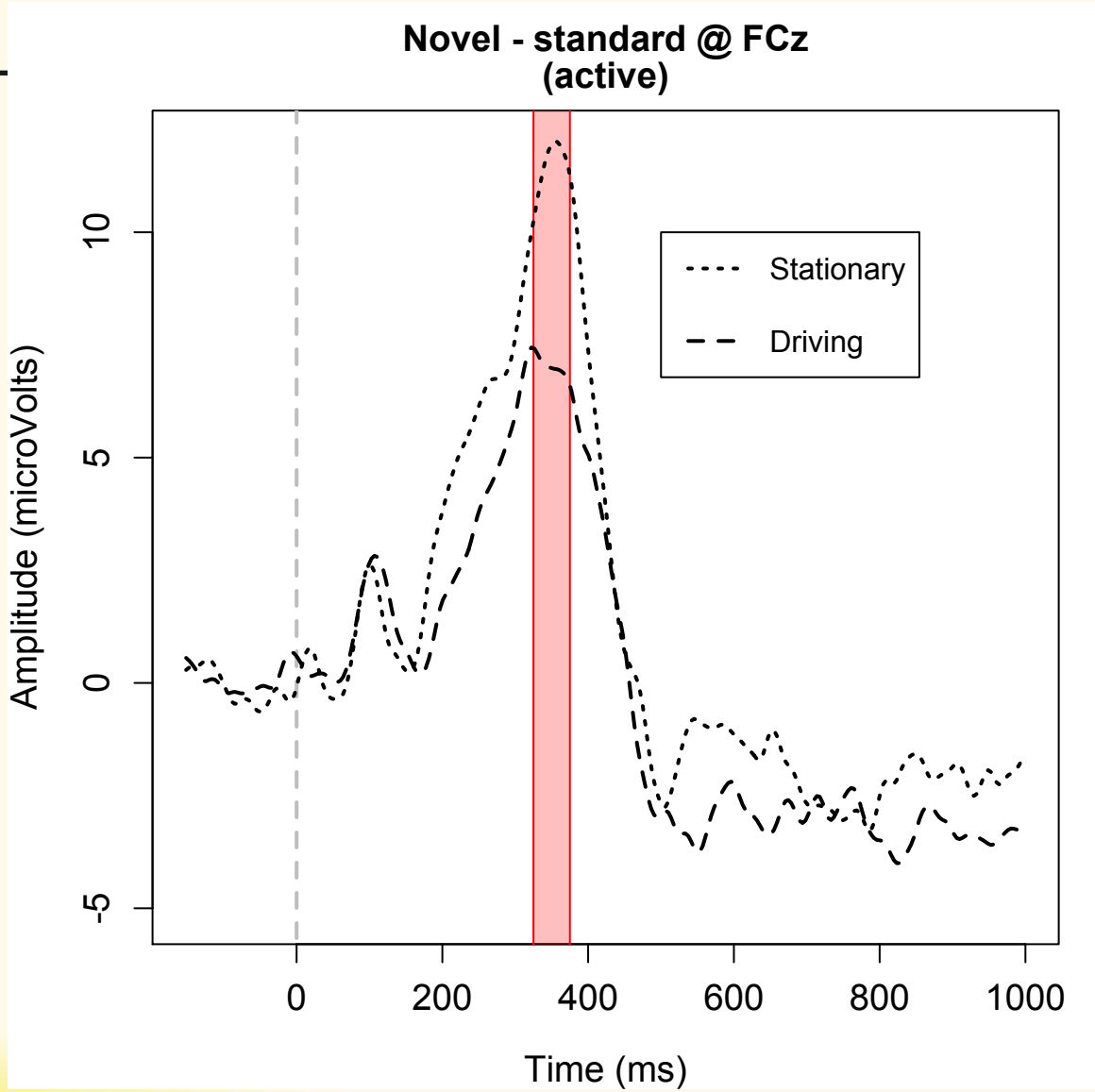

Novel - standard @ FCz
(active)
Amplitude (microVolts)
10
5
0
-5
0
200
400
600
800
1000
Time (ms)



See also: Wester et al. (2008) Accident Analysis & Prevention



See also: Wester et al. (2008) Accident Analysis & Prevention

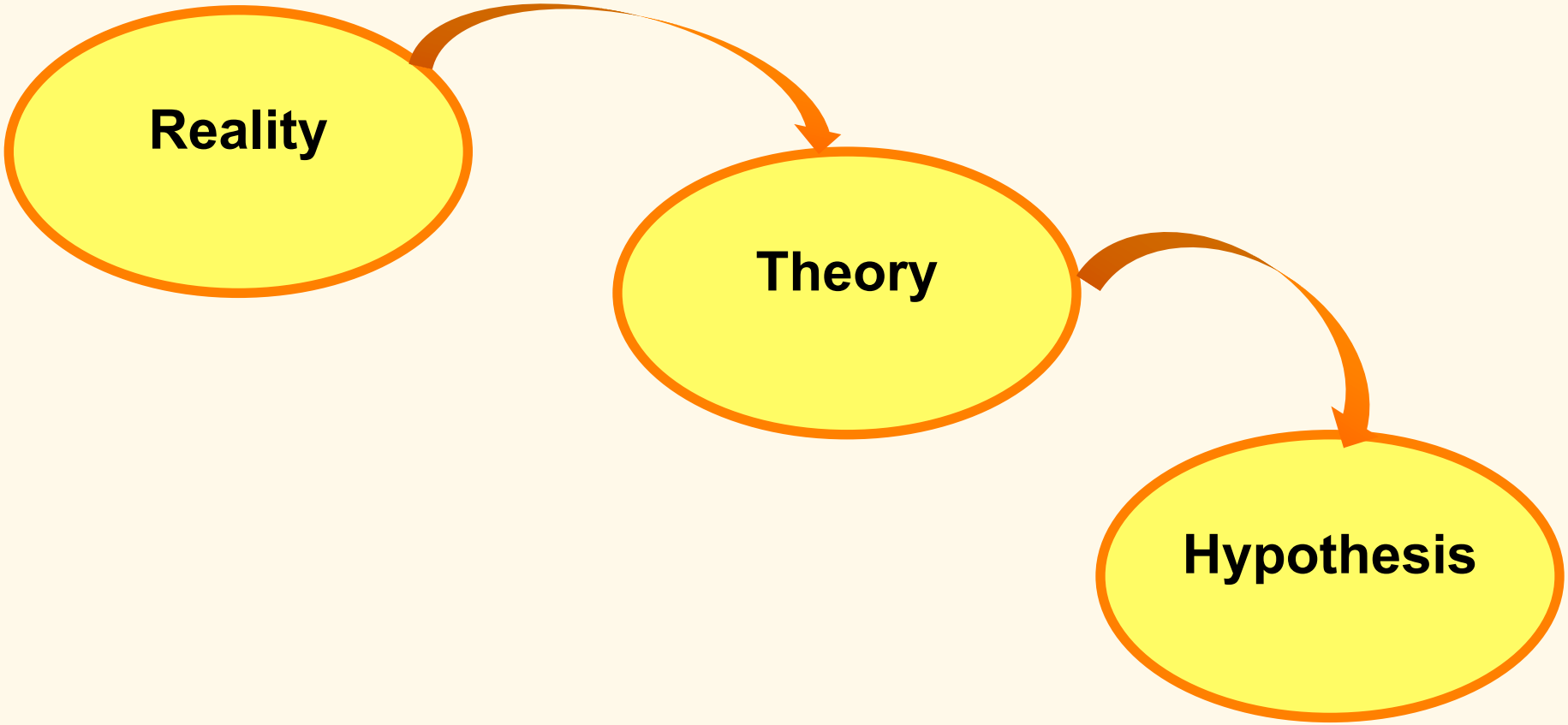


See also: Wester et al. (2008) Accident Analysis & Prevention

Theory conclusion

- **Workload (and driving) reduces susceptibility to sounds**
- **This can be measured in an ERP procedure**

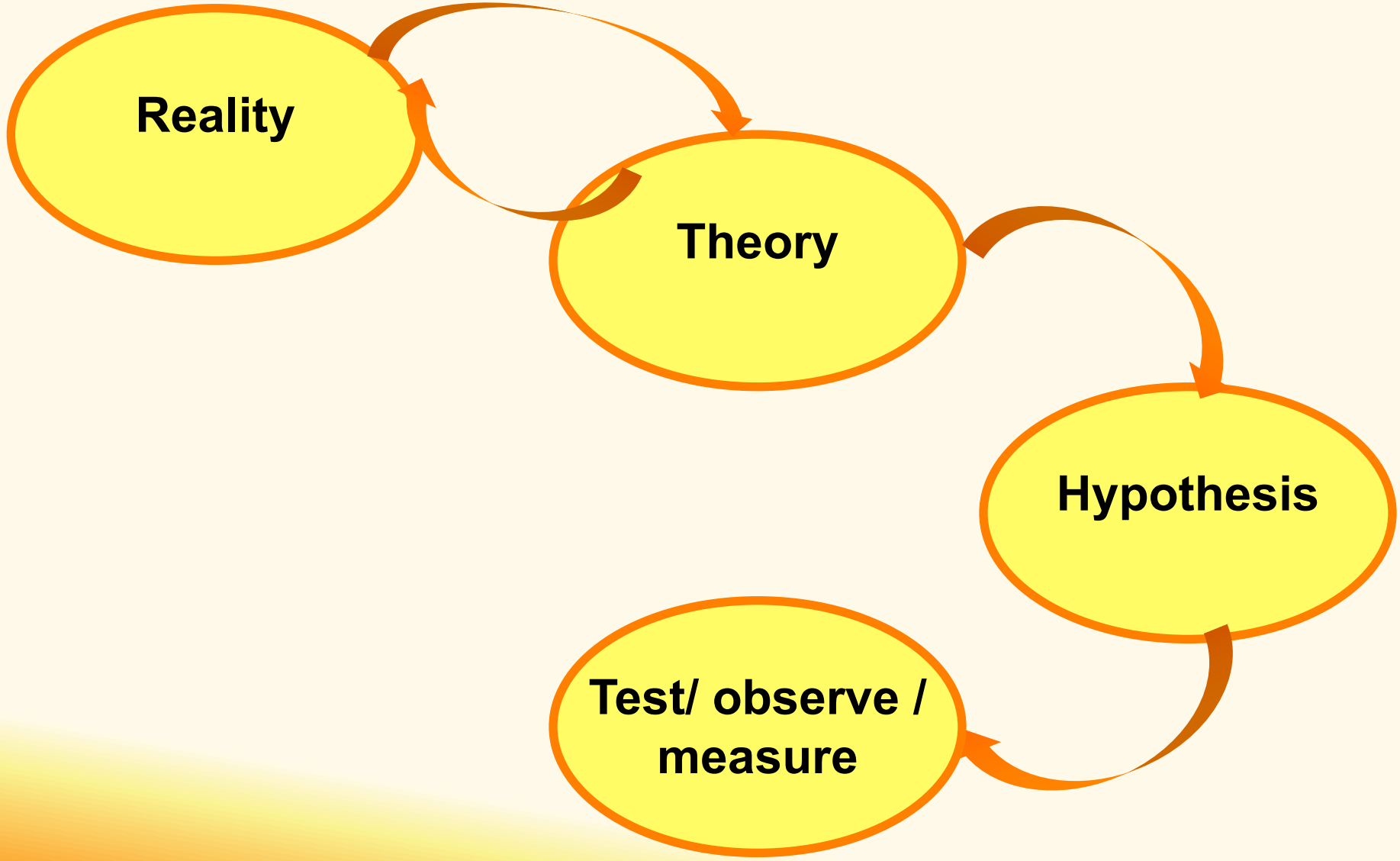
Step 3: Hypothesis



Hypothesis

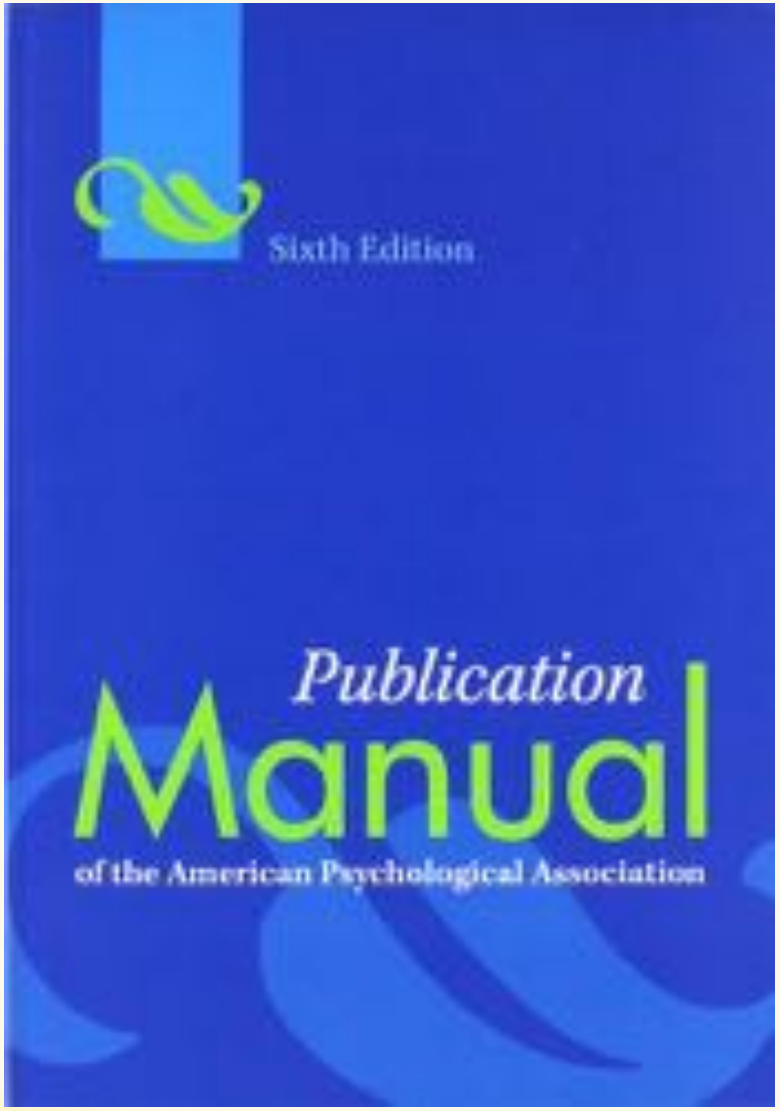
- **If autonomous is similar to:**
  - **Stationary: No reduction (in ERP at FCz)**
  - **Driving: strong reduction**
  - **Something in between: some reduction**

Step 4: Test / observe / measure



Methods section of paper

- Participants
- Materials/Stimuli
- Design
- Procedure
- Measures



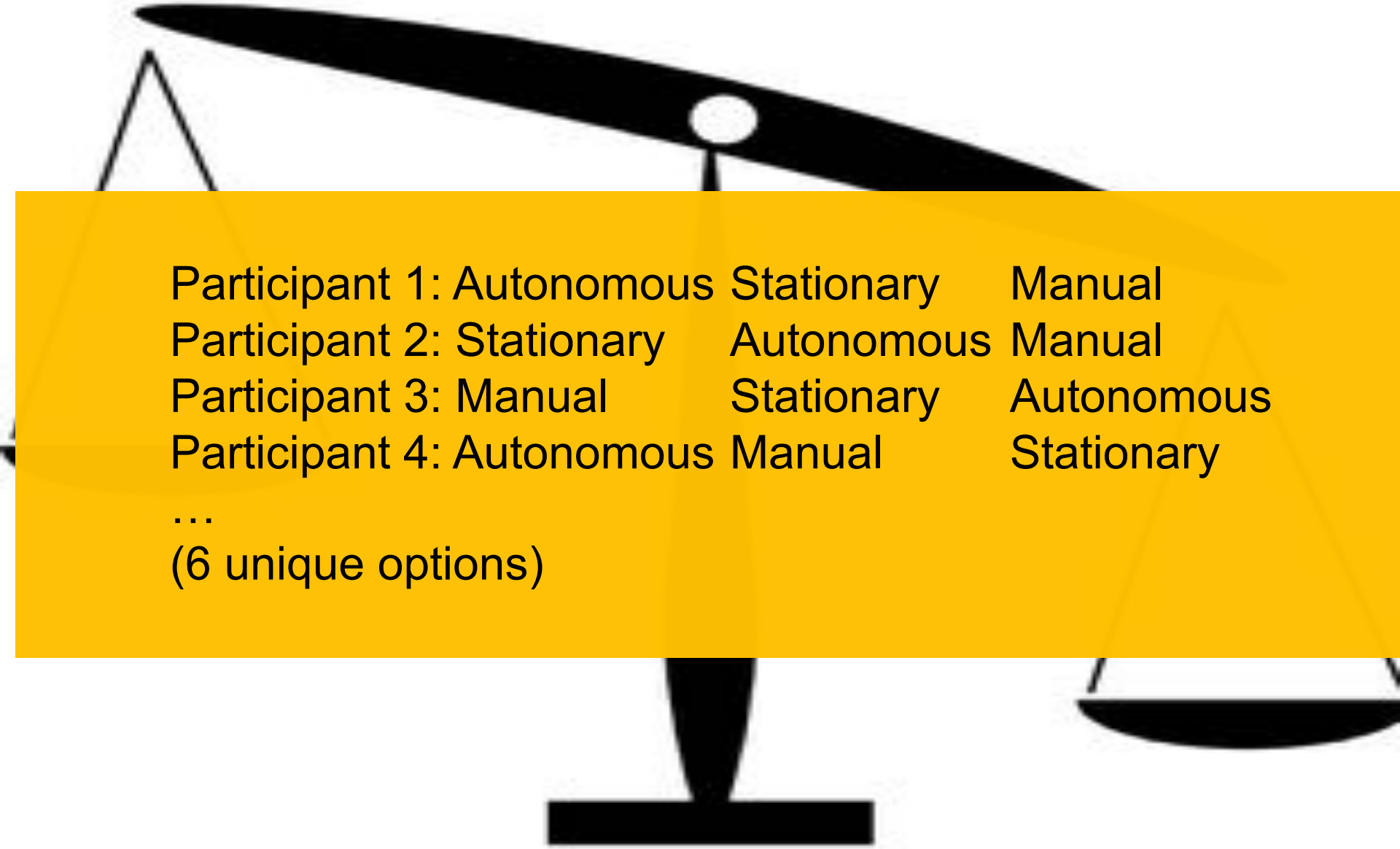
Participants, materials

- 18 participants
- Informed consent
- See paper for details

Design

- ***Within-subjects***: driving condition
  - Stationary
  - Autonomous
  - (Manual) driving
- ***Between-subjects***: response mode
  - Active condition
  - Passive condition

Condition order: *Counterbalanced*



Participant 1:	Autonomous	Stationary	Manual
Participant 2:	Stationary	Autonomous	Manual
Participant 3:	Manual	Stationary	Autonomous
Participant 4:	Autonomous	Manual	Stationary

…

(6 unique options)

Procedure

1. Consent signing & explanation
2. Practice driving
3. Practice oddball
4. Apply EEG cap
5. 12 trials of 3 conditions
6. Questionnaire

- 120-150 minutes total

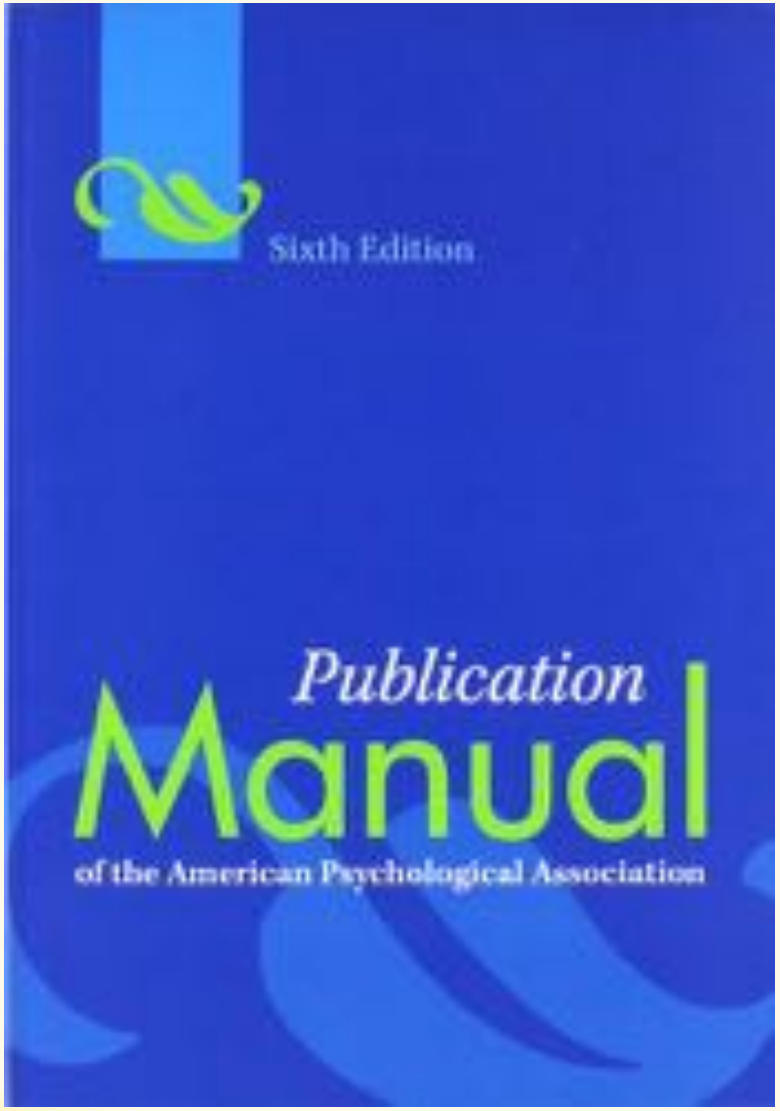
Measures

- ERP (see paper for details on calculation)
- Active condition: Reaction time button press

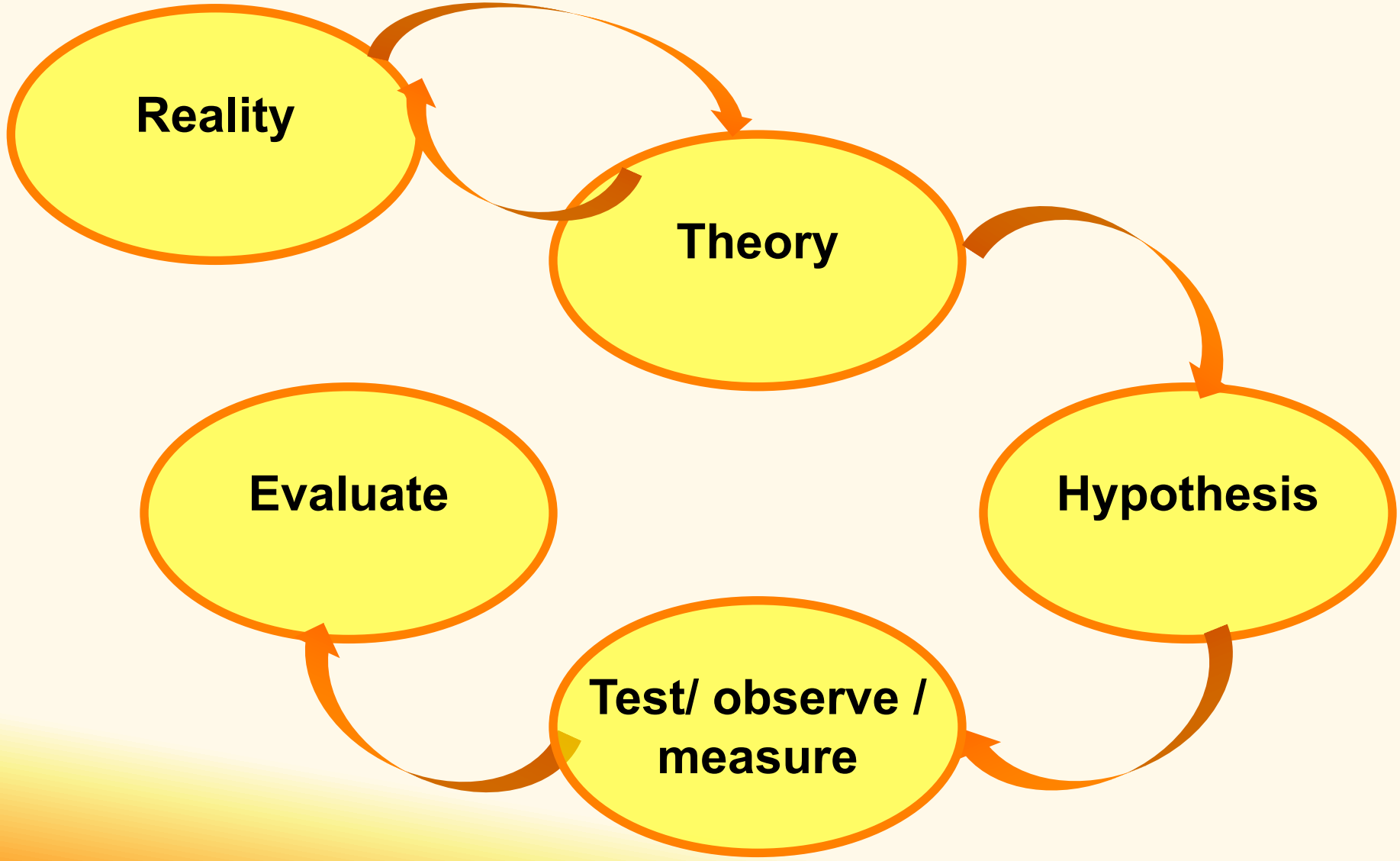
- (also reports criteria for hypothesis acceptance / rejection, effect sizes, etc)

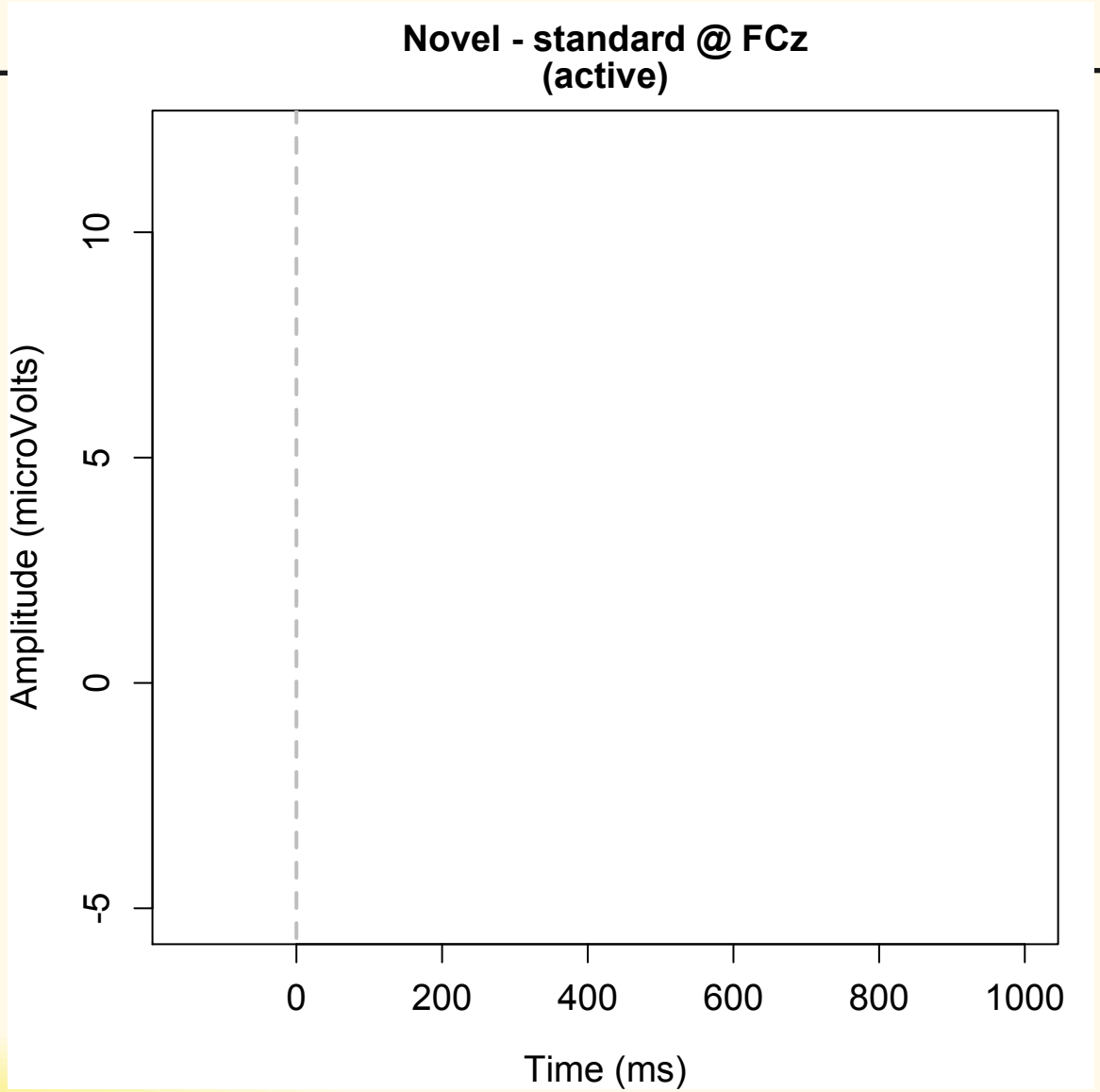
Methods section of paper

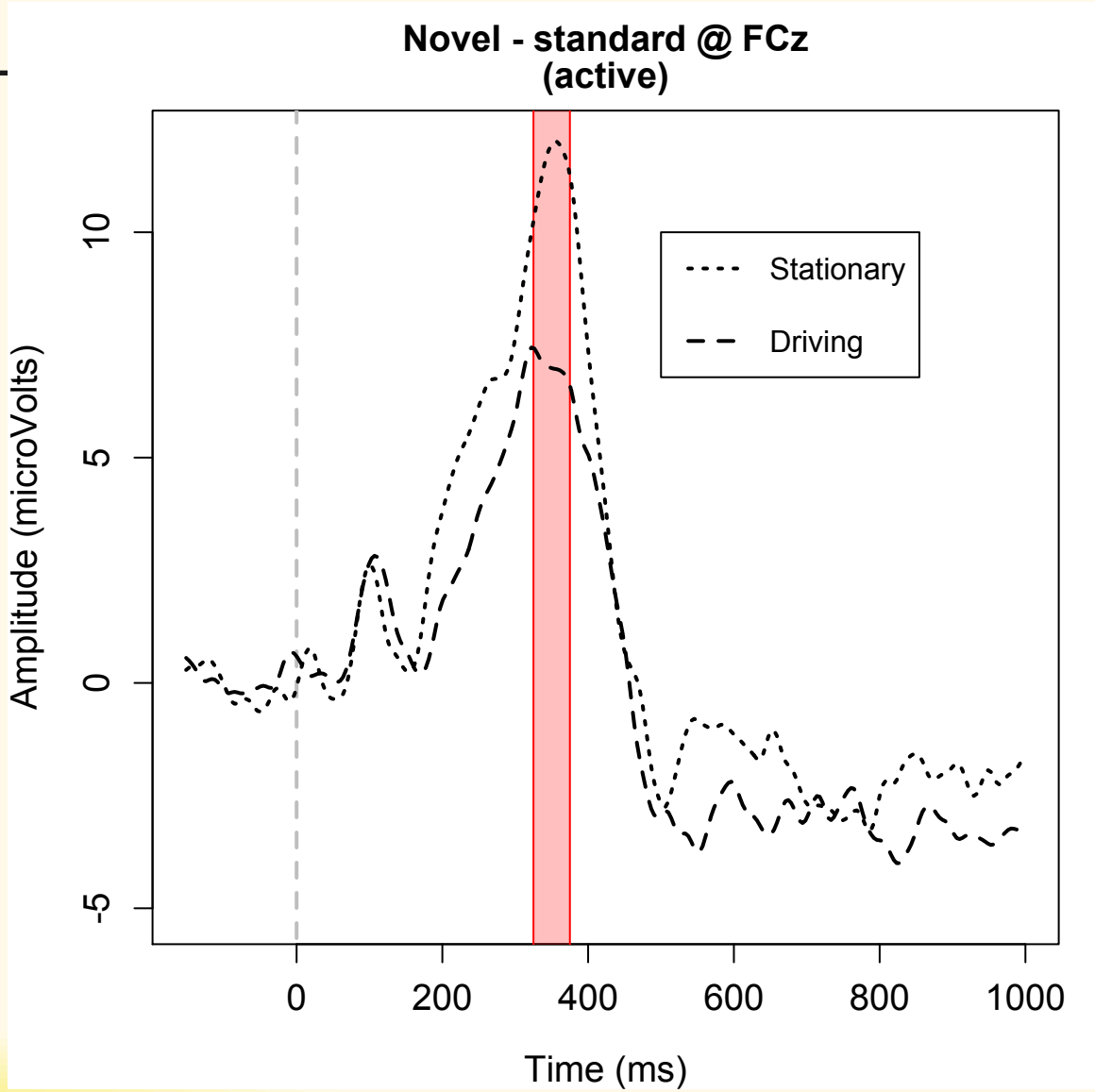
- Participants
- Materials/Stimuli
- Design
- Procedure
- Measures

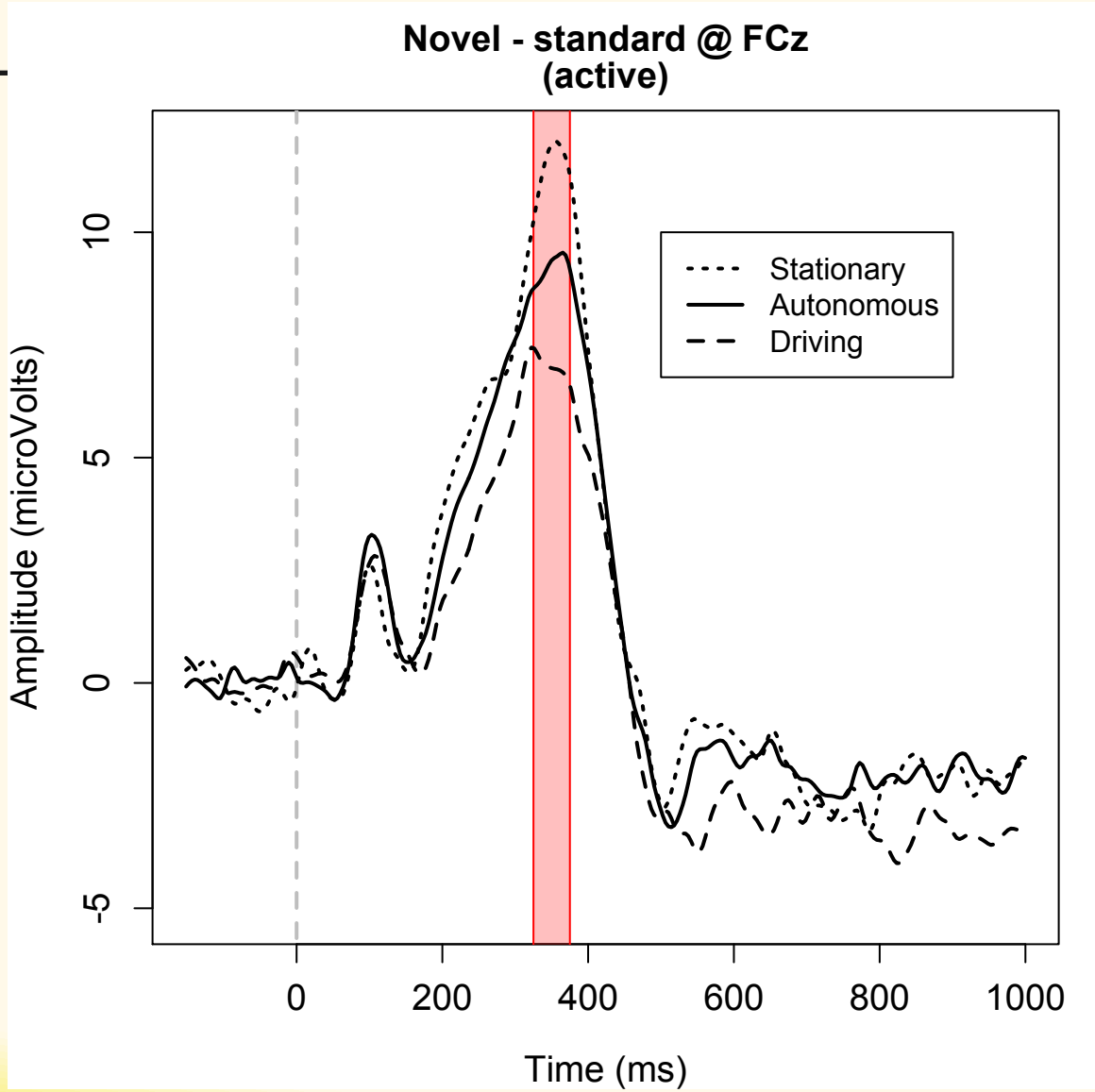


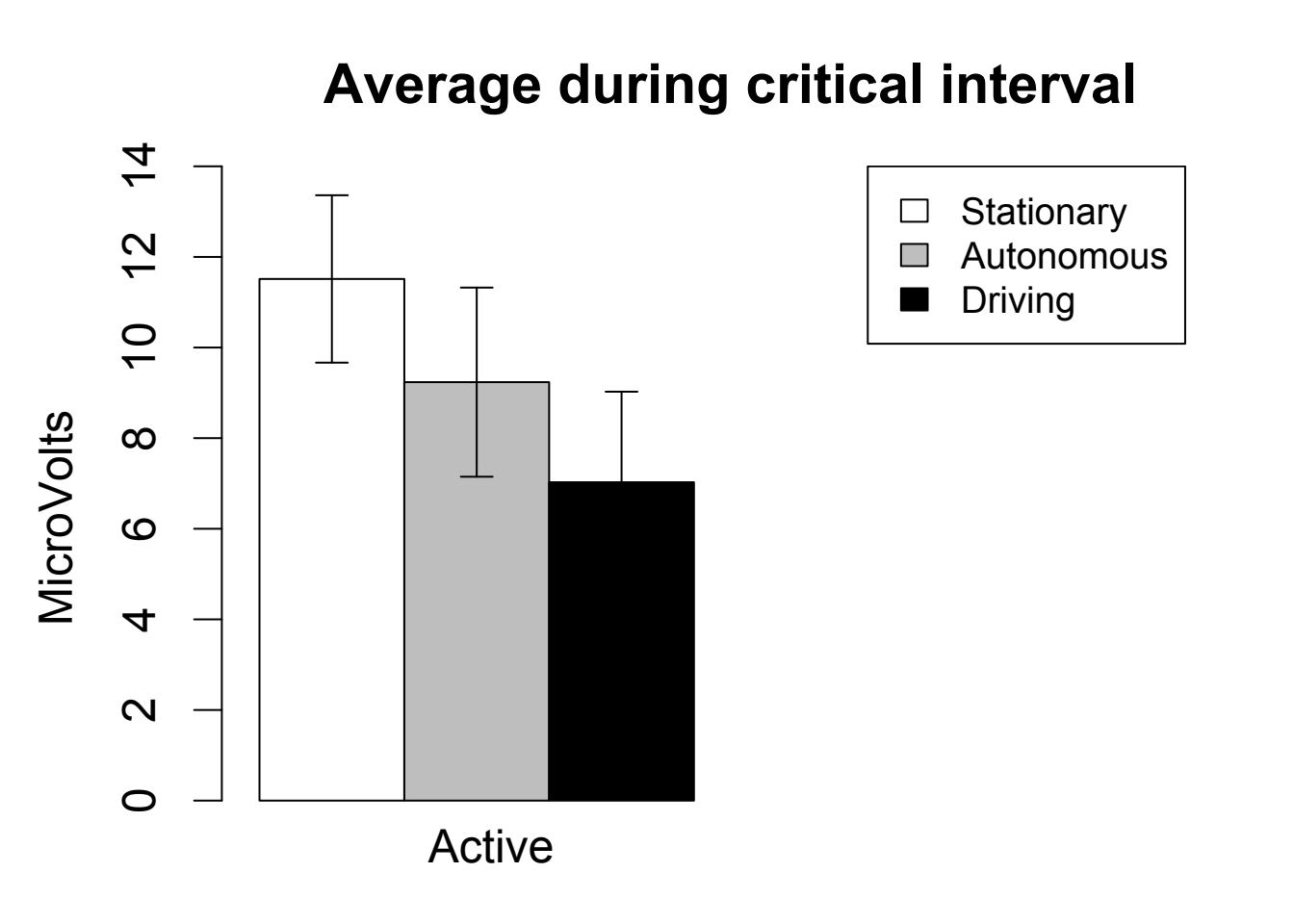
Step 5: Evaluate

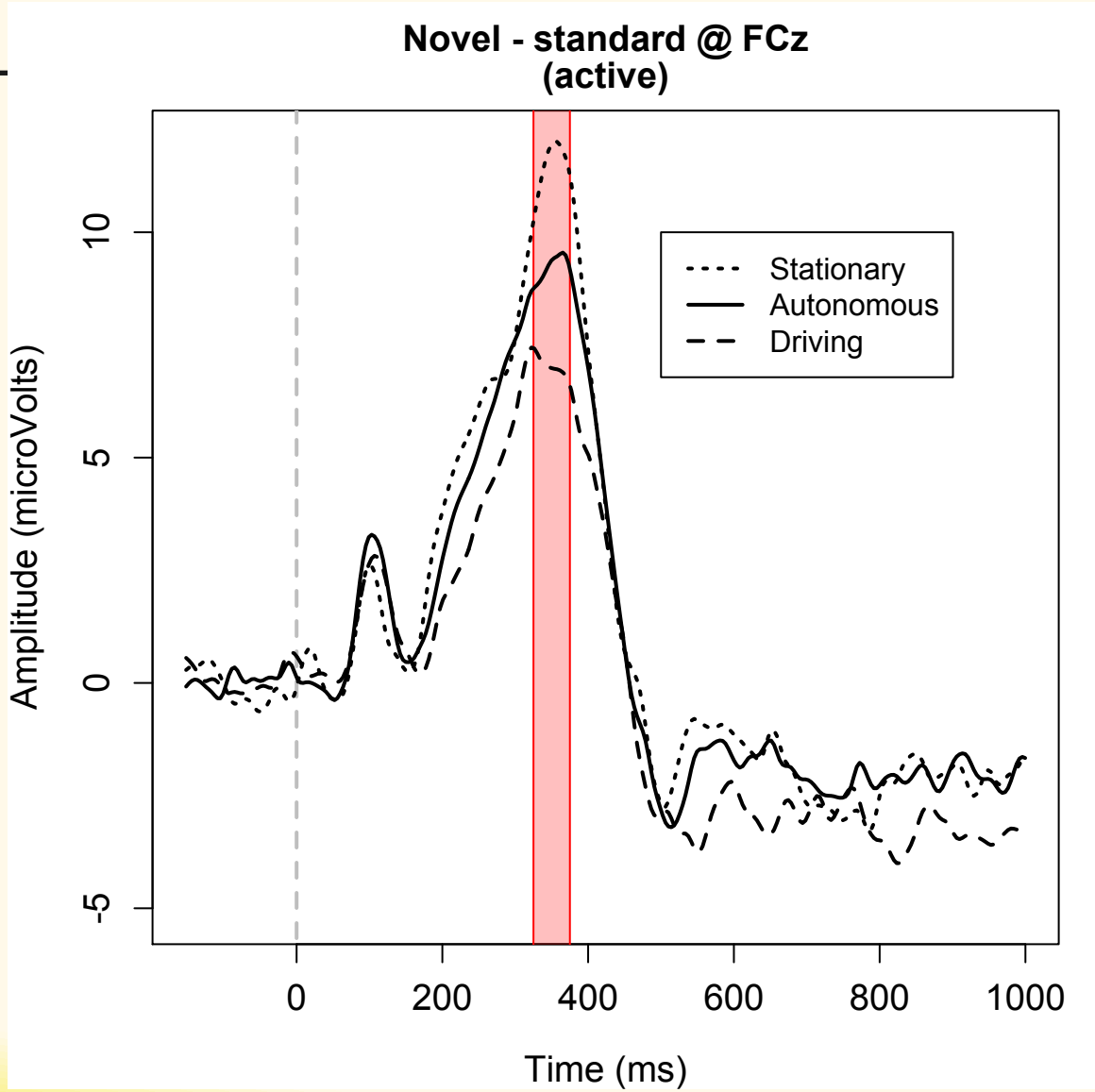


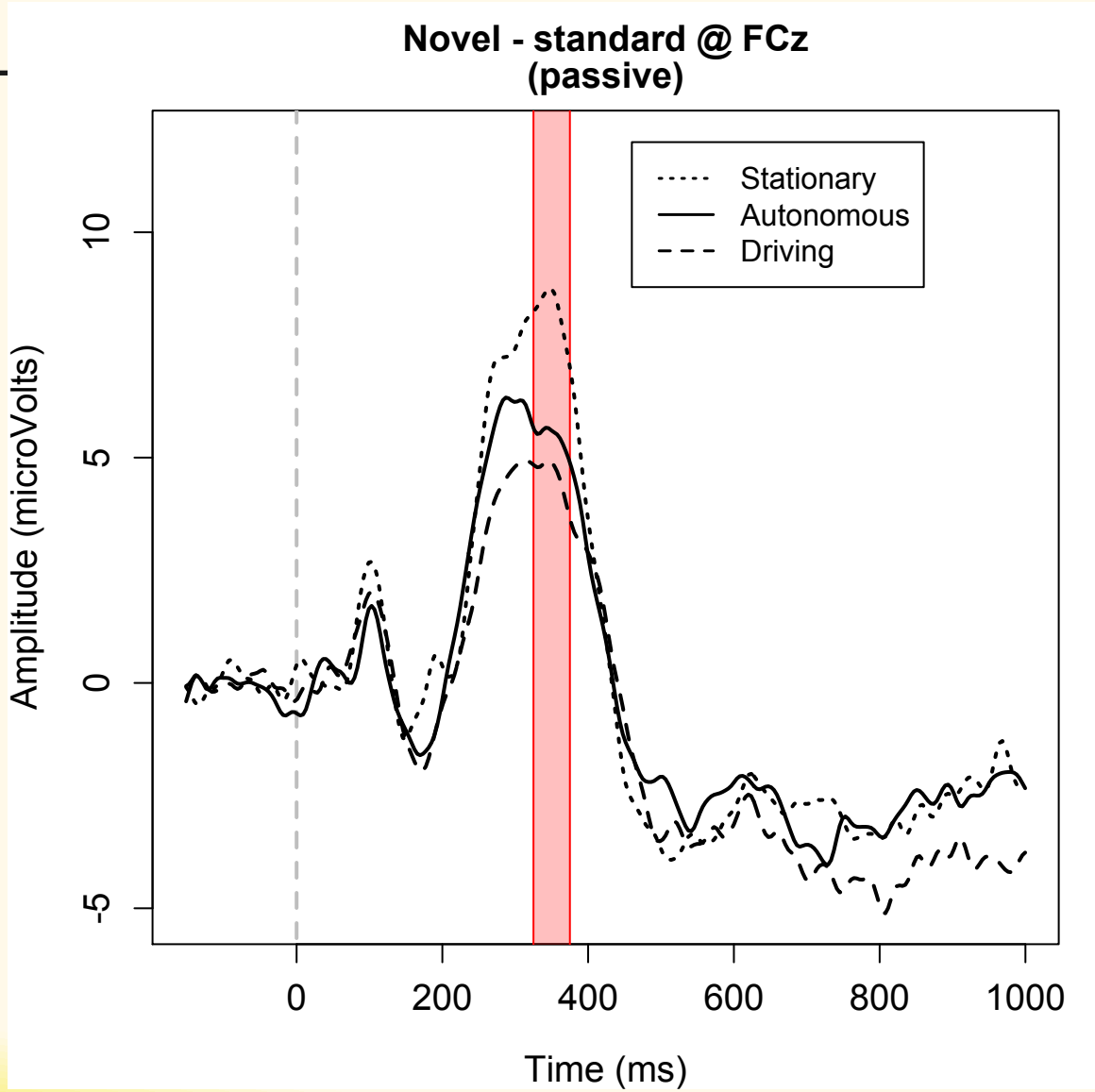

Novel - standard @ FCz
(active)
Amplitude (microVolts)
10
5
0
-5
0
200
400
600
800
1000
Time (ms)

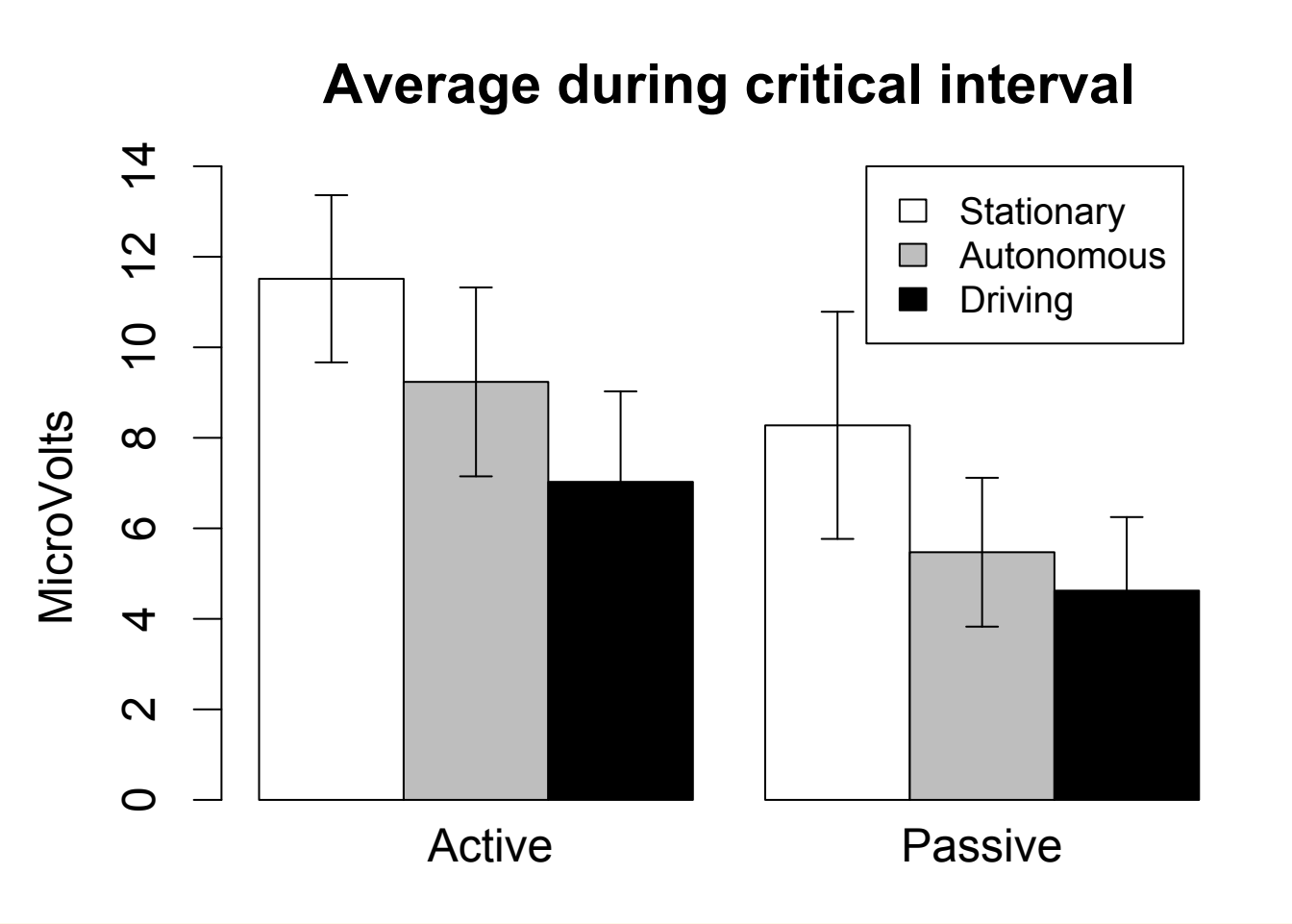

Novel - standard @ FCz
(active)
Stationary
Driving
Amplitude (microVolts)
10
5
0
-5
0
200
400
600
800
1000
Time (ms)


Novel - standard @ FCz
(active)
Amplitude (microVolts)
10
5
0
-5
0
200
400
600
800
1000
Time (ms)
Stationary
Autonomous
Driving

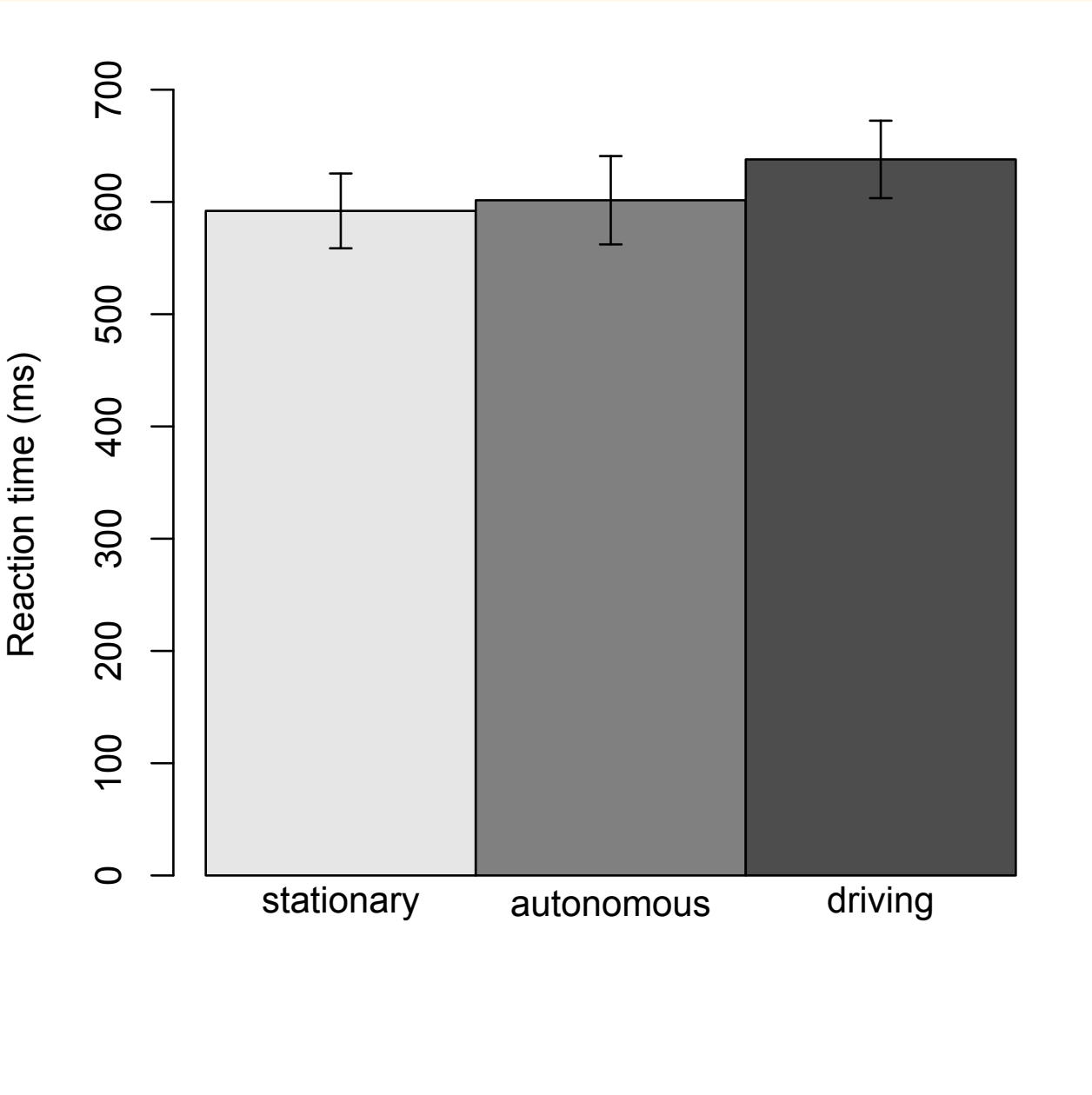

Average during critical interval
MicroVolts
0
2
4
6
8
10
12
14
Active
Stationary
Autonomous
Driving


Novel - standard @ FCz
(active)
Amplitude (microVolts)
10
5
0
-5
0
200
400
600
800
1000
Time (ms)
Stationary
Autonomous
Driving

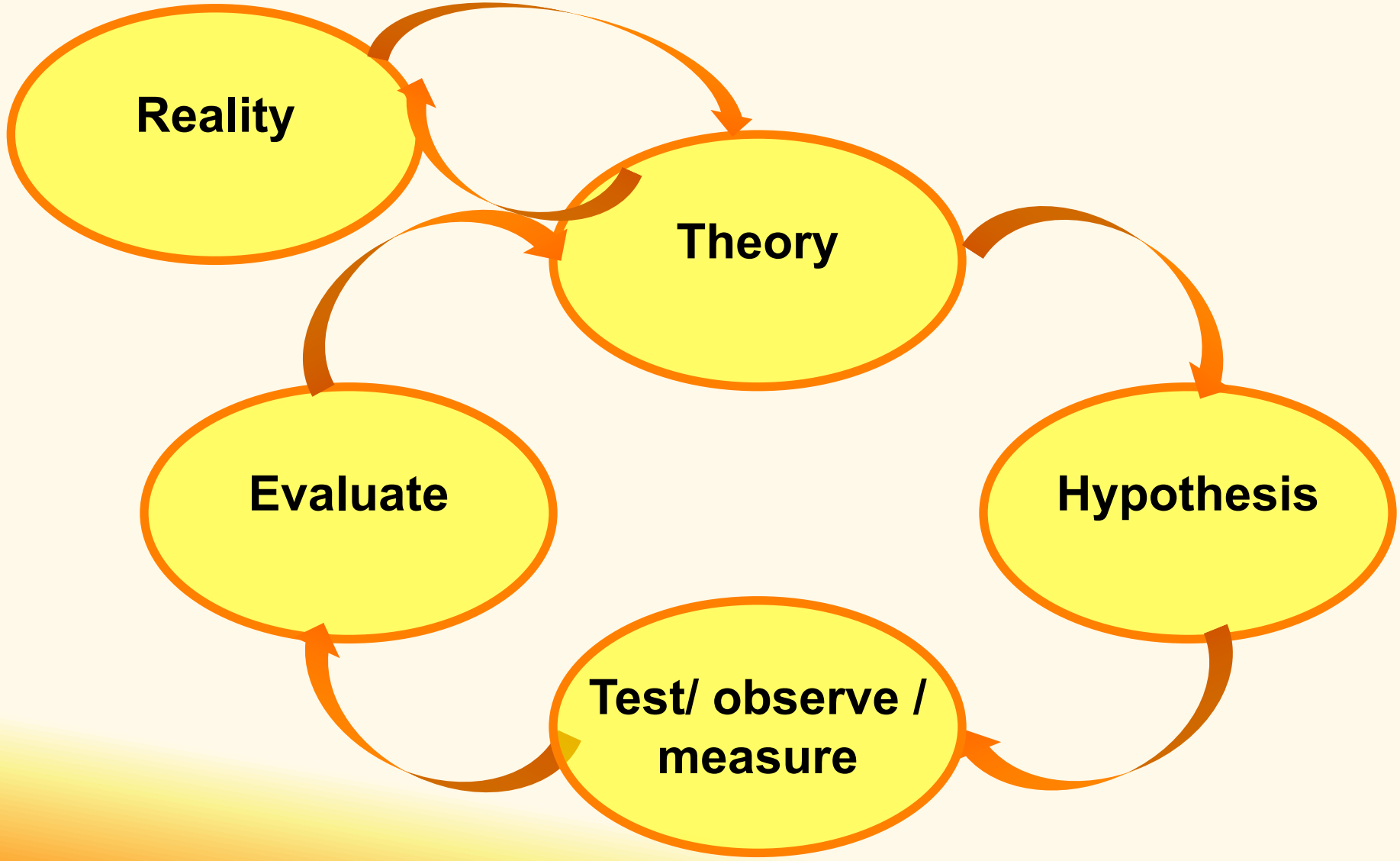

Novel - standard @ FCz
(passive)
Stationary
Autonomous
Driving
Amplitude (microVolts)
10
5
0
-5
0
200
400
600
800
1000
Time (ms)



- Distribution (means and error bars) seem to cover different ranges
- Therefore, conditions look different
- But… statistical test is needed to test
- (details too complicated for now → but you will practice with statistics, see assignment)


Reaction time (ms)
0
100
200
300
400
500
600
700
stationary
autonomous
driving

Step 6: Relate result to theory and reality



Revisit hypothesis

- **If autonomous is similar to:**
  - **Stationary: No reduction**
  - **Driving: strong reduction**
  - **Something in between: some reduction**

Revisit hypothesis

- **If autonomous is similar to:**
  - ~~**Stationary: No reduction**~~
  - **Driving: strong reduction**
  - ***Something in between: some reduction***

- Better during active condition

Interpretation

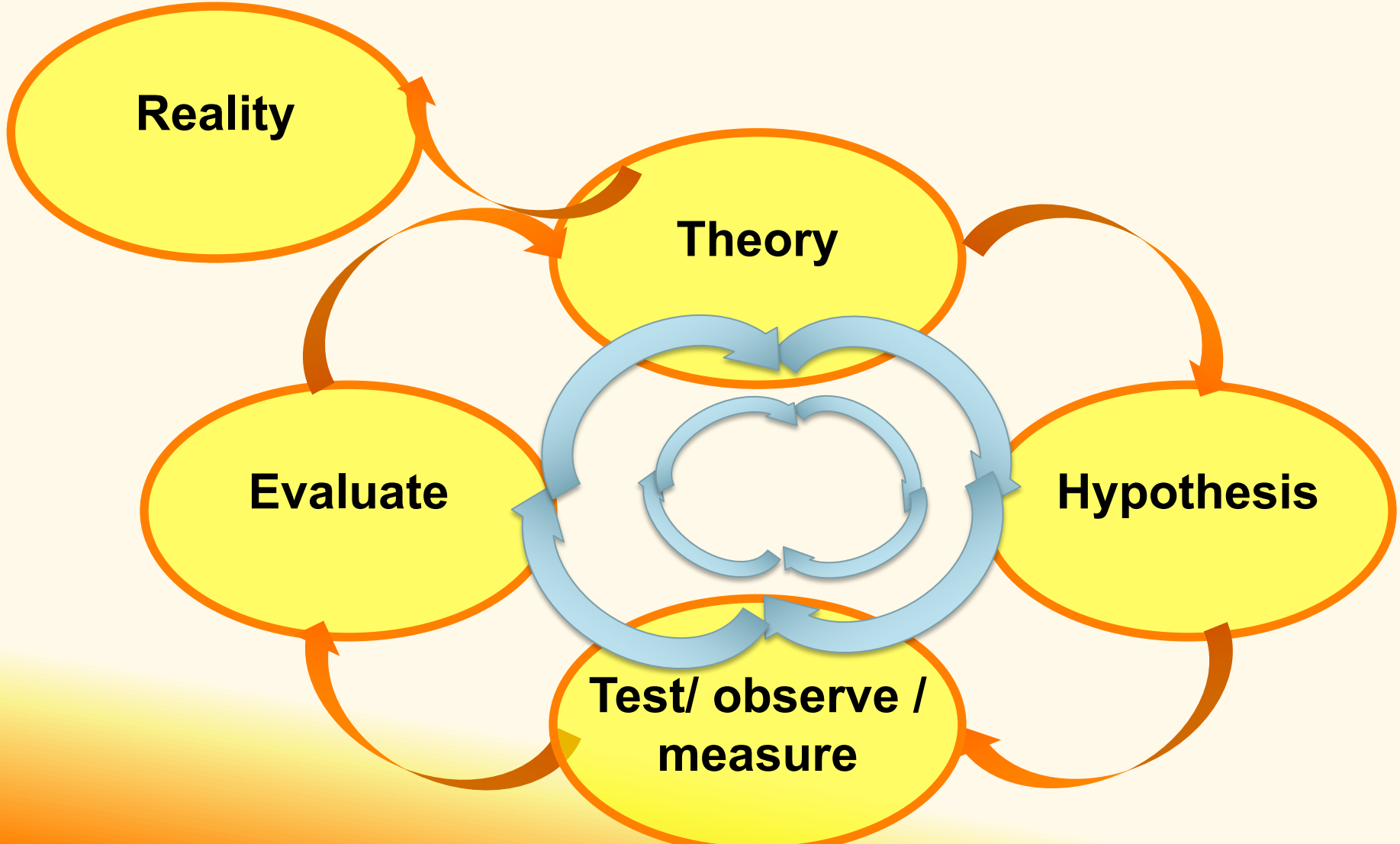
- **Susceptibility to alerts reduced under autonomous driving conditions**
- **Especially when listening *passively***

Implication for reality

- **Systems rely on alerts, but… people might not notice them**

Limitations

- Not tested under distracted conditions



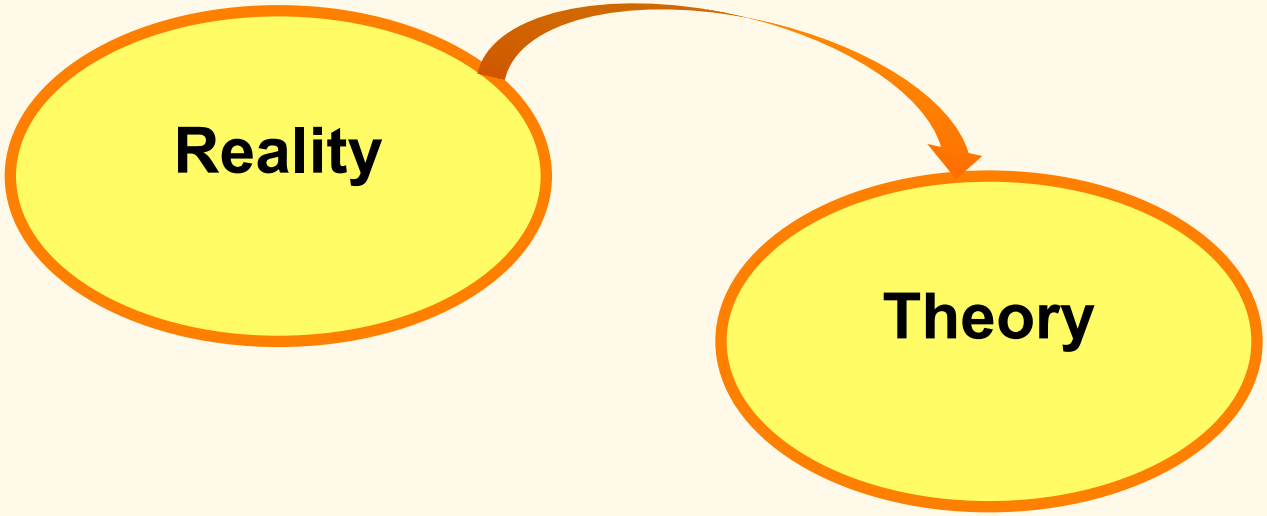
2 follow-up studies

1. Is fP3 reduced under other distracting tasks?

2. Is fP3 reduced when combining distracting tasks with driving?

- Interim results in:
  Janssen, Van der Heiden, Donker, Kenemans (2019) Measuring susceptibility to alerts while encountering mental workload. AutomotiveUI '19 Adjunct Proceedings

Step 2: Theory

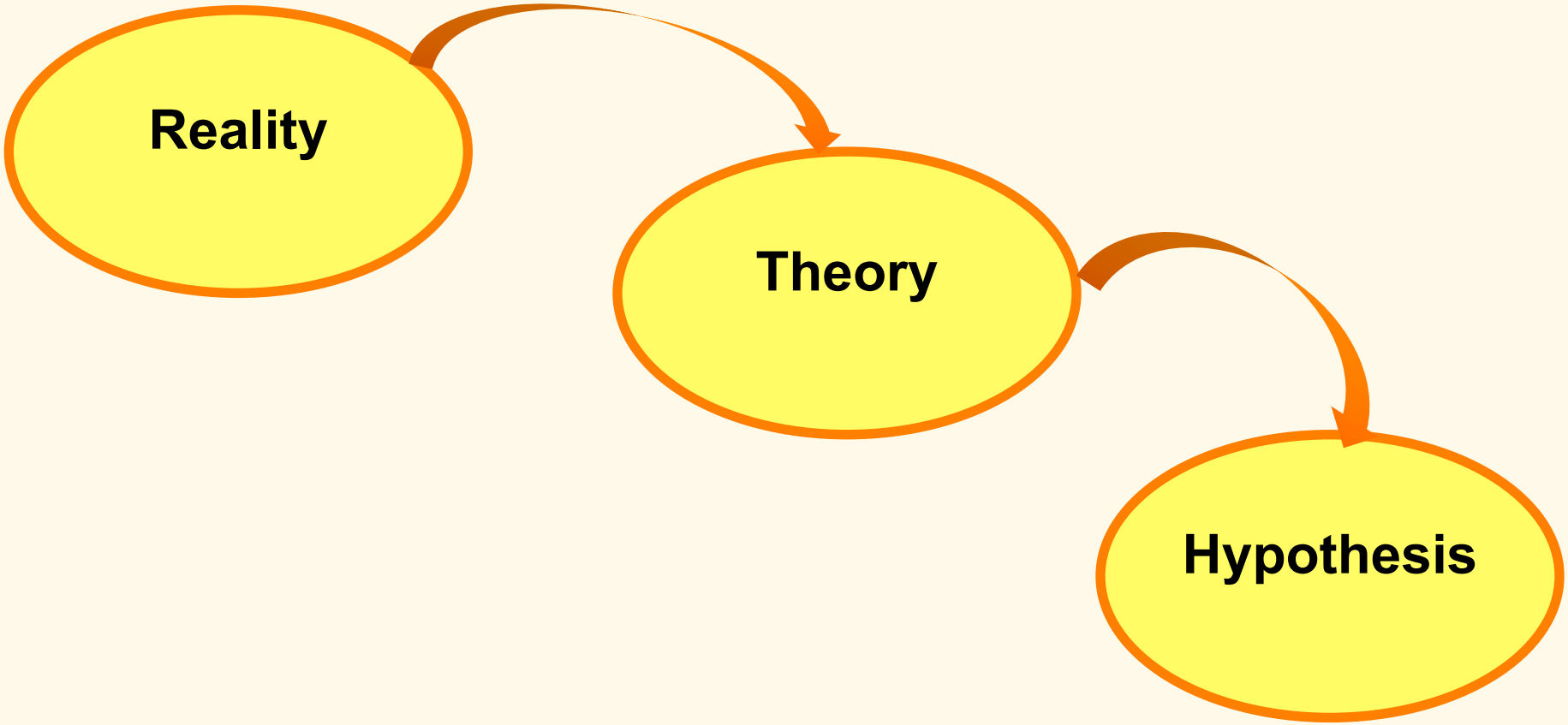


Theory

- **Idea:**
  - **Mental workload might be cause of reduction of fP3**

- **Workload can be induced through verb generation task**

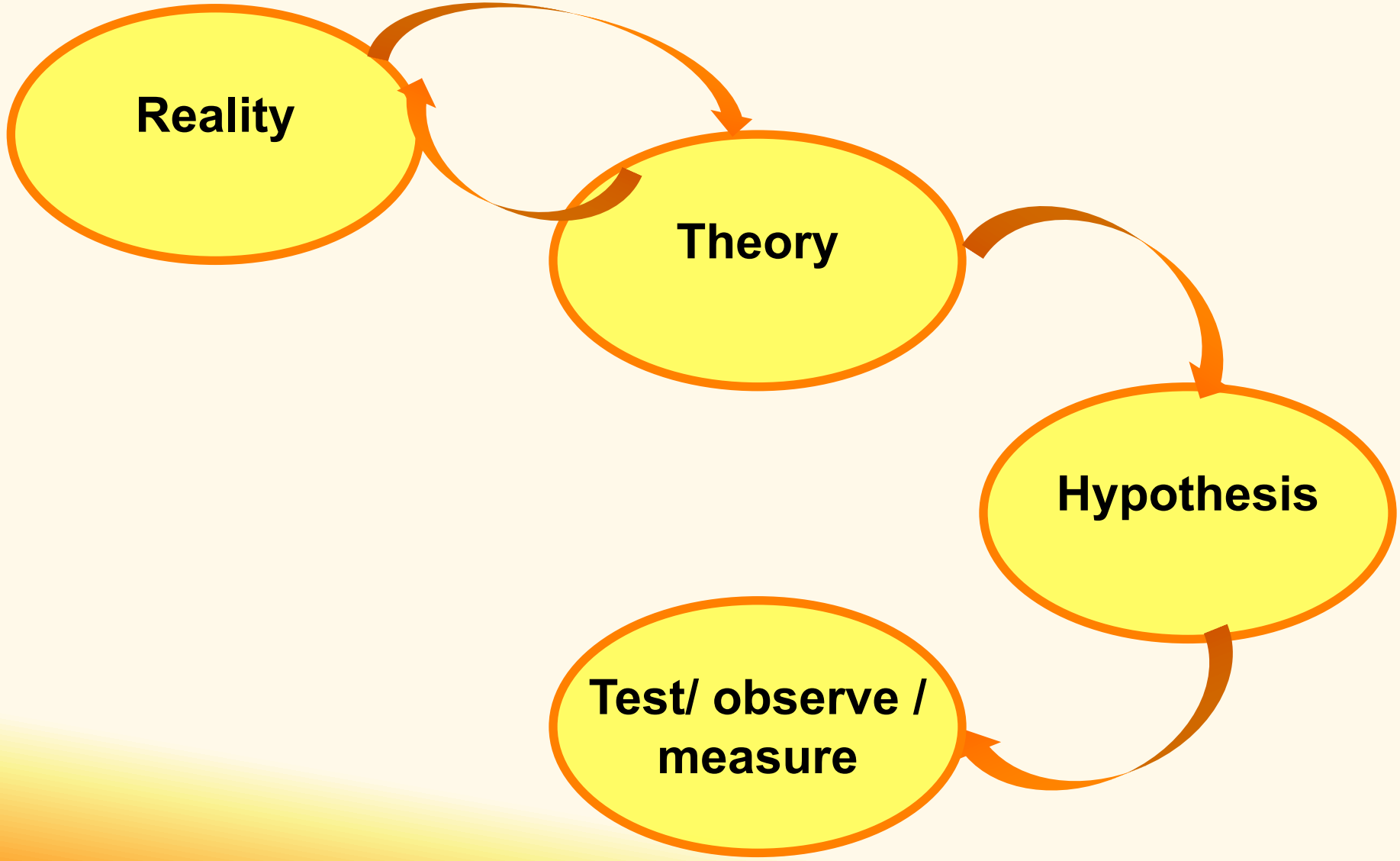
Step 3: Hypothesis



Hypothesis

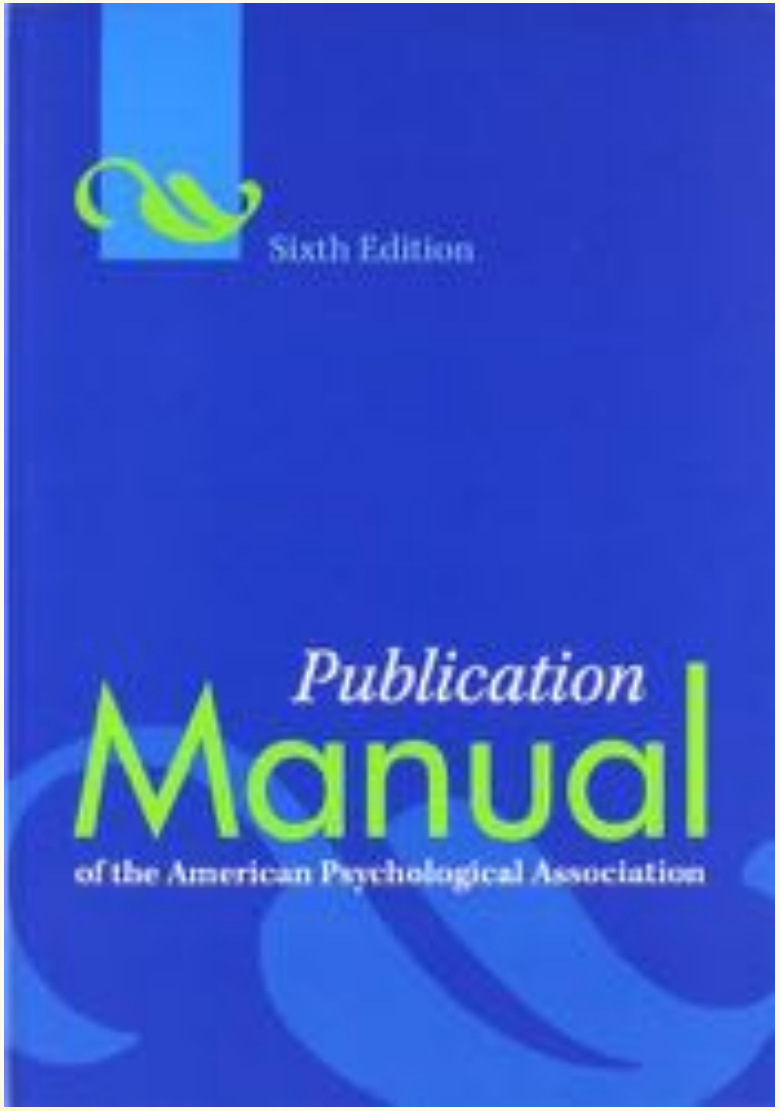
- **If fP3 reduction is due to mental workload:**
  - **fP3 should be reduced while thinking about a verb**
  - **This might depend on when we probe fP3 response ("How far is someone in the thinking process")**

Step 4: Test / observe / measure



Methods section of paper

- Participants
- Materials/Stimuli
- Design
- Procedure
- Measures



Participants

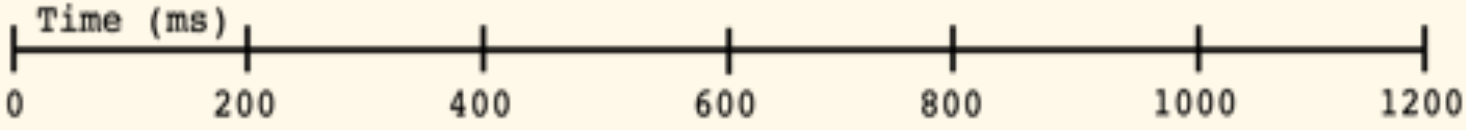
- 13 participants
- 8 M; 5 F
- Age: M = 23 years (SD = 2.6 years)
- Spoke Dutch
- Written, informed consent

Materials

- **Carefully selected list of nouns:**
  - **Roughly same length, short**
  - **Same “imaginability” score**
  - **Dutch**
  - **Spoken at same dB level by text-to-speech**
  - **Made same duration (speeded-up, slowed down): 400 ms**

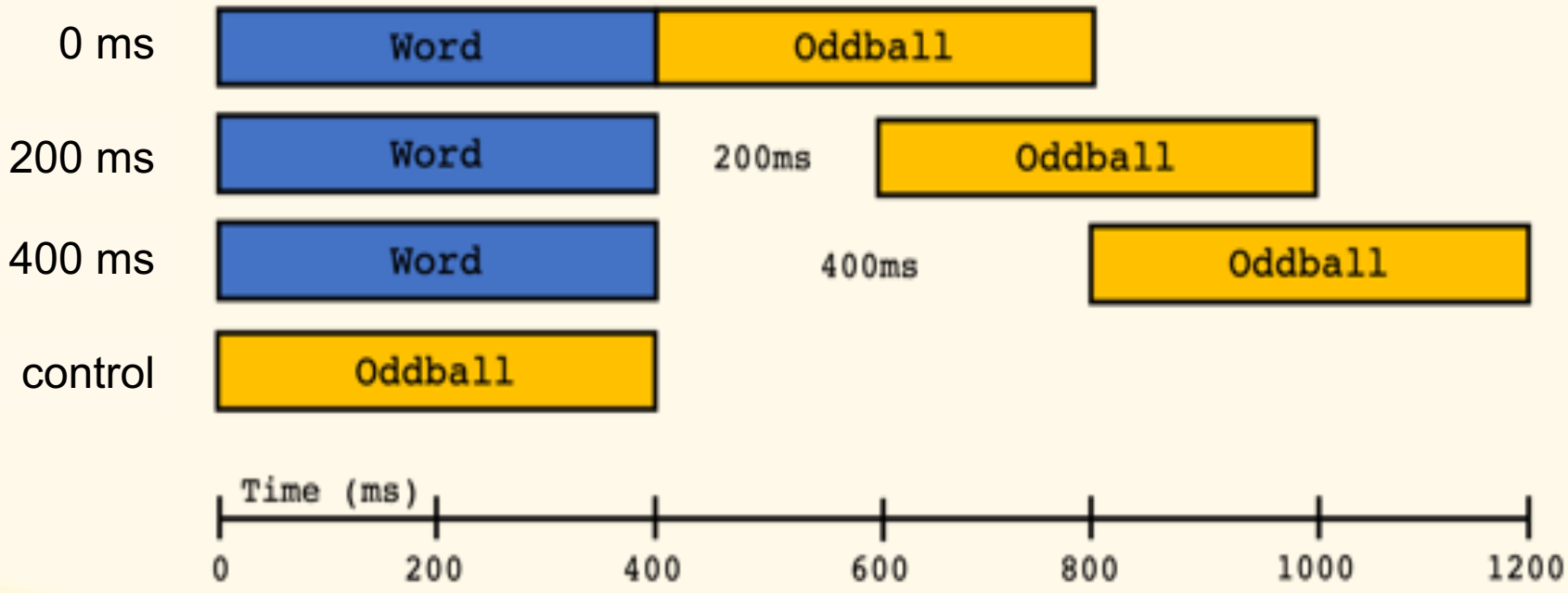
Design

- **Within-subjects manipulation: when is oddball probed relative to verb presentation?**



Design

- **Within-subjects manipulation: when is oddball probed relative to verb presentation?**

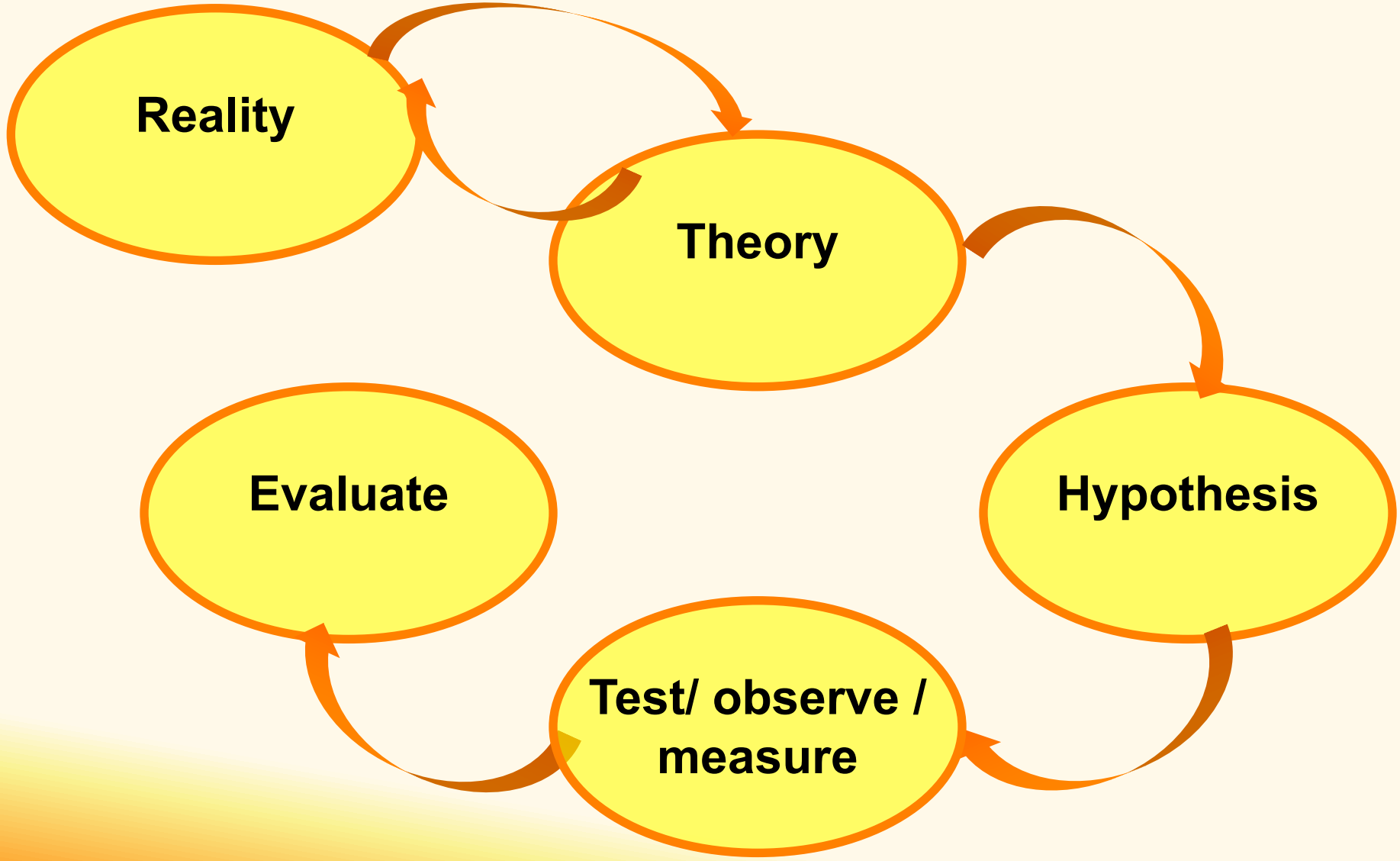


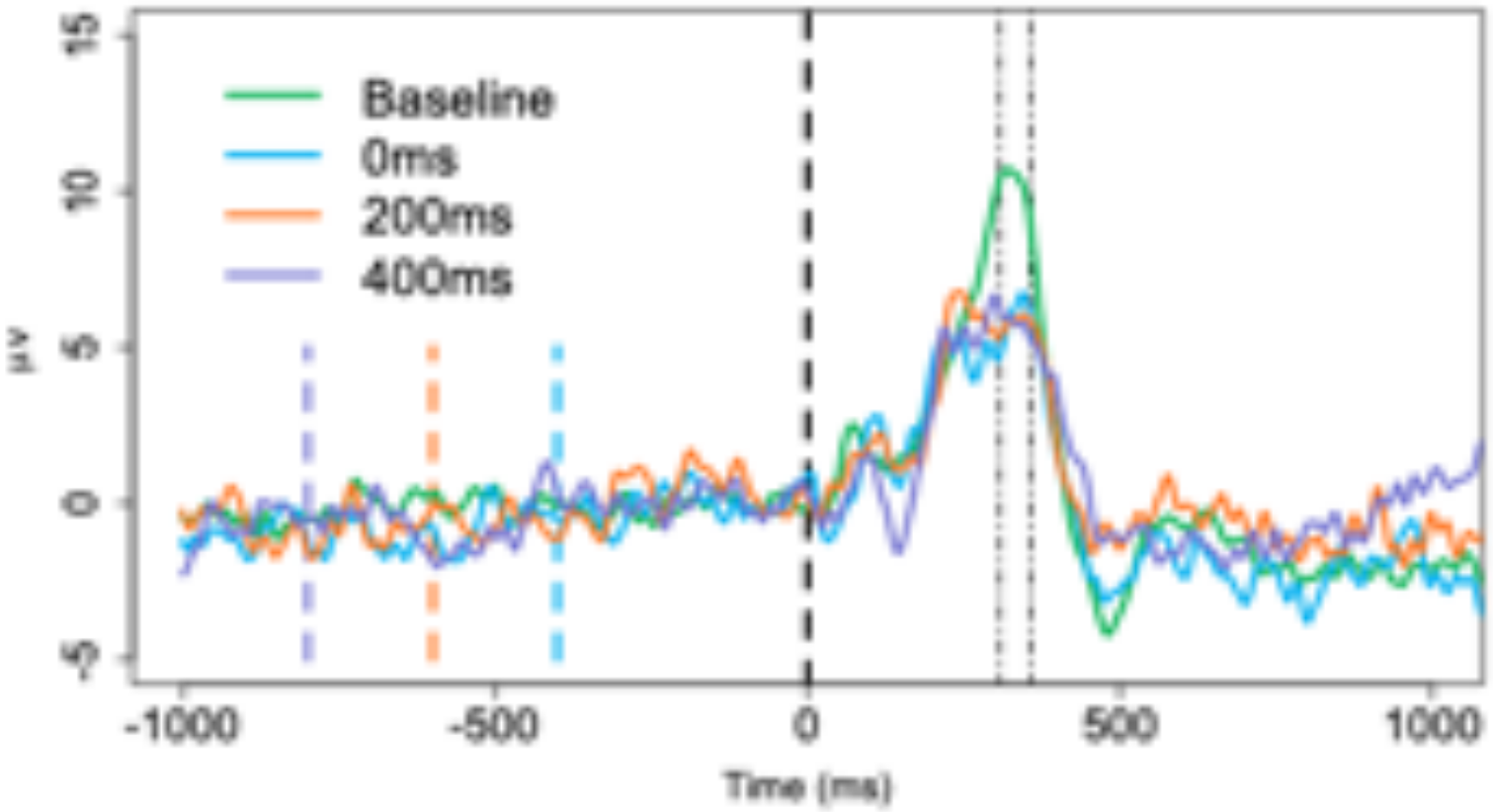
Procedure, measures

- 12 blocks
- Each block:
  - 80 oddball probes (64 standard, 16 novel)
  - 24 nouns (8 per ISI stimulus condition)

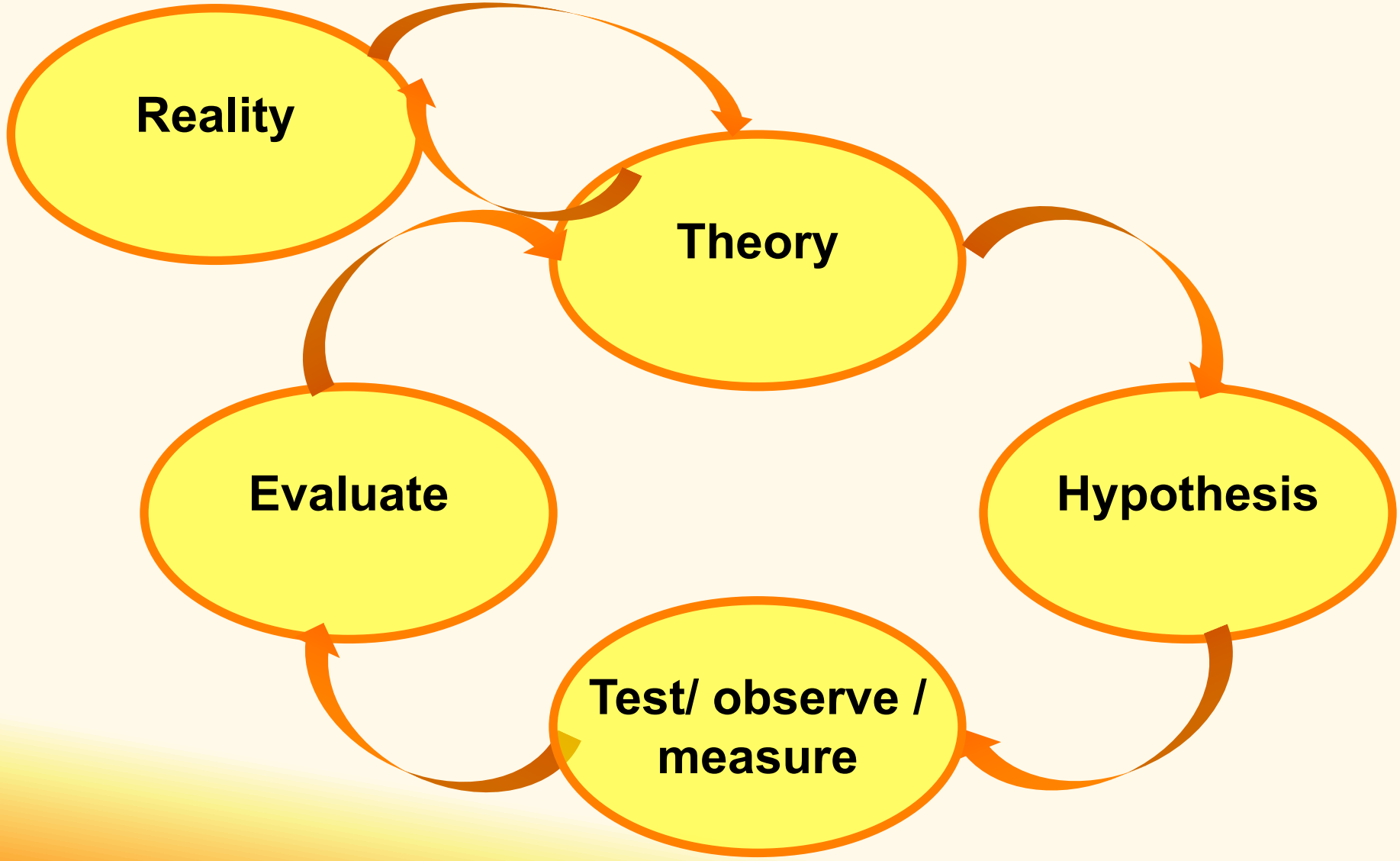
- Total novel probes per condition: 48

Step 5: Evaluate




Baseline
0ms
200ms
400ms
μV
15
10
5
0
-5
-1000
-500
0
500
1000
Time (ms)

Step 6: Relate result to theory and reality

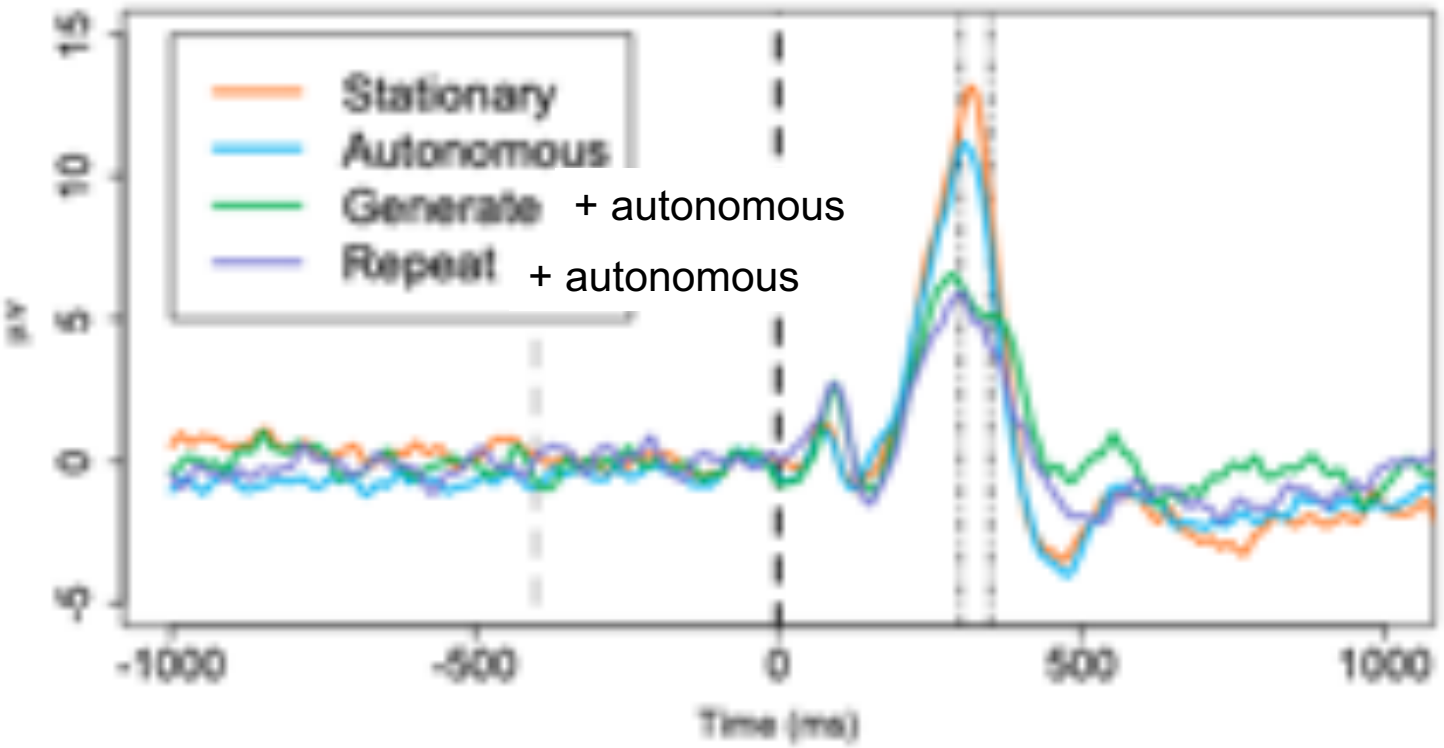


Interpretation

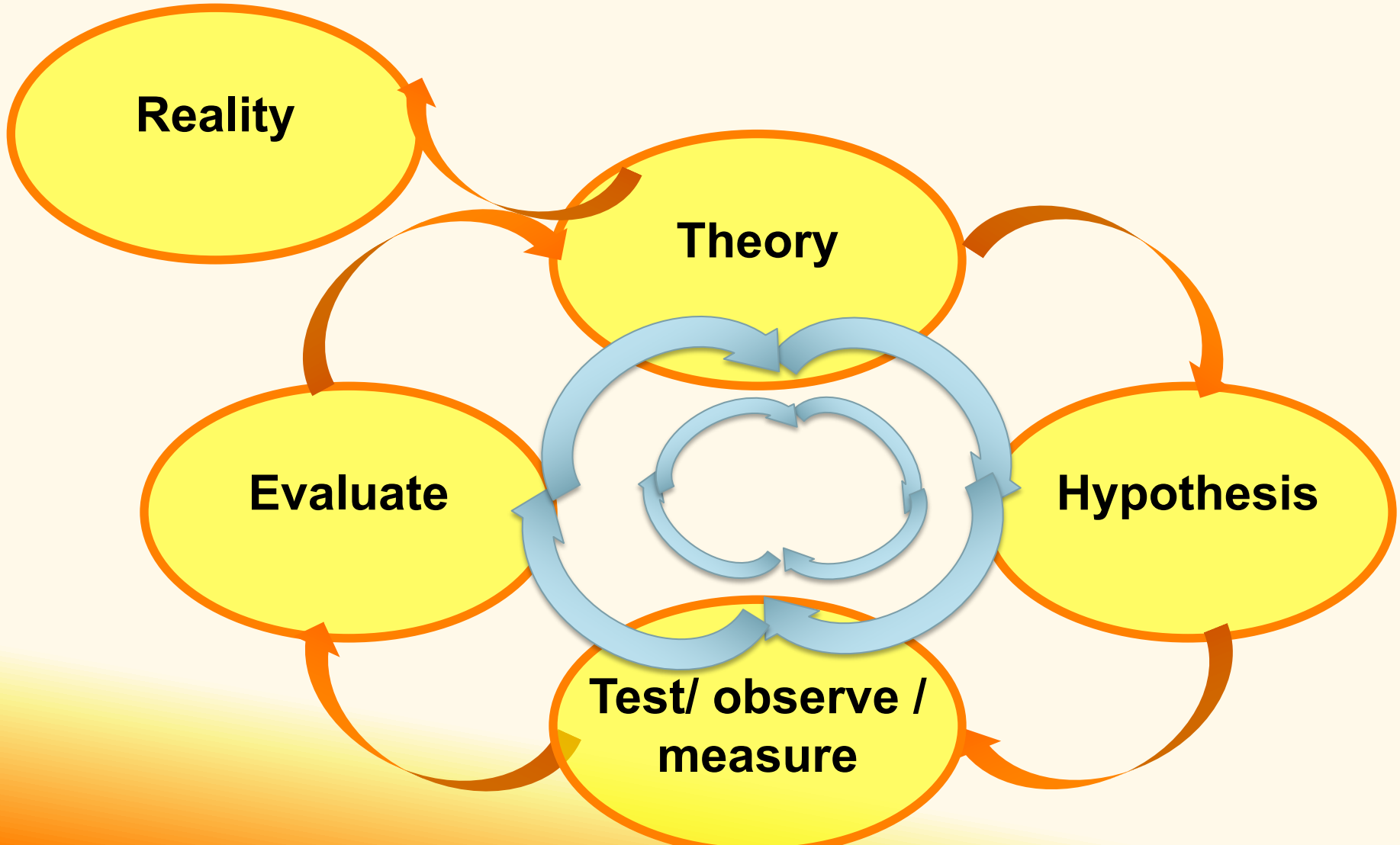
- **Mental distraction reduces susceptibility to alerts. No visual stimulation needed**

Next steps…

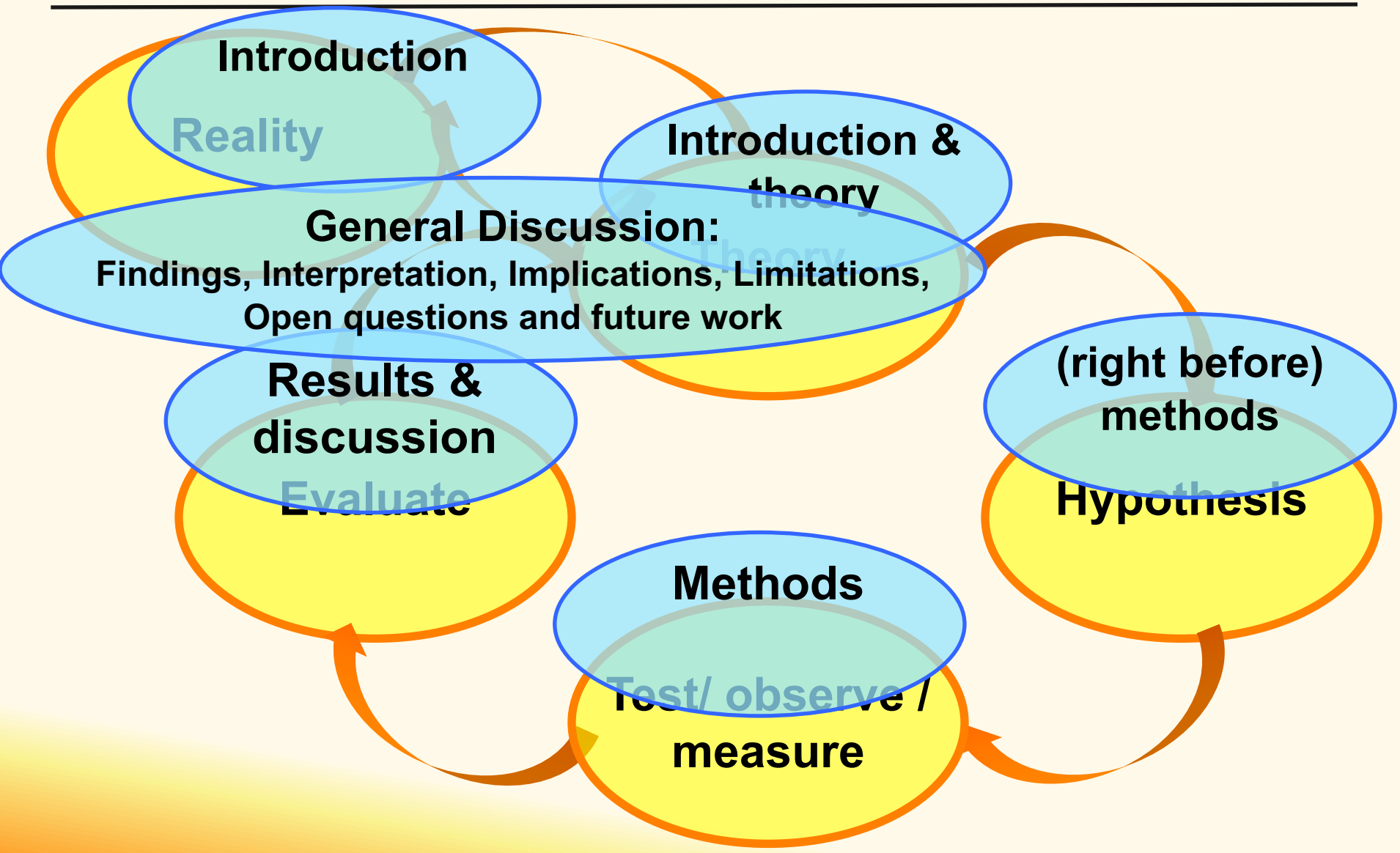
Interim results



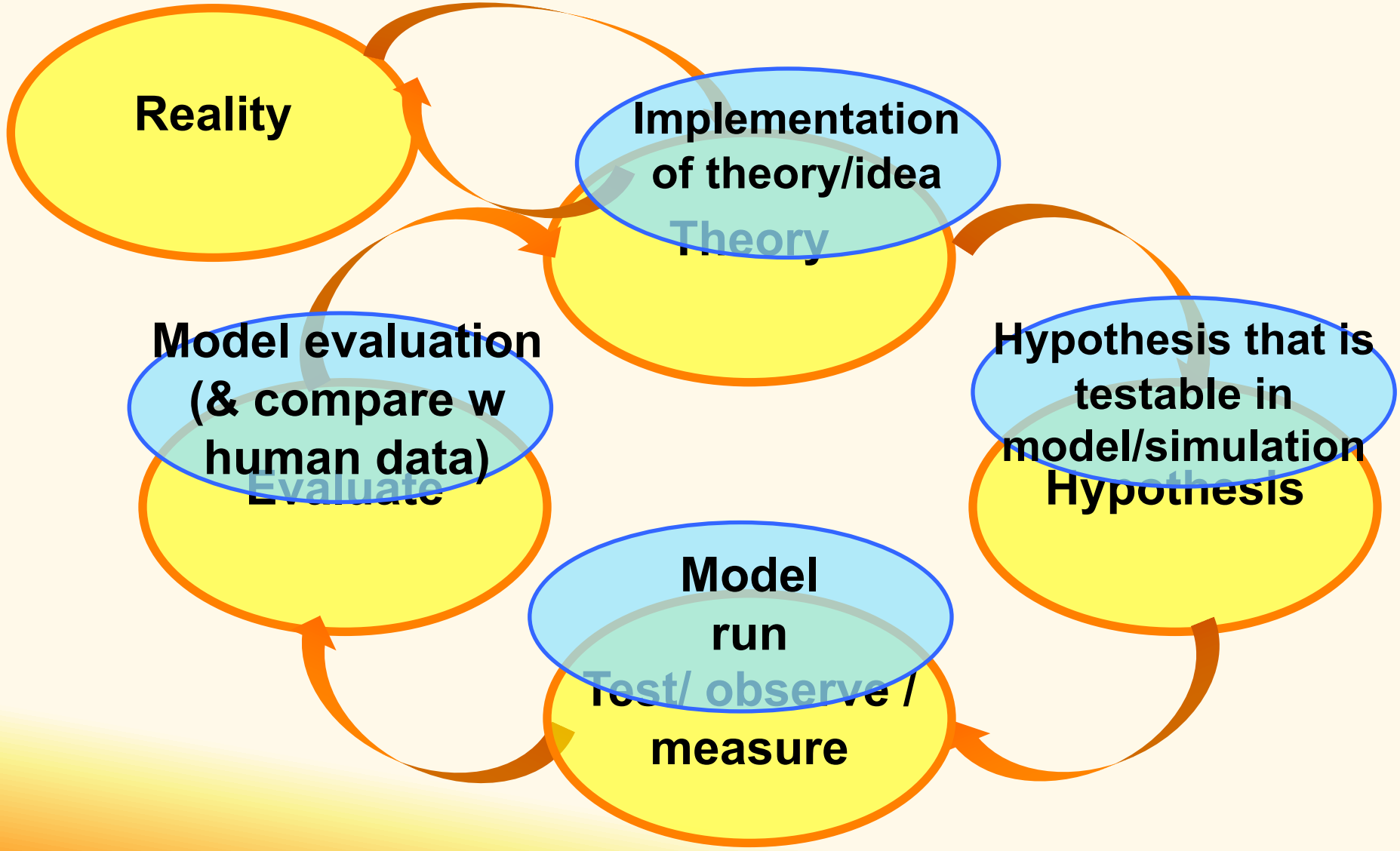
Iteratively go through cycle



Empirical Cycle: In report



Empirical Cycle: for AI professionals



Intermezzo: important terms

Important terms you should be able to use

- **Empirical cycle**
- **Manipulation:**
  - **Controlled change by experimenter to test *causal* effects.**
  - **Manipulate to test effect of *factor X* on *variable Y*.**

Important terms you should be able to use

- **Independent variable:**
  - **The one *controlled* through a manipulation**
  - **Described in *design* section of methods**
- **Dependent variable:**
  - **The one *measured*, and of which the value is dependent on manipulation of independent variable**
  - **Describe in *measures* section of methods.**
- **Experiments manipulate a *factor* (e.g., "type of driving"; independent variable) that has different *levels* (e.g., station, autonom., manual), and this manipulation might have a *measurable effect* on dependent variable**

Important terms you should be able to use

- **Within-subjects design**
  - **All participants experience all levels of a factor (typically in different orders)**
- **Between-subjects design**
  - **Different groups of participants experience different levels. Not everyone does everything**
  - **Example: A/B testing (think: Google website test)**
- **Mixed design:**
  - **Combination of within- and between- manipulations**

Important terms you should be able to use

- **Confounds:**
  - **Something that you failed to control and that co-varies with independent variable**
  - **Making it impossible to draw valid conclusions**
  - **Example:**
    “Who is the better teacher? Chris Janssen, Floris Bex, Dong Nguyen, Rosalie Iemhoff….”

4 type of validity (Cairns, 2016)

1. **Construct**
2. **Internal**
3. **External**
4. **Ecological**

More about details of experiment;
Relatively more objective assessment

More about relevance to real-world;
Relatively more subjective assessment

Empirical studies

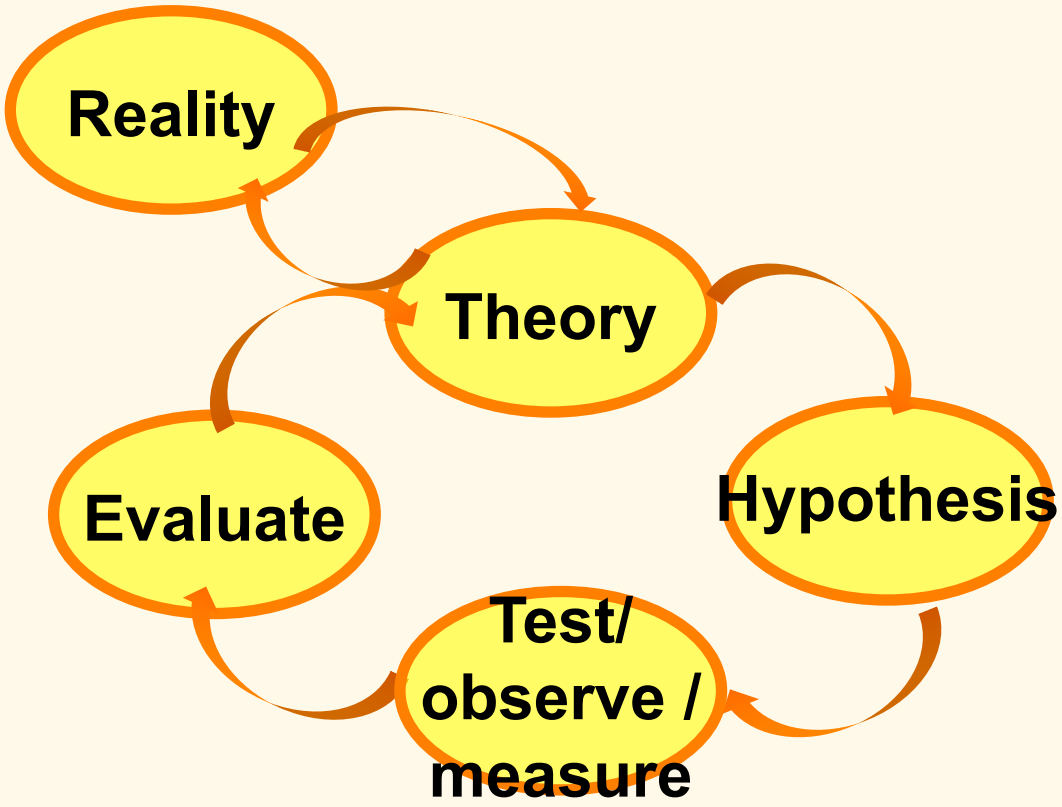
- More than just applicable to psychology..

Example 2: Robot vacuum cleaner

https://www.youtube.com/watch?v=A0Z79ycisDU

Map-based robot

Which vacuum is best: world-based (Chris) or map/model-based (Remo)?



- Participants
- Materials/Stimuli
- Design
- Procedure
- Measures

https://www.youtube.com/watch?v=H9DM586ANhA

Example 3: (testing) Game design

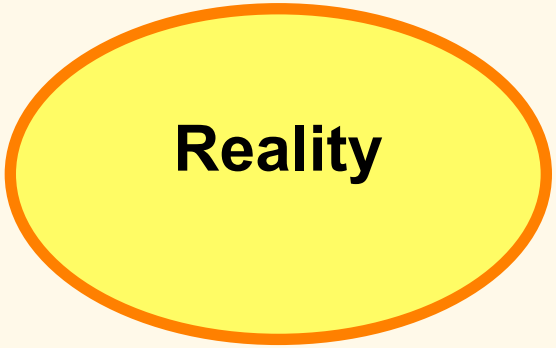
Al Zayer, Tregillus, and Folmer (2016). PAWdio: Hand Input for Mobile VR using Acoustic Sensing. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play* (CHI PLAY '16). ACM, New York, NY, USA, 154-158. DOI: https://doi.org/10.1145/2967934.2968079

(Winner best note at CHI Play 2016)

Example 3: (testing) Game design

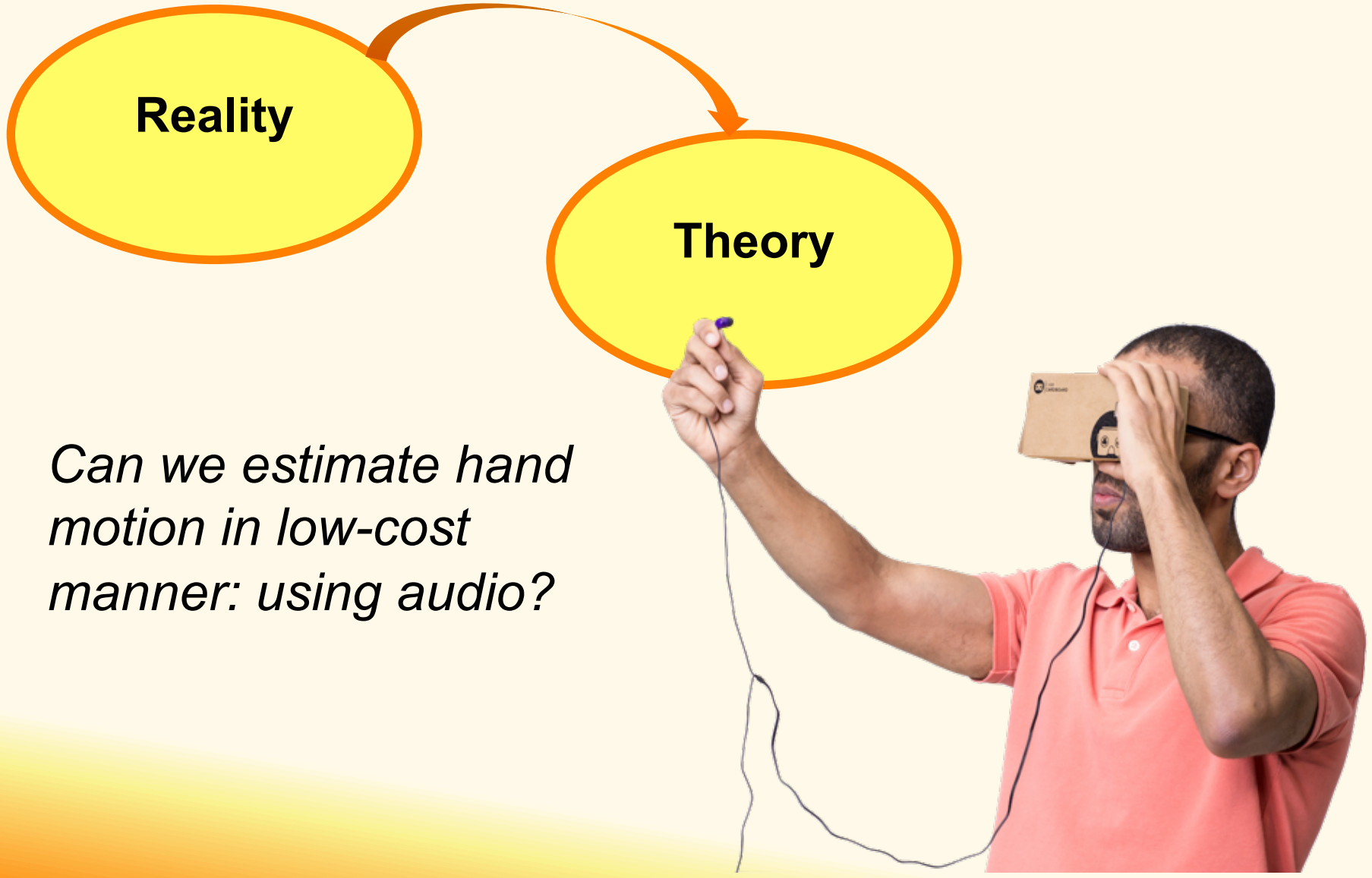
https://dl.acm.org/citation.cfm?doid=2967934.2968079

Step 1: Reality



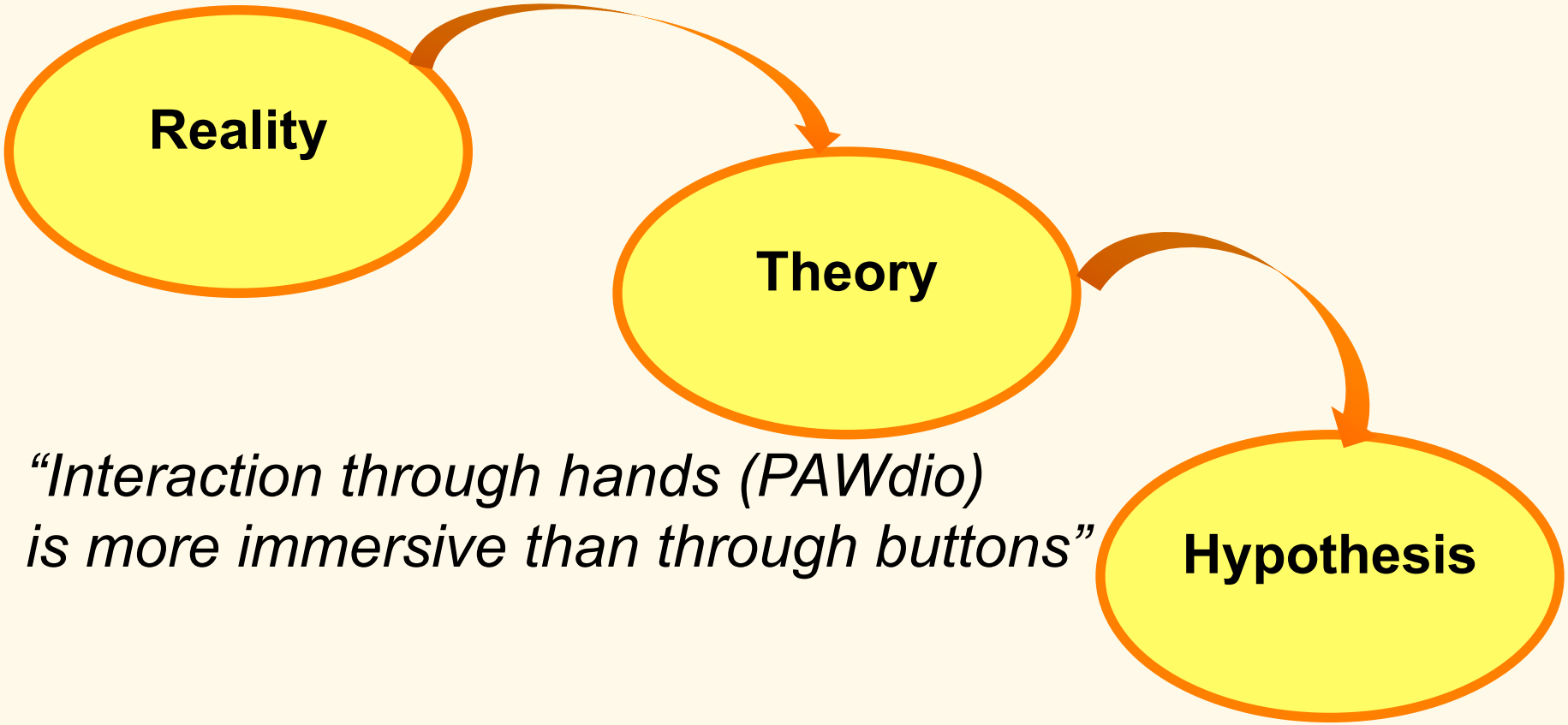
Can devices to estimate distance and motion be made cheaper while remaining successful?
For example, to use on phone?

Step 2: Theory



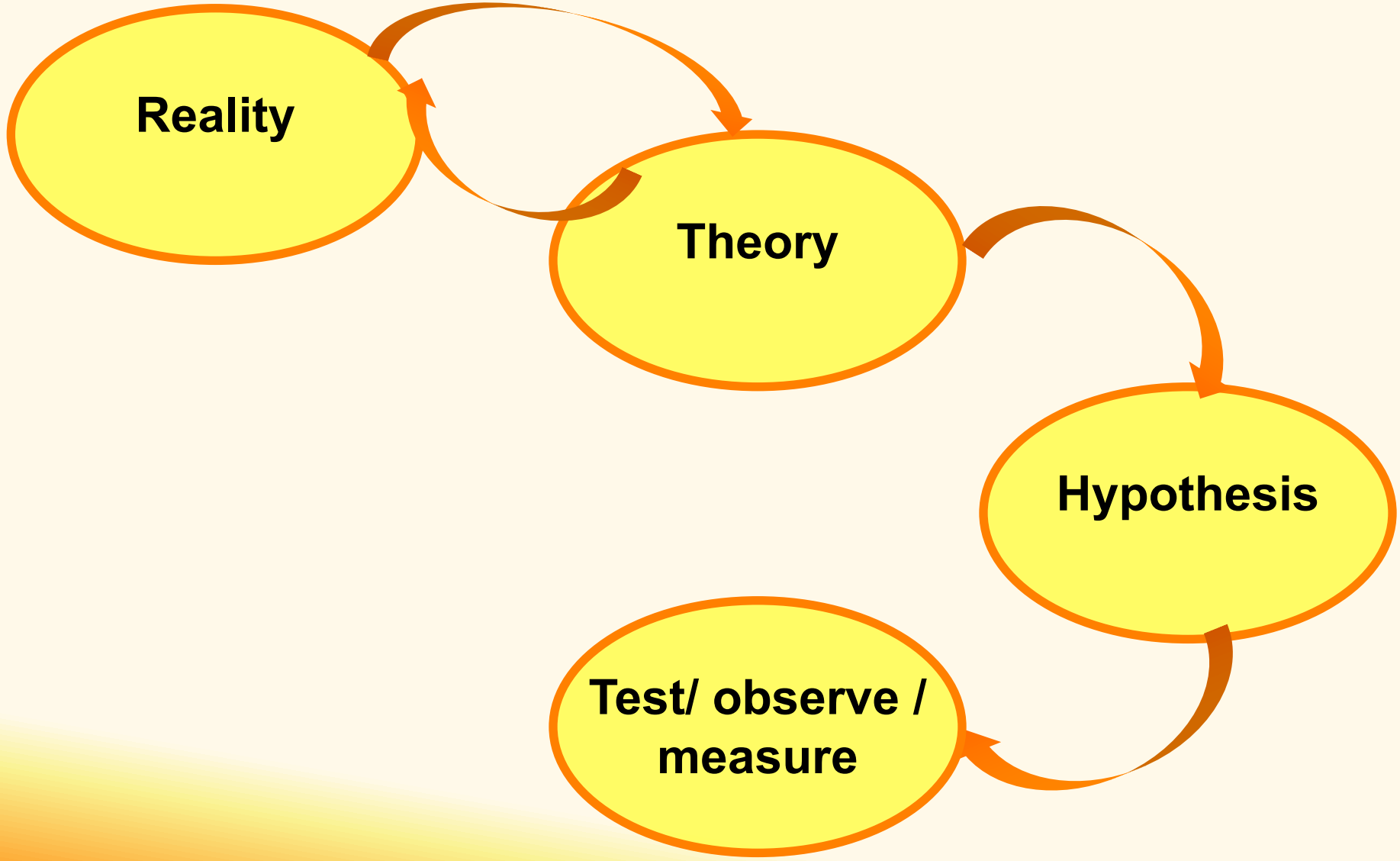
Can we estimate hand motion in low-cost manner: using audio?

Step 3: Hypothesis



“Interaction through hands (PAWdio) is more immersive than through buttons”

Step 4: Test / observe / measure



Step 4: Test / observe / measure

- Participants: 18 (see paper for details)
- Materials: apple grabbing game
- Design: within-subjects (PAWdio vs button)
- Procedure:
  - Consent: Unknown
  - Tutorial. Instruction PAWdio: “move hand slow”
  - Then experimental conditions (order unknown).
  - 30 minutes total

Step 5: Evaluate

Reality

Evaluate

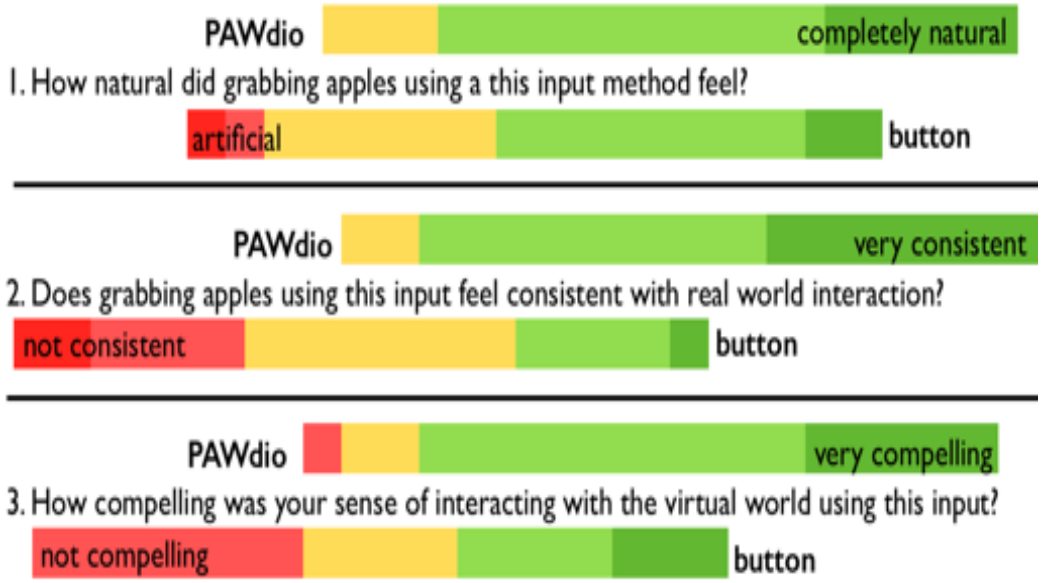
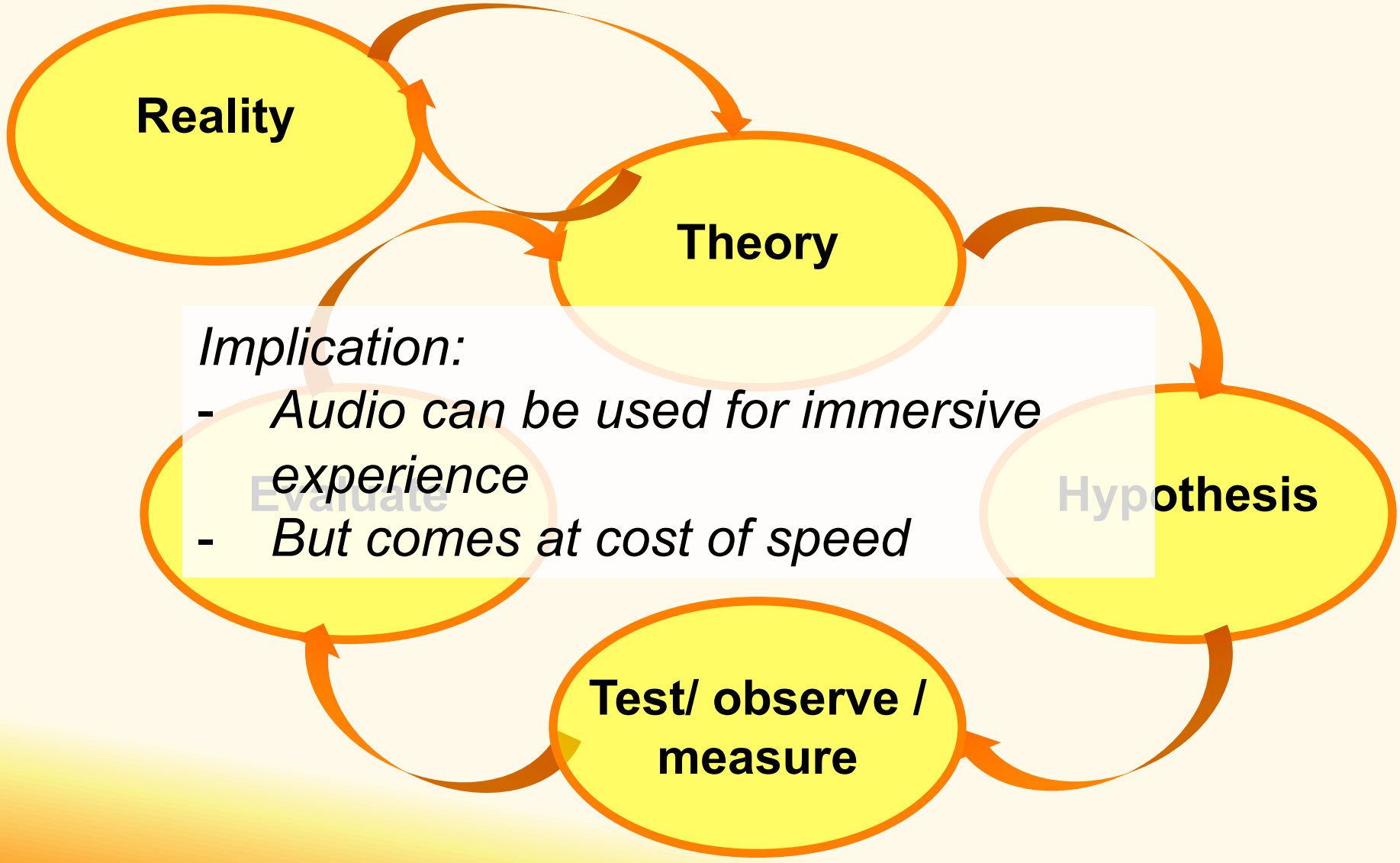


Figure 3. Evaluation of immersion for both input techniques

Measuring:

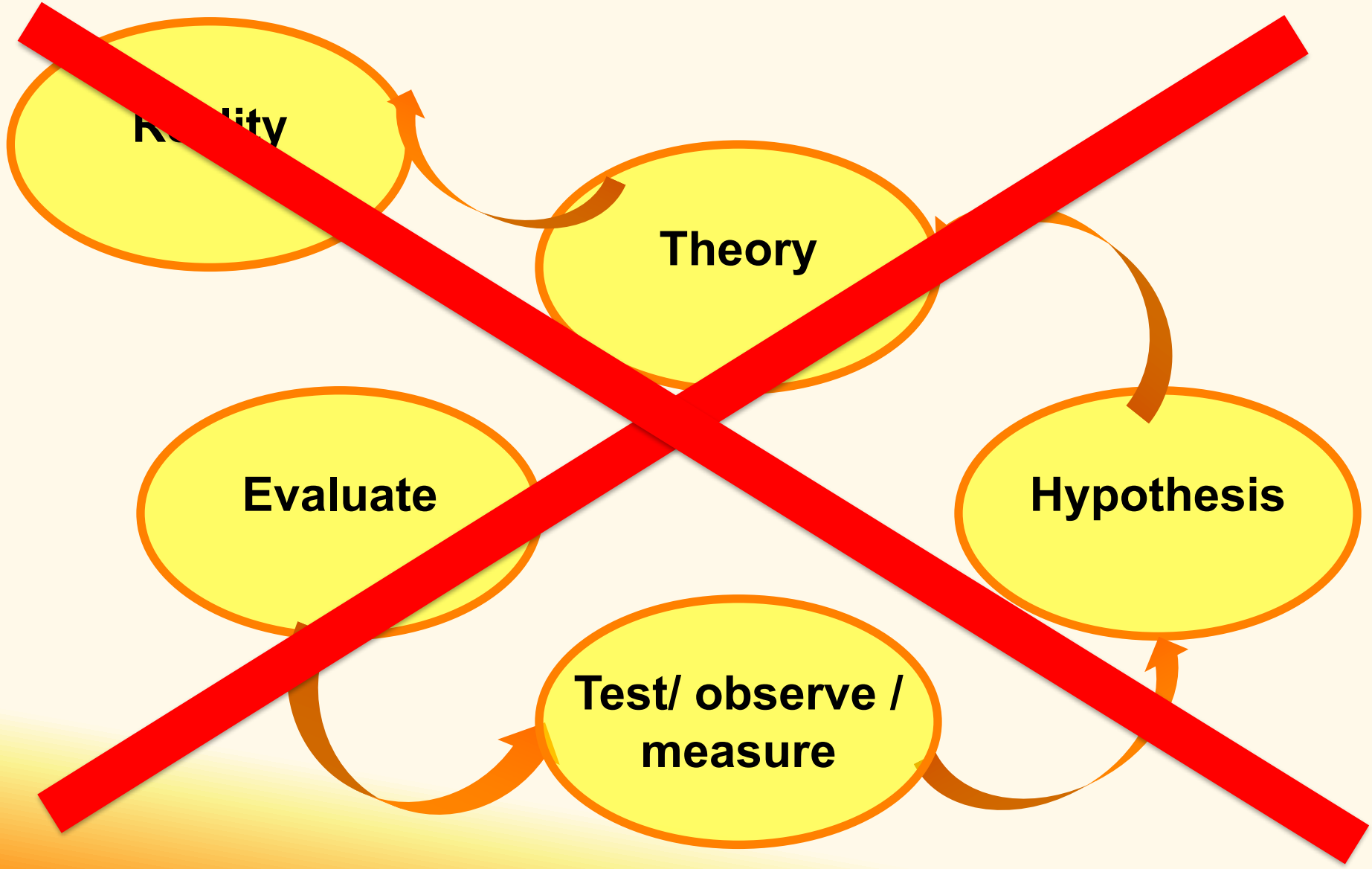
- *Performance metrics (see paper: PAWdio's is slower)*
- *Immersion and subjective experience*

Step 6: Relate result to theory and reality



Example 4: Big data

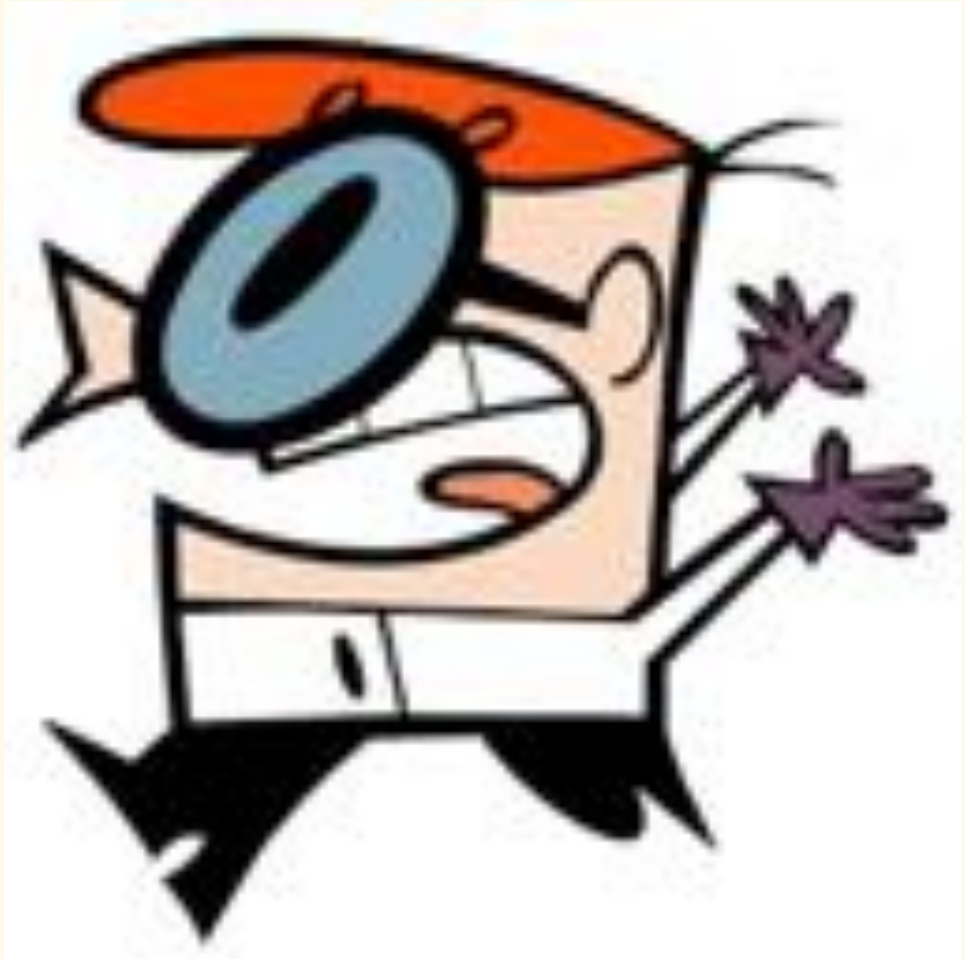
Where some big data research goes wrong…

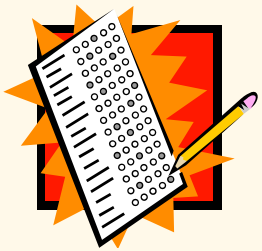


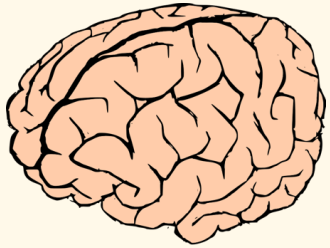
Why is this (most of the time) wrong

- Data is collected without a reason WHY this should be done
  - At cost of privacy, data storage (energy costs), time
- Not clear whether data is best measure
  - Construct validity at risk
- Statistical techniques sometimes incorrectly applied
  - Hypothesis testing requires an *a priori* hypothesis
  - Hypothesis sometimes presented as if it was there all along
- Net result: findings are presented as theory based, whereas they should be presented as ***exploratory***

Relevant for your thesis research….

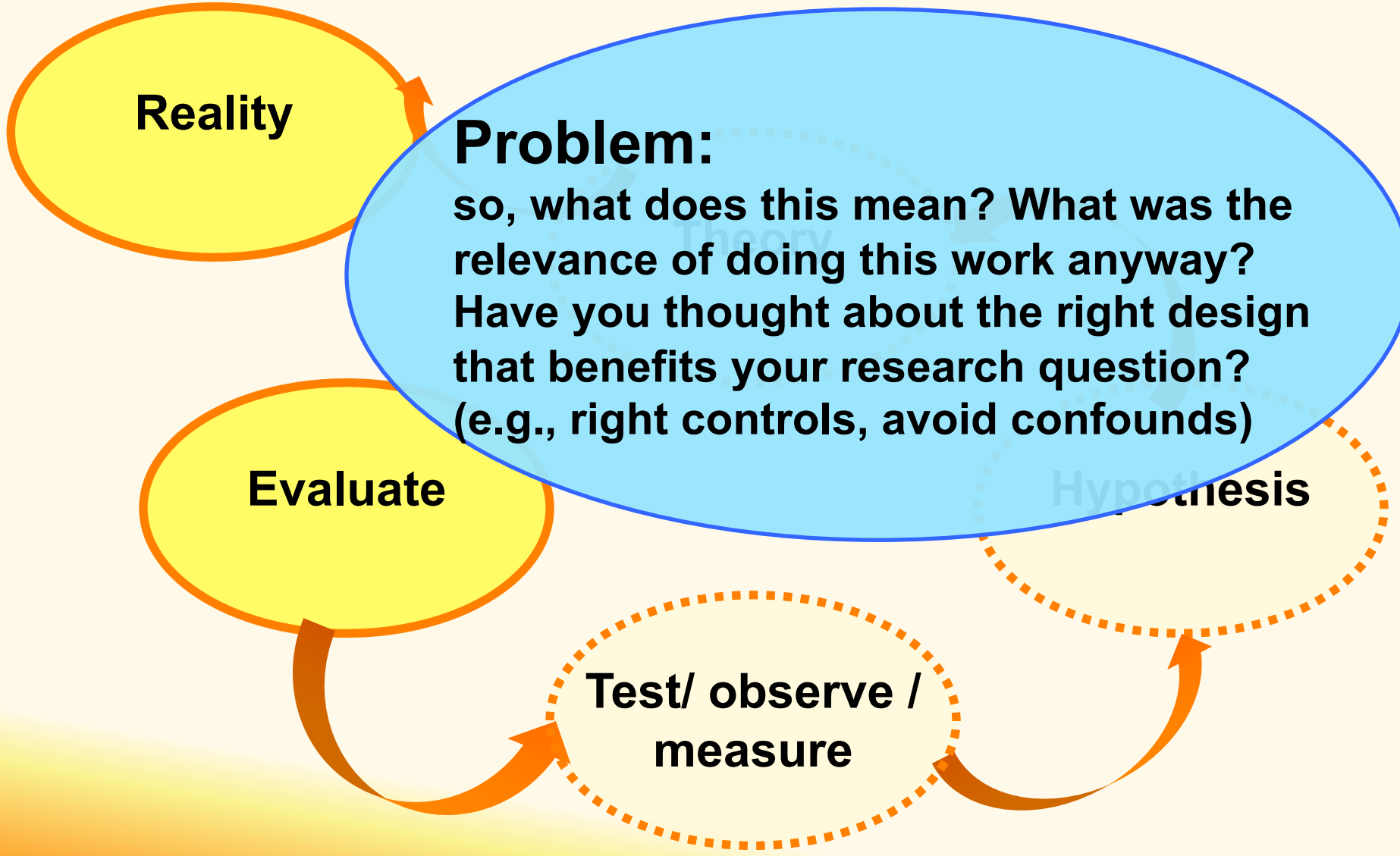




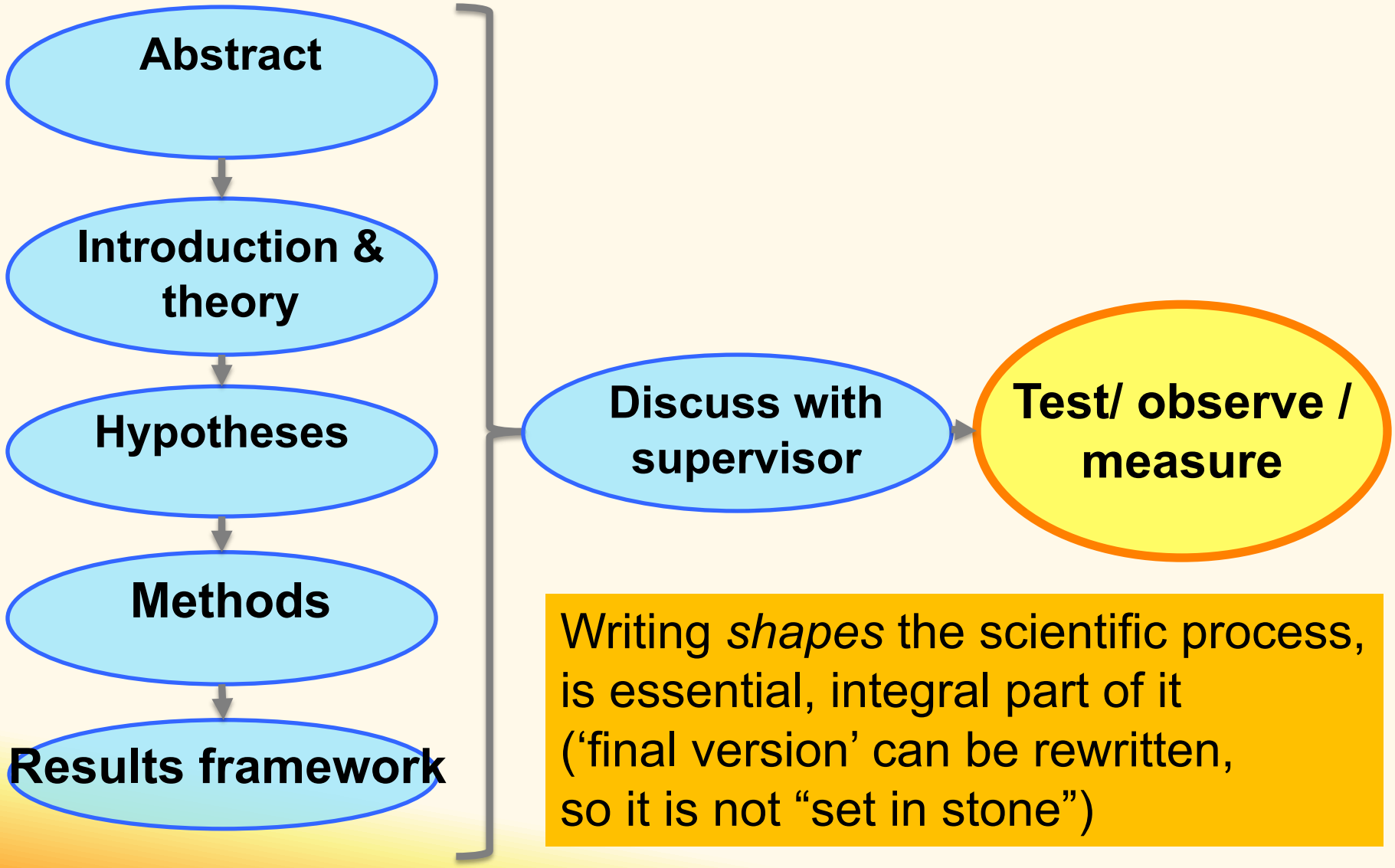


“I collected my data.
How do I now analyze these?”

Why you are in trouble if you don't know how to analyze data later…



Cairns' solution: write first (see article)



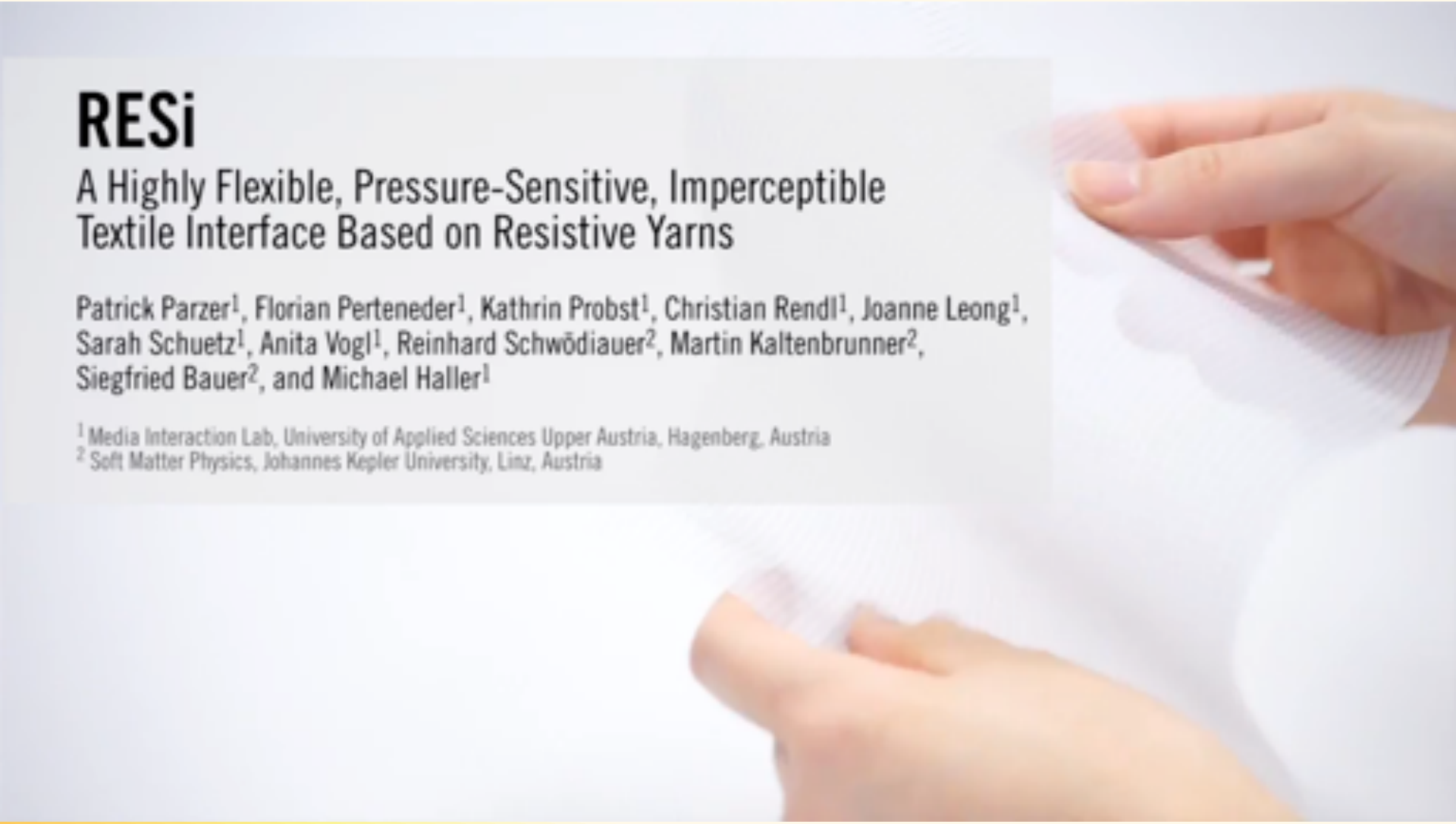
Writing *shapes* the scientific process, is essential, integral part of it ('final version' can be rewritten, so it is not "set in stone")

Example 5: Product design

http://mi-lab.org/projects/resi/
Parzer, Perteneder, Probst, Rendl, Leong, Schuetz, Vogl, Schwoediauer, Kaltenbrunner, Bauer, and Haller (2018). RESi: A Highly Flexible, Pressure-Sensitive, Imperceptible Textile Interface Based on Resistive Yarns. In *The 31st Annual ACM Symposium on User Interface Software and Technology* (UIST '18). ACM, New York, NY, USA, 745-756. DOI: https://doi.org/10.1145/3242587.3242664

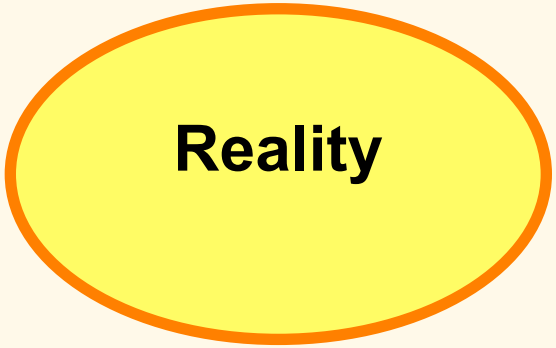
(winner best paper at UIST 2018)

Example 5: Product design



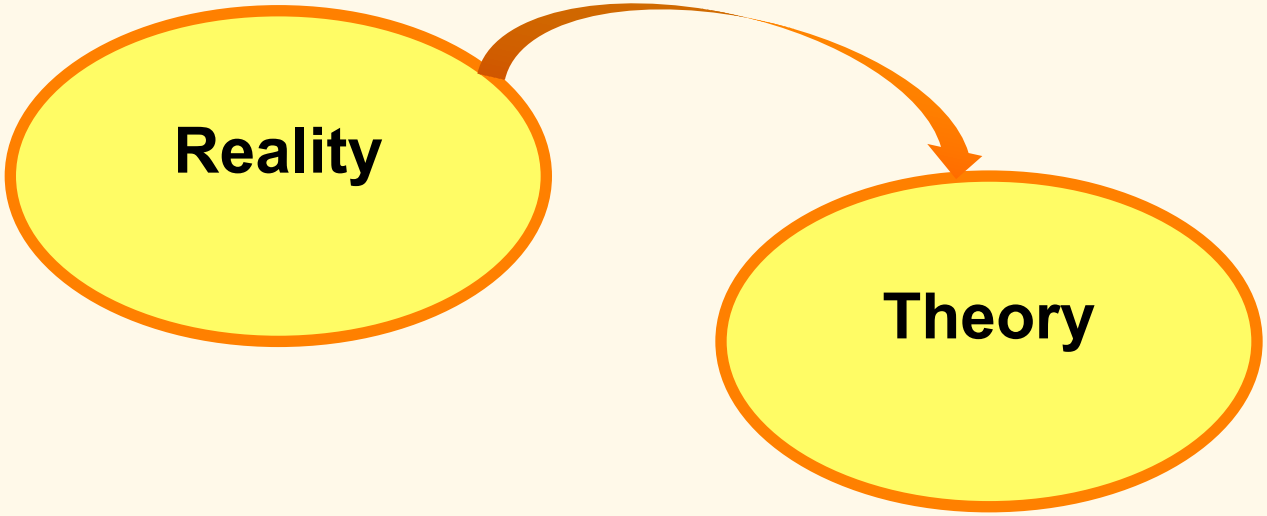
http://mi-lab.org/projects/resi/

Step 1: Reality



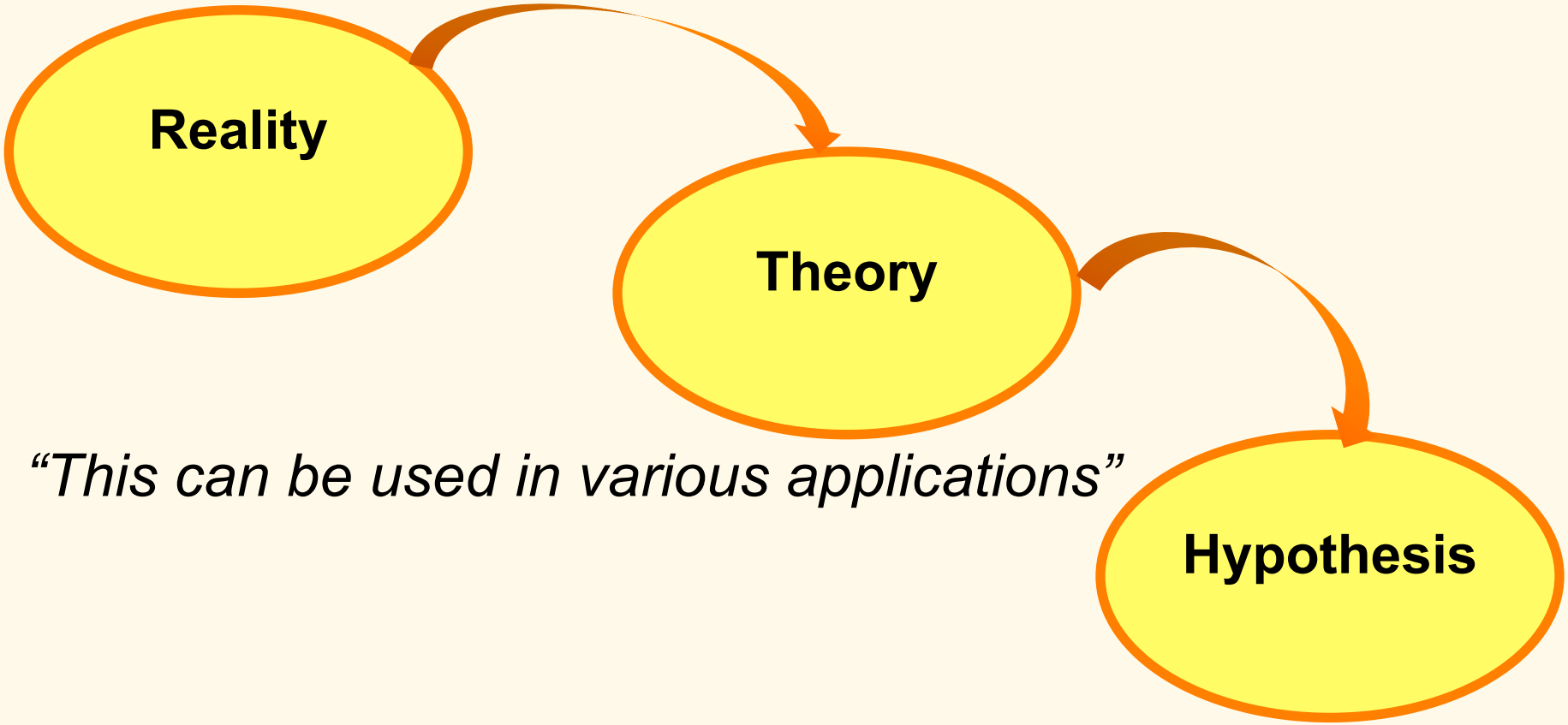
Why can't we control machines from things we already wear?

Step 2: Theory



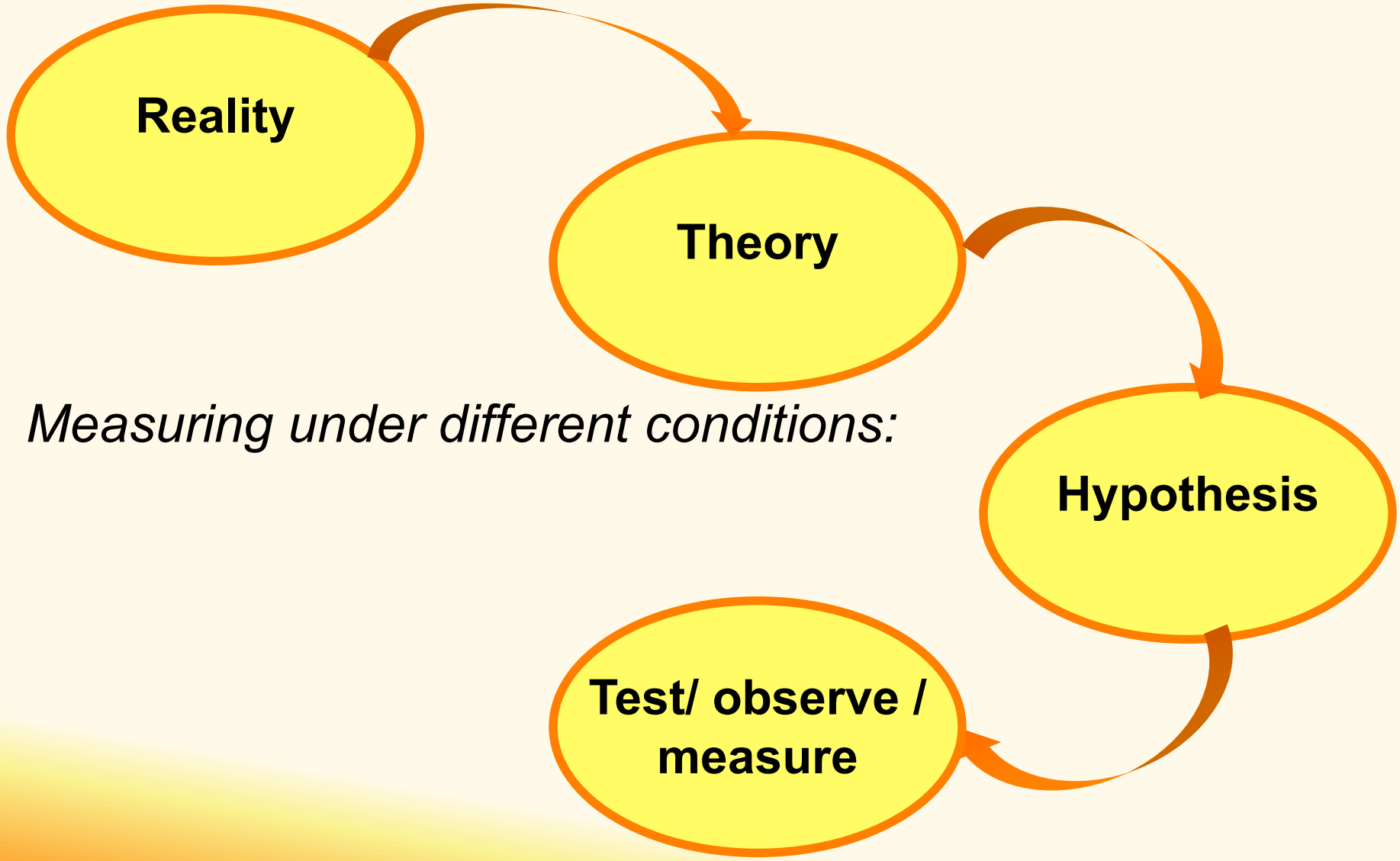
Threads can be made that conduct electricity
(lots of technical details on conditions under which it might work)

Step 3: Hypothesis

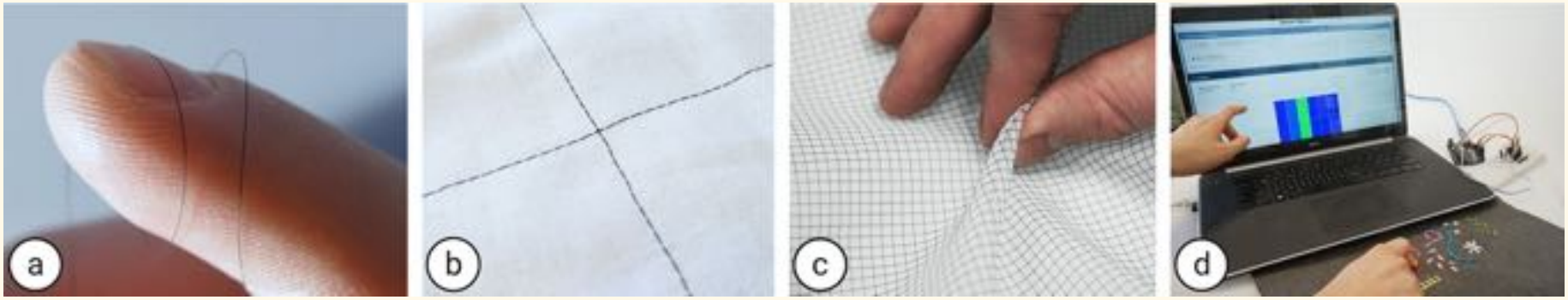


"This can be used in various applications"

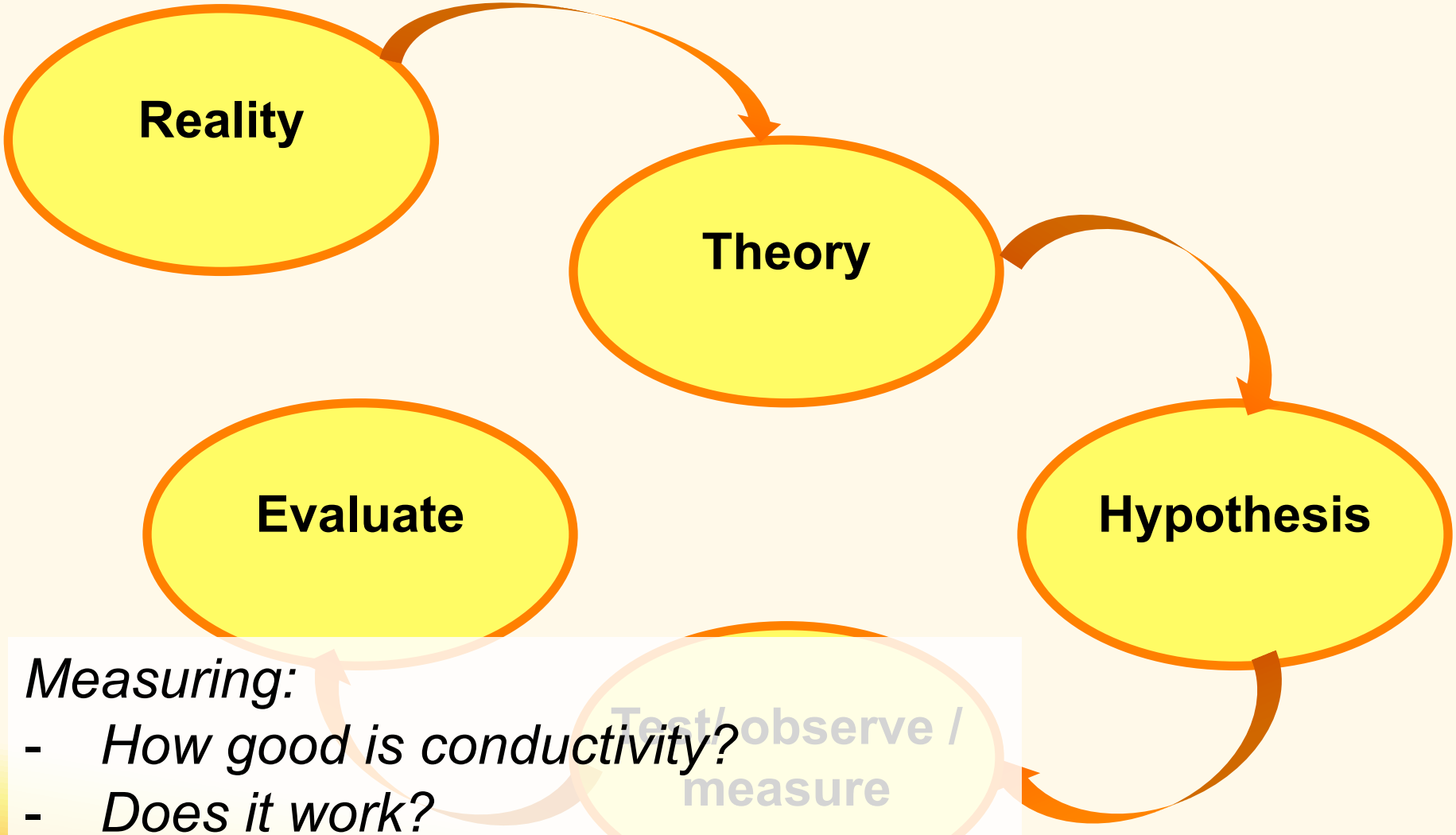
Step 4: Test / observe / measure



Measuring under different conditions:


a
b
c
d

Step 5: Evaluate



Measuring:

- *How good is conductivity?*
- *Does it work?*
- *When exactly, how? Details!*

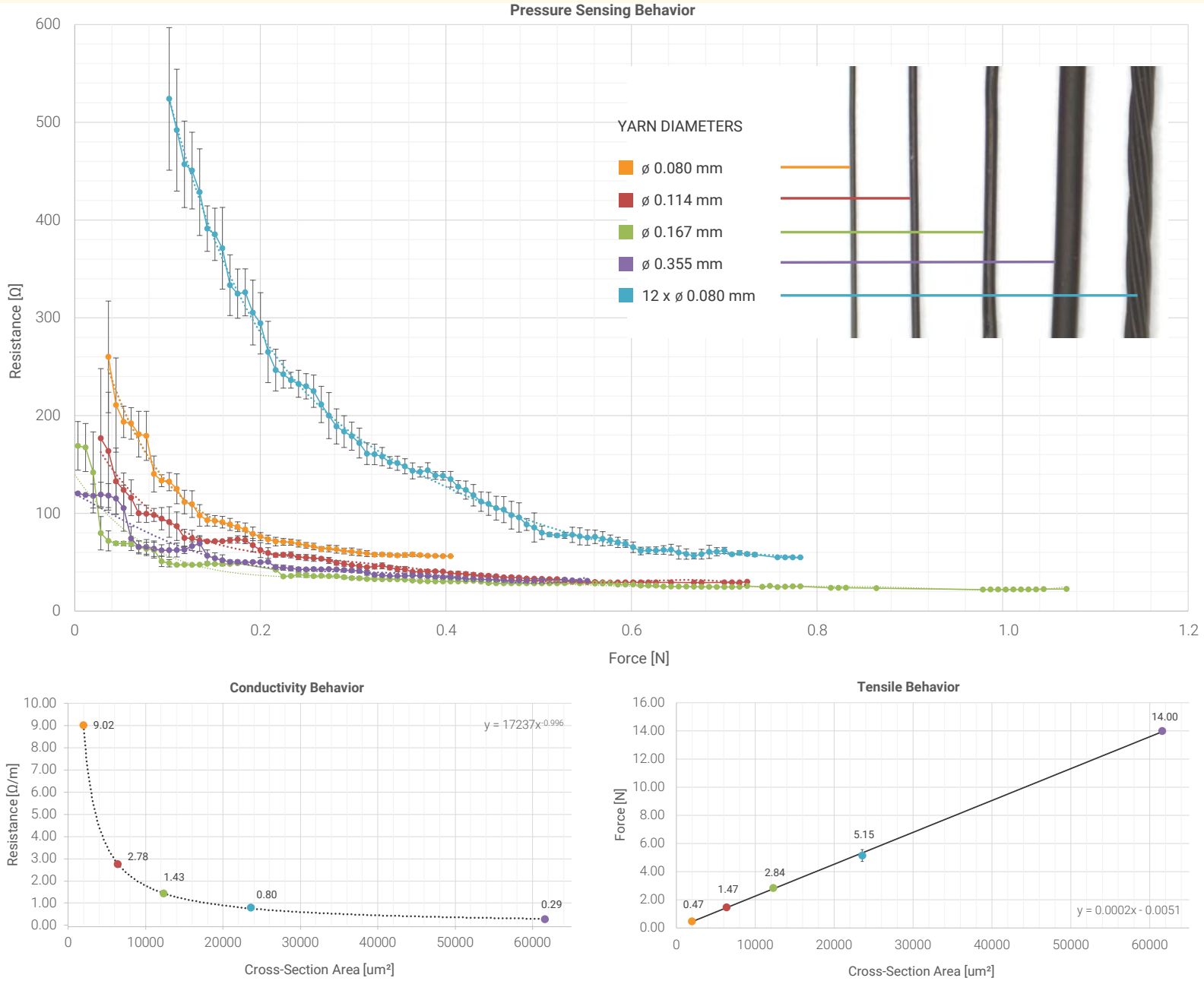
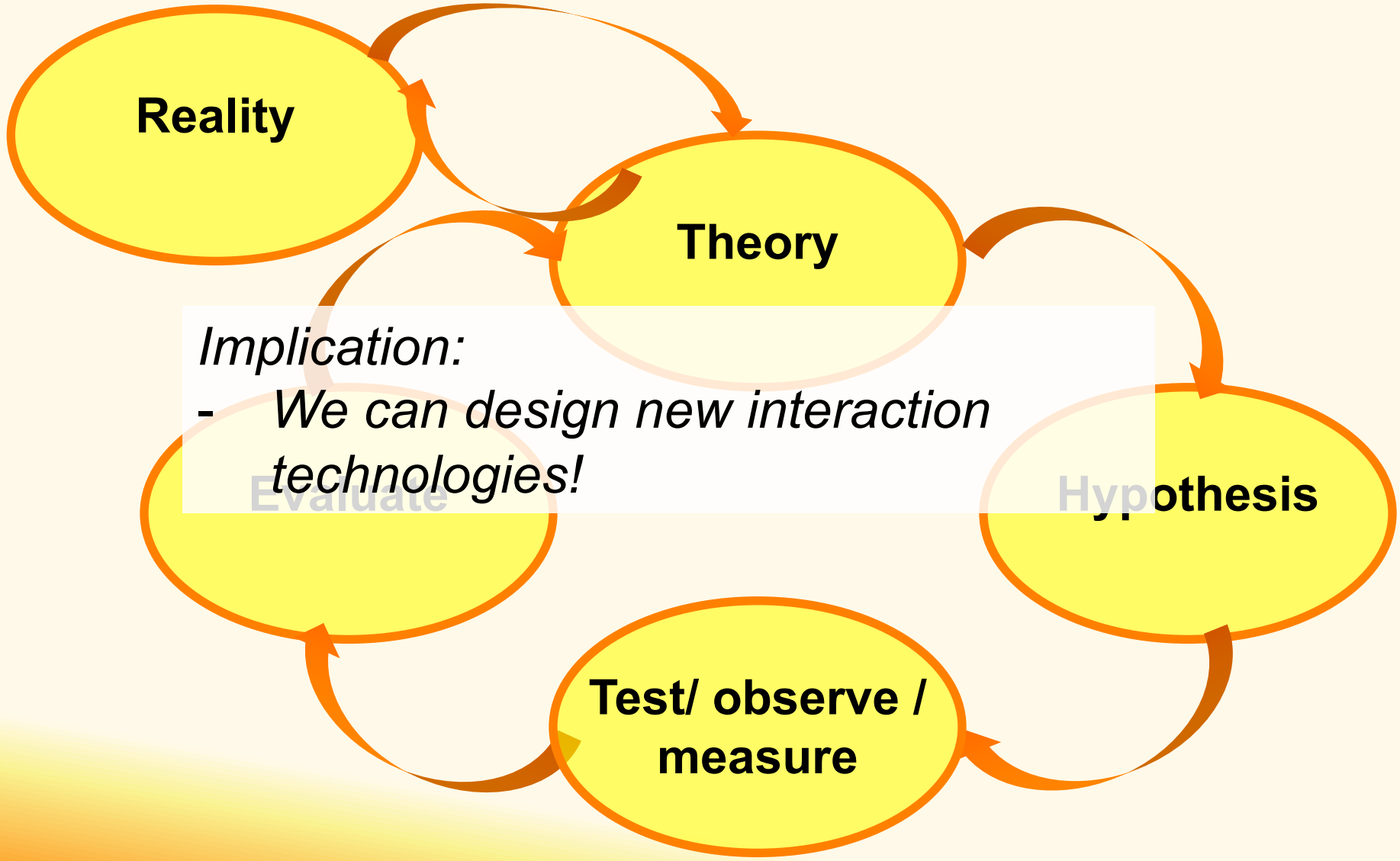


Figure 5: Pressure Sensing Behavior (top) of pairs of resistive yarns pressed perpendicular against each other. Conductivity Behavior (bottom left) shows that the resistance of the yarns correlates with the cross-section area of the metallic core. Tensile Behavior (bottom right) shows that the average fracture force correlates with the metal core cross-section area of the yarn.

Step 6: Relate result to theory and reality



Implication:

- *We can design new interaction technologies!*

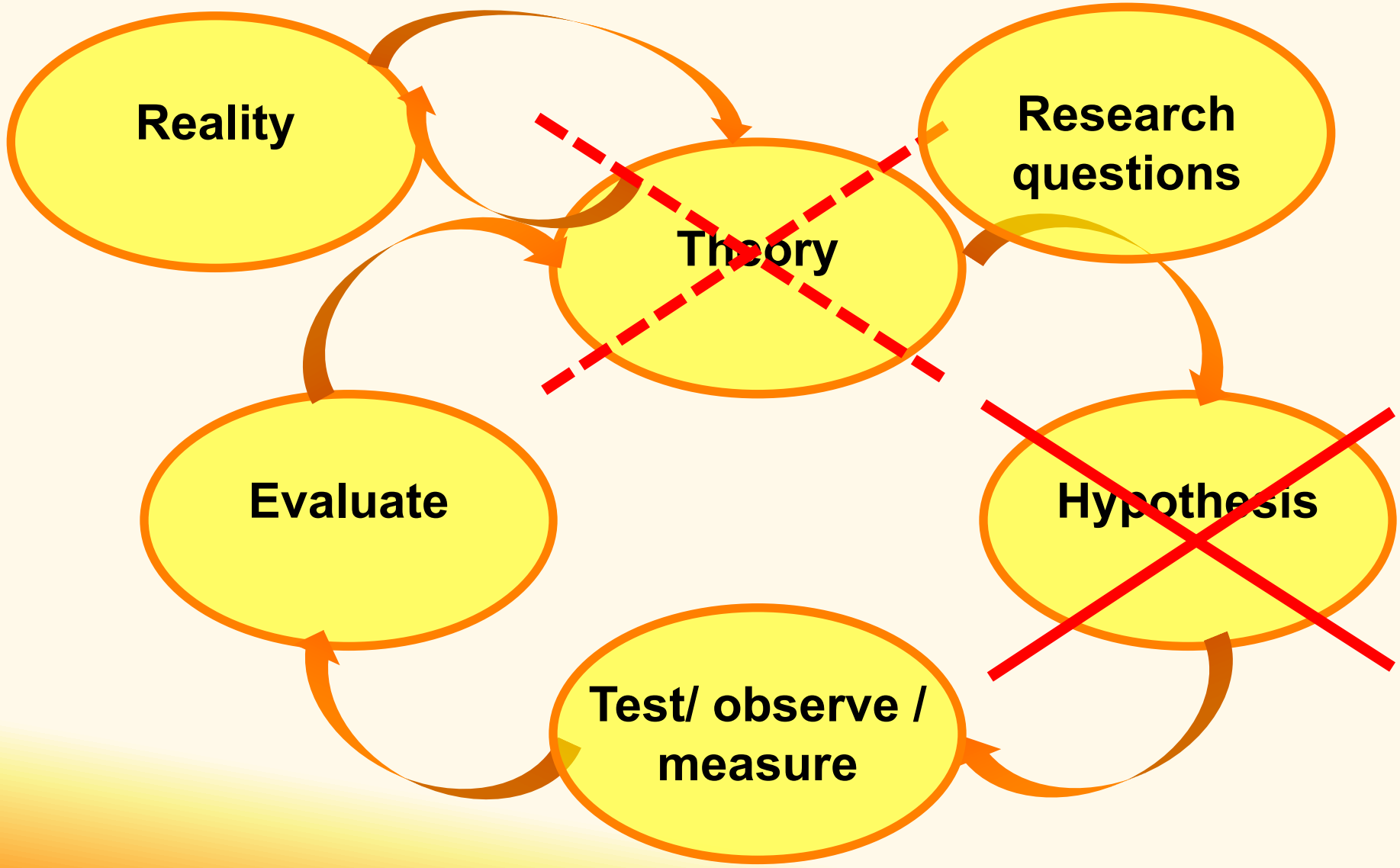
Why is (part of) this also experimentation?

- *Roughly* follows steps of empirical cycle:
  - Based on preceding work
  - Used empirical manipulation
  - Tests systematically (e.g. “experiments” with technology)
  - Evaluates result
  - Relates it back to theory and practice

Example 6: Experience of smart watches

Cecchinato, Cox, & Bird, J. (2017). Always on (line)?: user experience of smartwatches and their role within multi-device ecologies. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 3557-3568). ACM.

Different empirical approach:
Observe what is going on, then form theory

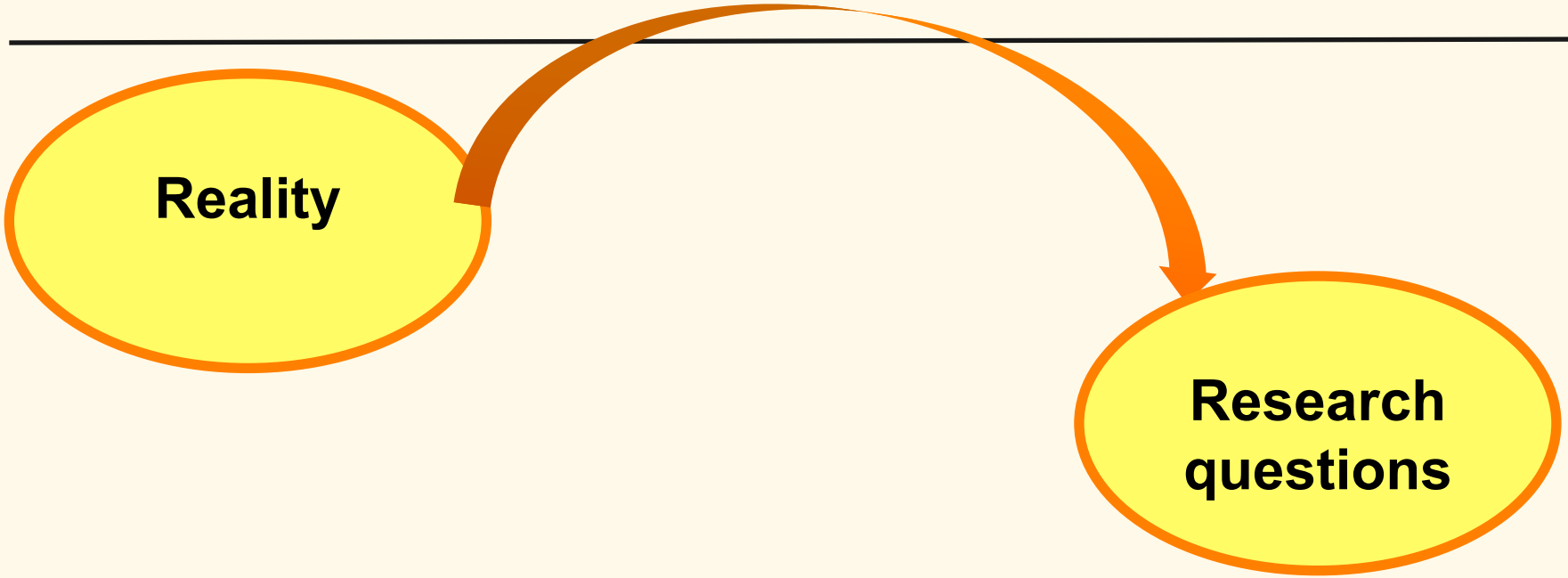


Step 1: Reality

Reality

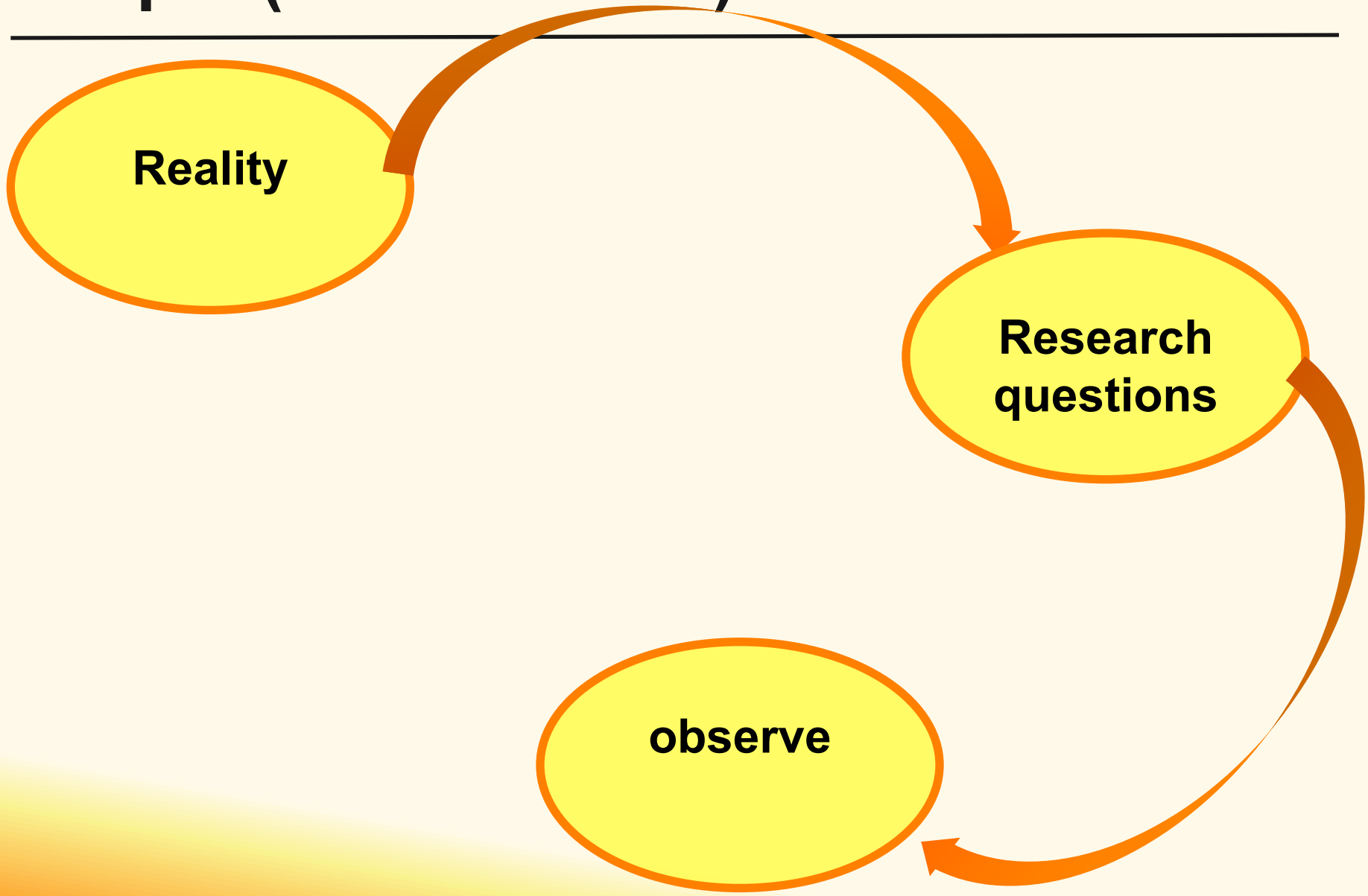
Many notifications on smartphones and watches. How do people handle those?

Step 2 (for this work): Research questions / scope



- What are benefits and challenges of wearing a device that is always online?
- How do smartwatches fit into user's wider multi-device ecologies?
- How can the design of smartwatches be improved in order to better fit user needs?

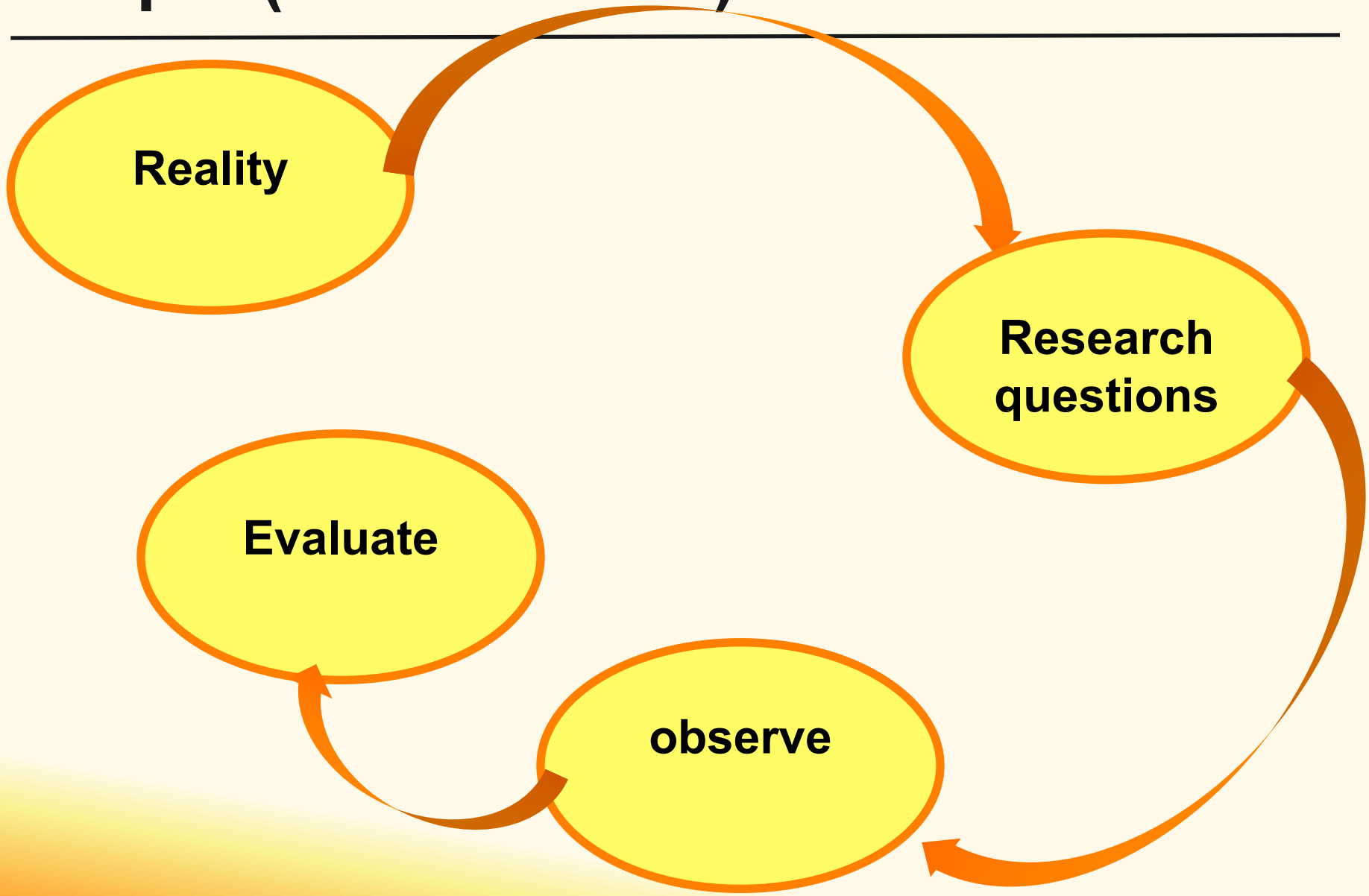
Step 3 (for this work): observe



Method: Mixed qualitative method

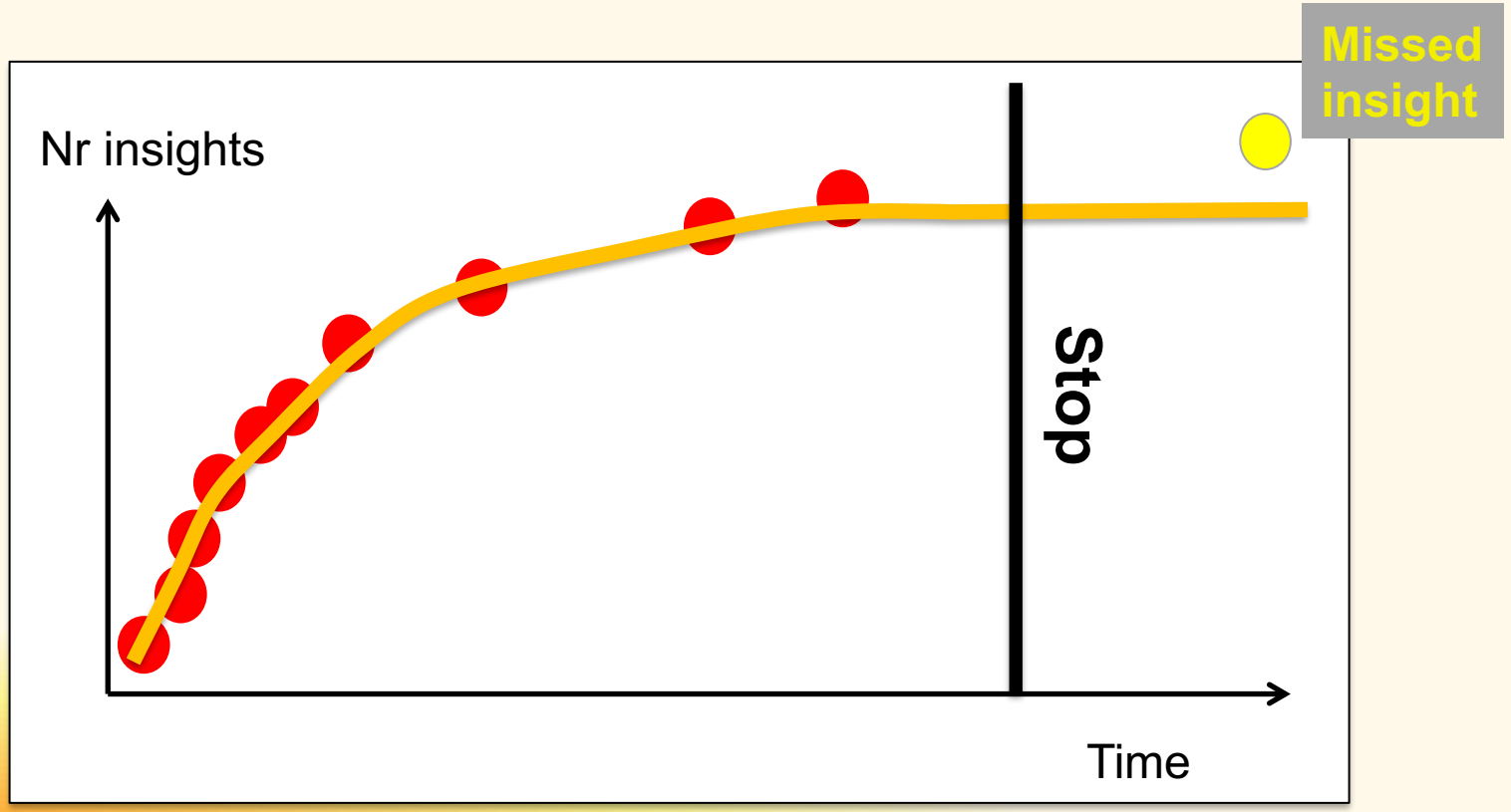
- **Diary-based Auto-ethnography**
  - **2 months of daily diary entries on experiences and interactions**
- **17 interviews**

Step 4 (for this work): Evaluate



Evaluation: thematic analysis

- Systematic approach to find new data points
- Analysis (and observation / data-collection) continues until “saturation is reached”

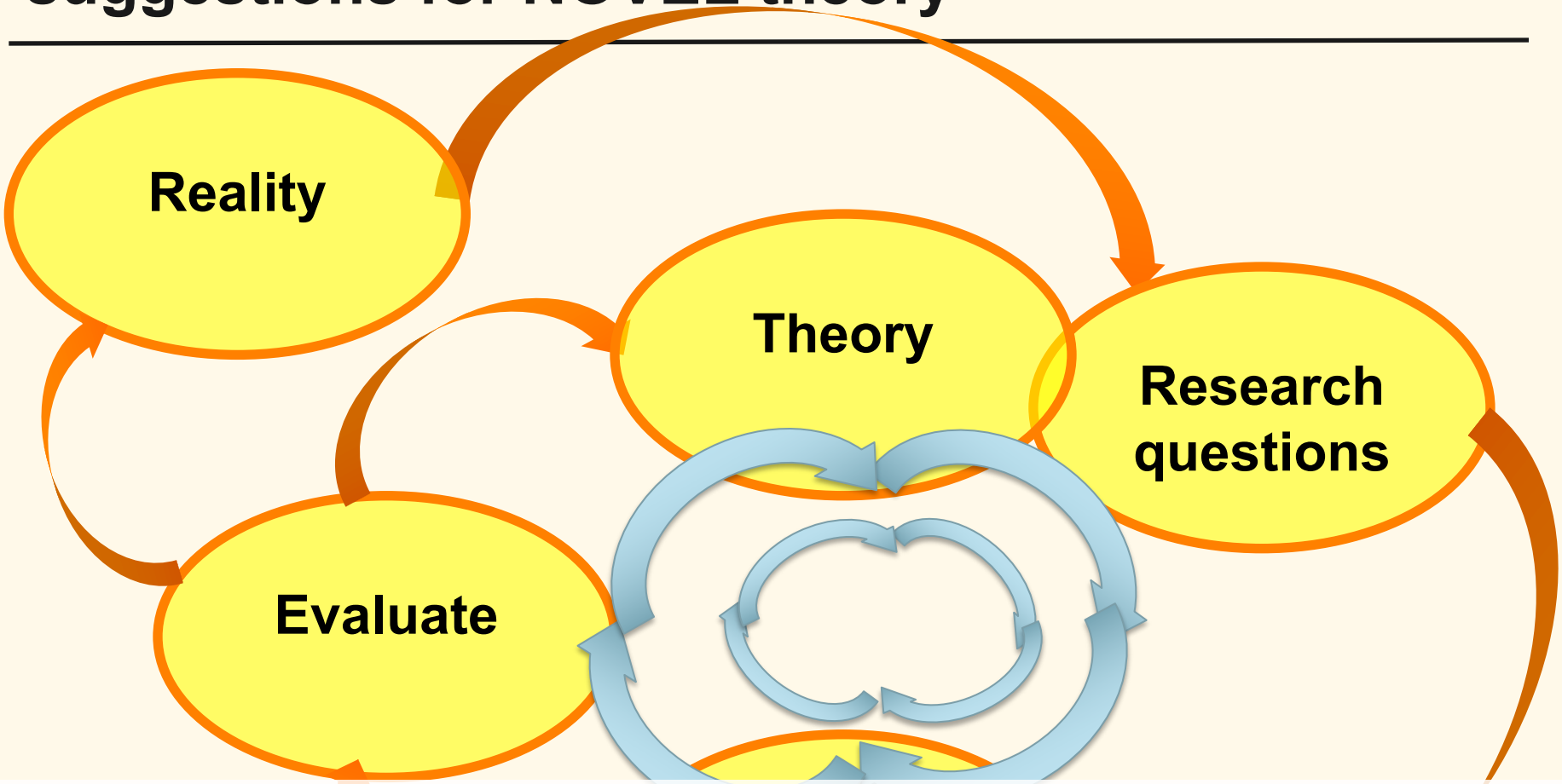


Data representation

Presentation of ideas, sorted by themes. Supported with quotes from users. For example:

- **Theme: receiving notifications**
  - **Subtheme: "How: glanceable information"**
  - **General text describing the finding, then quote:**
  - **"*It's made my life easier in terms of managing notifications and seeing which ones are duds and which ones are actual- ly ones I want to deal with.*" [P9]**
  - **"*I don't let myself be interrupted.* [...] *So the Pebble serves as an extension of the phone that lets me know that some- thing is going on and has the added benefit of finding out in a more convenient way what that something is. It gives me a little bit of distance;* [...] *I can't answer it because I have no microphone, but at least knowing what's going on can be useful. I can decide is it worth running to grab the phone.*" [P13]**
  - **Statement on generalization: "All participants but one [P6] also agreed that glancing at their smartwatch was less rude and more socially acceptable than looking at their smartphone. In particular, .."**

Step 5 (for this work): Relate result to reality and suggestions for NOVEL theory



e.g.: data suggests people use watch different in different contexts, and reflects in glance time. Experiments can test how:

- *Glance time varies between experimental conditions*
- *How this impacts performance, experience*

Why is (part of) this also experimentation?

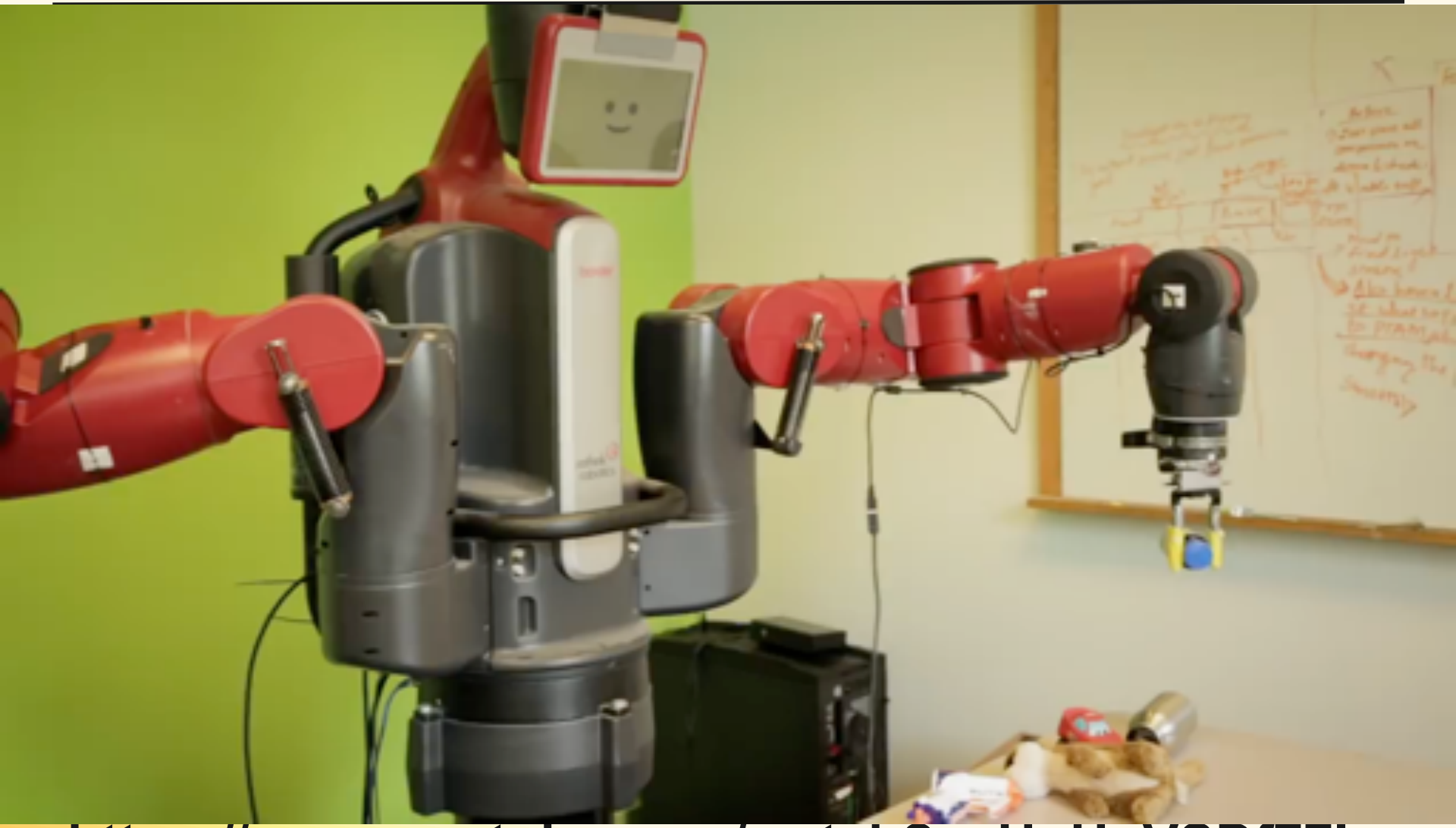
- **Empirical studies** **(“going out there and measure things”)**
- **With intention to answer research question**
- **Uses scientific tools and methods to study question**
- **Difference: no theoretical motivation, hypothesis or manipulation**

Qualitative studies…

- Sometimes seem “easy”
  - “No statistics”
- Are very hard to do well!
  - Unsure how much data is needed
  - Data collection and processing is intense
  - Subjective component, therefore more subjective assessment of whether it was done “correct”
- But… valuable in creating new ideas!

Experiments: Just humans?

Experiments: not just humans…
Active manipulation for learning



https://www.youtube.com/watch?v=HuHuVSDf77I

Today's topics

- **Why is experimentation useful?**
  - **In science**
  - **In practice / industry**
- **What constitutes experimentation? (see also: Cairns, 2016)**
  - **Empirical cycle**
  - **Experimental design**
  - **Statistical analysis**
  - **Experimental write-up**
- **Discussed using examples**
- **Topics we discuss along the way: manipulation, causality, validity (4 types), confounds & control, (in-)dependent variable, factor, condition, level, within- and between-subjects, counterbalancing….**
- **If you want to know more..**
- ***Appendix: more info that is useful for exam!***

If you want to know more

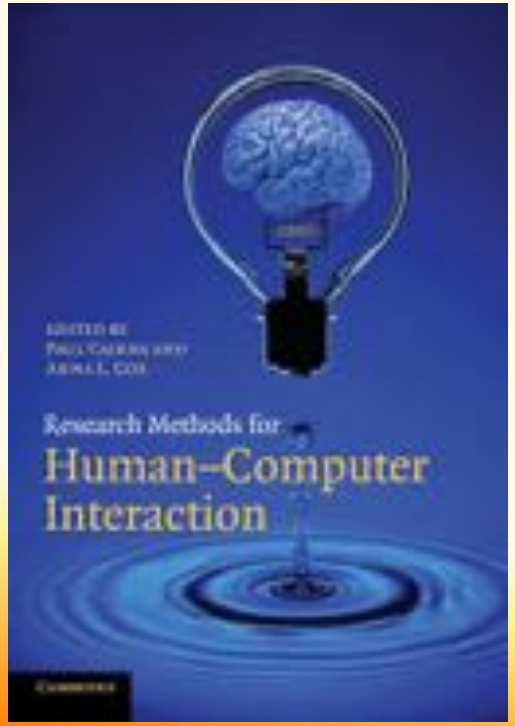
- **Term 3:**
  "Experimentation in psychology, linguistics, and AI"
  - **More hands-on experience with experimentation**
  - **More background on various steps involved**

If you want to know more: books on methods

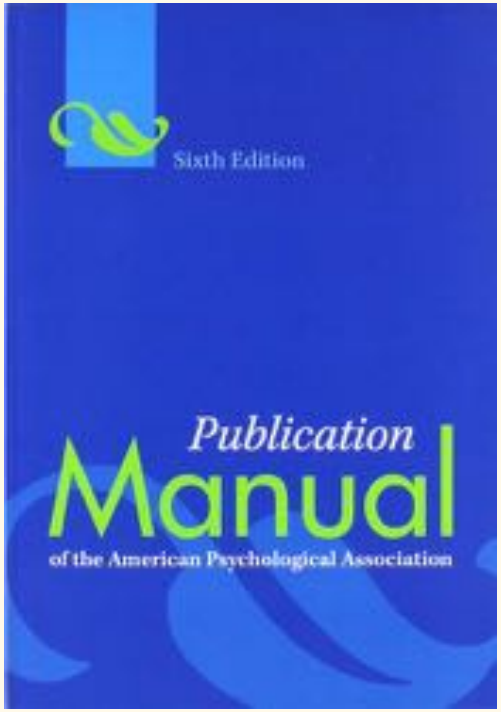
INTERACTION DESIGN FOUNDATION

https://www.interaction-design.org/literature

Many papers on various research methods (incl experiments by Cairns)

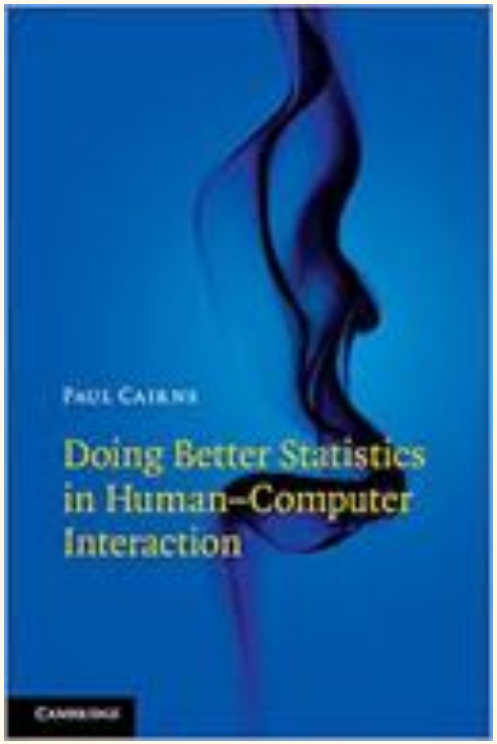


Basic intro to research methods, with applied perspective

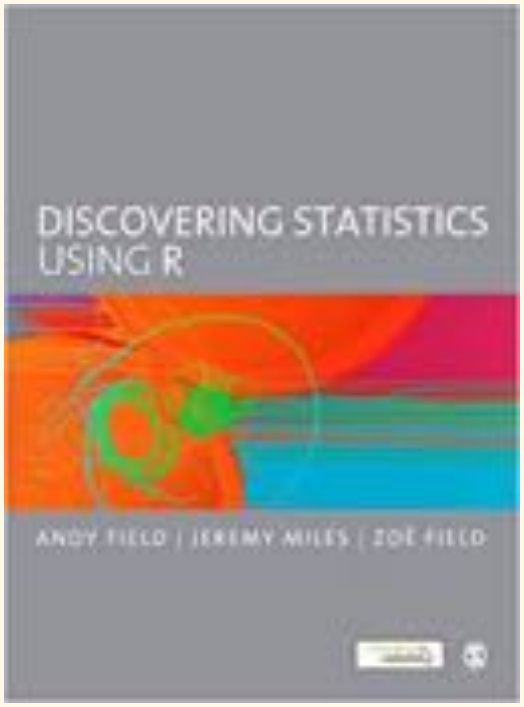


How to write an empirical methods section

If you want to know more: books on stats



Intro: why statistics?
Applied (& tech)
perspective and examples



Statistician's perspective
+ R implementation

See also: resources provided in appendix of MAIR assignment

My classes: methods covered

- 1/10: Cognitive modeling
- 3/10 Experimentation
- **8/10 Scientific writing**
- **10/10 Designing & Evaluating AI and Automated systems**

My classes: preparation & deadlines (see blackboard)

- 1/10: Anderson, J. R. (2002). Spanning seven orders of magnitude: a challenge for cognitive modeling. *Cognitive Science*, *26*(1), 85–112.
- 3/10: Cairns, P. (2016) Experimental Methods in Human-Computer Interaction. In Soedergaard, Dam (Eds.) *The encyclopedia of Human-Computer Interaction* (2nd edition). Online available at: https://www.interaction-design.org/literature/
- **8/10:**
  - **Before 7/10 at noon, hand in "Fantasy abstract" via Blackboard individually**
  - **Before lab: read lab assignment (of part 2)**
- **10/10:**
  - **Extra office hour to ask questions about assignment in building "Langeveld", Heidelberglaan 1, room H0.12 (follow signs to "floor 0", then to "section H")**
- **Week of 12/10: Chris in China (no e-mail)**

In 30 minutes…

Multitasking with Alexa: How Using Intelligent Personal Assistants Impacts Language-based Primary Task Performance

Justin Edwards, UC Dublin

Langeveldbuilding, room H0 (or follow us)
13:15 – 13:45

Questions?

Chris Janssen

c.p.janssen@uu.nl
www.cpjanssen.nl

Study questions lecture (not bullet proof)

- What is empirical cycle? What are steps? (be able to draw & interpret them)
- What sections are typical for an empirical report? How do these tie to the stages of the empirical cycle?
- What sections are there typically in an (empirical) methods section?
- Why are empirical studies important for science and practice? Why is this more than just for psychology
- Why is early writing important?
- Why is it often incorrect to "start" with data analysis (e.g., as in some big data approaches)
- What are: manipulation, causality, validity (4 types), confounds & control, (in-)dependent variable, factor, condition, level, within- and between-subjects, counterbalancing….

All questions: Be able to apply these principles/concepts/ideas to case studies

Study questions Cairns (not bullet proof) (1/3)

- What are the “three pillars” that underpin good experiments? And why is each pillar important and relevant?
- What is the main idea of “new experimentalism”?
- What is the role of experiments beyond testing theory (see text around Hacking, 1983 quote)
- Why is prediction central to good experiments?
- Why is causality important? And how is this achieved in experiments?
- What is meant with “GIGO”?
- What is (in-)dependent variable?
- What are 4 types of validity and why is each important? What are compromises for each type?
- What are ways to control for confounds?
- How do high ecological validity and experimental control relate?

All questions: Be able to apply these principles/concepts/ideas to case studies

Study questions Cairns (not bullet proof) (2/3)

- Why is statistics not simply a ”tag on”?
- Why do statistics alone not provide good evidence?
- What is Cairns’ “gold standard statistical argument?” Why is prediction important for statistics?
- What is ”fishing”?
- Why use the KISS principle for experimental design? What are Cairns’ recommendations (simple rules) for experimental design?
- Why is a criterion for p-value needed? What are typical criteria?
- Why is p= 0.001 not an indication that something is ”more convincing” or “more important” than p=0.05?
- Why are sometimes smaller samples with significant results more convincing evidence?
- What is the threat/risk of small samples

All questions: Be able to apply these principles/concepts/ideas to case studies

Study questions Cairns (not bullet proof) (3/3)

- Why is a good write-up important for empirical studies? Why should many sections be written-up in advance?
- What are typical sections of a report?
- What is the value of the introduction / literature review for empirical work?
- What is a weakness of experiments?
- …
  All questions: Be able to apply these principles/concepts/ideas to case studies

Example exam questions from previous years

Some notes on these questions:

- Recent lecture, so these are questions that might get asked, not actual exam
- Questions vary in their difficulty. Some focus on the lecture content, others on your understanding of the articles.
- Some stay close to study questions that I gave with my classes (see previous slides), others ask you to apply your knowledge to a novel situation

Example exam questions (1)

Draw the steps of the empirical cycle in a diagram (including arrows).

Example exam questions (2)

For the following example sentences from a scientific report below, answer two questions:

I. To which step of the cycle they belong

II. In what section of the scientific report you would find them

A. “Coffee dosage was manipulated between subjects.”

B. “Previous work has investigated how eating chocolate while studying for an exam helps you stay motivated”

C. “Smartwatches are everywhere nowadays”

D. “Reaction time was faster in the Pepsi condition (*M*= 12s, *SD* = 0.8s) compared to the Coke condition (*M* = 15s, *SD* = 1.2s).”

Example exam questions (3)

Cairns (2016) discusses 4 types of validity. For the following example, please indicate *which* validity type is breached and *why*.

“John tests the immersion of his new game ‘AI RULEZZ’. He develops his own immersion questionnaire, and finds that players rate the immersion of his game higher than the immersion of another game called ‘LOOZAHS’.”

Example exam questions (4)

Jane tests how coffee dosage affects the reaction time in a driving experiment. She tests four levels of coffee dosage: no coffee, 1 cup, 2 cups, or 5 cups.

For this example:

What is (or are) the independent variable(s)?