UNIVERSITY OF EDINBURGH - CNCC FINAL CONFERENCE
OCTOBER 2-4, 2009

THE EXTERNALIST CONSTRAINTS ON SELF-EVALUATION

Joëlle Proust & Anna Loussouarn

Institut Jean-Nicod, Paris

Institut | Nicod

CNRS L'ECOLE DES HAUTES ETUDES EN SCIENCES SOCIALES ENS

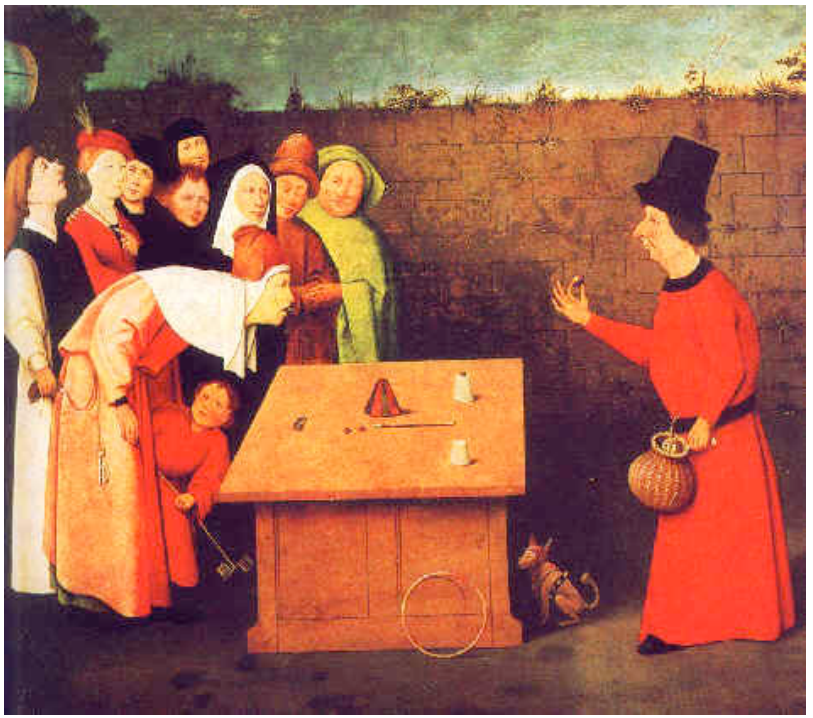
EUROPEAN SCIENCE FOUNDATION
SETTING SCIENCE AGENDAS FOR EUROPE

Hieronymus Bosch: The Conjuror

METACOGNITION

METACOGNITION
http:// mpscESF.free.fr

- D. Smith and M. Beran (U. of Buffalo)
- J. Perner (U. of Salzburg)
- J. Brandl (U. of Salzburg)
- H. Leitgeb (U. of Bristol)
- B. Renault (Salpétrière, Paris)
- J. Call (MPI Leipzig)
- J. Proust (ENS, Paris)



Collaborators

- Jean Lorenceau, DR Psychology (LENA)
- Damien Gabriel, postdoc Neuroscience (LENA)
- Anna Loussouarn, doct. Student Philosophy (IJN)

Special thanks to:

- Kevin O'Regan & Pascal Mamassian (LPP, CNRS, Paris)
- Malika Auvray (LIMSI, CNRS)
- Takashi Kusumi (Kyoto University).

Versions of externalism

- **Semantic externalism**: reference of proper names and natural kinds is determined by external causal and historical factors.
- **Externalism about mental contents:** intentional states have contents that depend on external causal and historical factors.

Externalism about self-knowledge: a puzzle

- Externalism seems incompatible with the view that thinkers have a privileged access to their thought contents.
- It seems logically impossible that a thinker misjudges the content of her own thoughts.
- It is logically possible, however, that a thinker is influenced to sense, infer, predict or retrodict various properties of her own mental actions (such as truth, intelligibility, exhaustivity)

Versions of externalism

- **Externalism about justification**:

- A subject may be entitled to hold a belief even though she doesn't have access to the reasons that justify her to form this belief.

(Dretske, 2000, Sosa, 2009)

Active externalism

- A brain is coupled with a complex spatial (material & technological) environment.
- This environment drives a person's cognitive processes (provides support to thought)
- External resources (external memory aids, linguistic statements, etc.) may be constitutive of one's thought contents.

Clark & Chalmers 1998

Clark, 2002

Goal of this presentation

- Discuss the extent to which externalist assumptions apply to the case of metaperception (evaluation of one's perception)
  - on the basis of the theory of adaptive control applied to self-evaluations.
  - On the basis of an experiment using the paradigm of « change blindness » as a primary task.

Plan

1- Conceptual clarifications: mental actions & self-evaluation

2 – Conceptual and nonconceptual externalism.

3 –Feelings, norms, and externalism

4 – Dynamics, loops and metaloops

5 – A metaperceptual experiment

6 – Discussion and conclusions

1- Conceptual clarifications: mental actions & self-evaluation

Externalizing metacognition

- Conceptual clarification:

- Metacognition consists in performing evaluations of an upcoming or just performed mental action.
- It decomposes in various subdomains such as:metaperception, metamemory, metareasoning, etc.

Examples of mental actions

Purely Epistemic

- **Perceptual attending**
- **Directed reasoning**
- **Directed memory retrieval**
- **Directed visualizing**
- **Directed imagining**

Non purely epistemic

- **Planning**
- **Reflective deciding**
- **Controlling emotion**
- **Preference management**

Self-Evaluation of type 1: Self-probing

- Before trying to act mentally, one needs to know whether
  - Some item is in memory (before trying to retrieve it)
  - One has epistemic competence in a domain (before one tries to predict an event)
  - One is sufficiently motivated to act in a certain way (when planning)

Self-Evaluation of type 2: Post-evaluation

- Performing a mental action entails the ability to evaluate its success
- One needs to know whether
  - The word as retrieved is correct
  - One's reasoning is sound
  - One does not forget a constraint while planning

Conceptual and nonconceptual externalism.

Conceptual externalism

- One might consider that metacognition involves conceptual knowledge.

- It is arguable, in particular, that perceivers rely on current theories about perception to judge how accurate their perception will be, or has been.
- In this sense, metacognitive beliefs determine the specific contents of individual self-evaluative episodes.

Ex. Scholl, Simons & Levin, 2004, in *Thinking and Seeing*

- « Under the grip of incorrect theories about how aspects of their minds work, people often fail to accurately predict their own behavior. Several of the most pernicious of these metacognitive errors involve the nature of visual awareness »

Vygotsky's scaffolding principle

- Socially available theories can be only partially grasped by thinkers, but nevertheless influence their judgments.
- These theories are used by subjects as guiding constraints in the selection of appropriate mental actions (choosing this or that MA) and in their MC prediction/evaluation)

A complication for conceptual externalism

- Although this view may be correct, it cannot be the whole story.

Problem : nonlinguistic metacognition

Some non-humans, with no technology or social help, are able to perform metamemory and metaperception-based mental actions (Smith et al. 2006, Hampton et al. 2000).

However: highly trained social animals: extensive techno-procedural knowledge!

It is assumed that metacognition might rely on feelings of uncertainty.

Feelings do not need to be grind to an internalist mill

- Metacognition as self-evaluation is often seen as relying on internally generated noetic feelings.
- These feelings were at first considered as internal, intrinsic signals for self-confidence, based on the belief about one's past ability in tasks of the same type (D. Hume, *Treatise*, I, 4, 1)
- Metacognitive feelings, however, do not need to be ultimately generated by a thinker.

Strategic role of feelings

- **In bodily action** (Pacherie, 2008):
- Feelings like:
  - Sense of agency and ability
  - Sense of effort
  - Sense of ownership
- help decide whether an action is feasible, painful, whether it is preferable to another, whether it requires effort or not.
- help evaluate the success of the action.

Strategic role of feelings

- **In mental action** (eg: directed recall, problem solving, perceptual decision), feelings are used to
- decide whether it is feasible, whether it is preferable to another, whether it requires effort or not.
- evaluate the prospective success of the action.
- appreciate the retrospective success, near success, or failure of the action.

(Carver & Scheier, 1998, Koriat 2000, Proust, 2009).

Two externalist considerations on noetic feelings

1- Noetic feelings are motivational expressions of underlying dynamic facts, rather than of (propositional) attitudes

- Prospective noetic feelings represent the **velocity** to or away from a represented mental goal.
- Post-evaluative noetic feelings also compare the observed and anticipated outcomes in order to determine the necessity of revising goals.

Two externalist considerations on noetic feelings

2- **Noetic feelings reflect various norms, inherent to a task and a context**

- **Fluency or intelligibility: perceptual, memorial, conceptual, inferential.**
- **Coherence**
- **Quantity or exhaustivity**
- **Relevance (or quality)**
- **Consensus**
- **Accuracy**

Which norm is enforced when?

- Subjects enforce context-relevant norms, without having to think about it.

- It just is an architectural fact, (< the constitutive character of epistemic norms), that there is a specific norm regulating control and monitoring in each mental act.

MC & externalism: basic question

- What makes dynamic coupling possible?

The existence of regulation and feedback laws.

- Idem for METACOGNITION

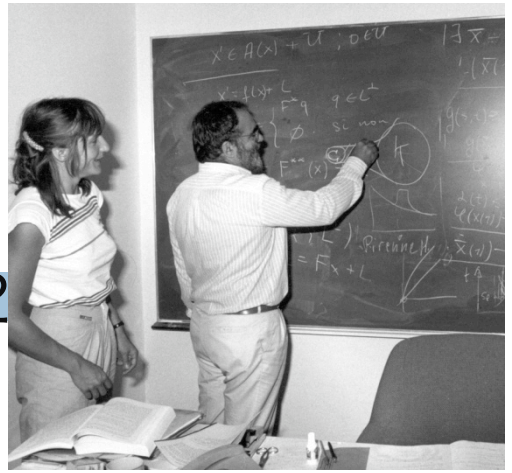
- Metacognition is in part constituted by dynamic facts to which mental agents are able to be attuned.

Metacognition, regulation, and feedback

General notions of dynamic control systems

Viability theory

(Frankowska et al. 1990, Aubin, 2001, 2



describes the mathematical structure of the dynamic couplings of two evolving systems (for example: an agent and his/her environment)

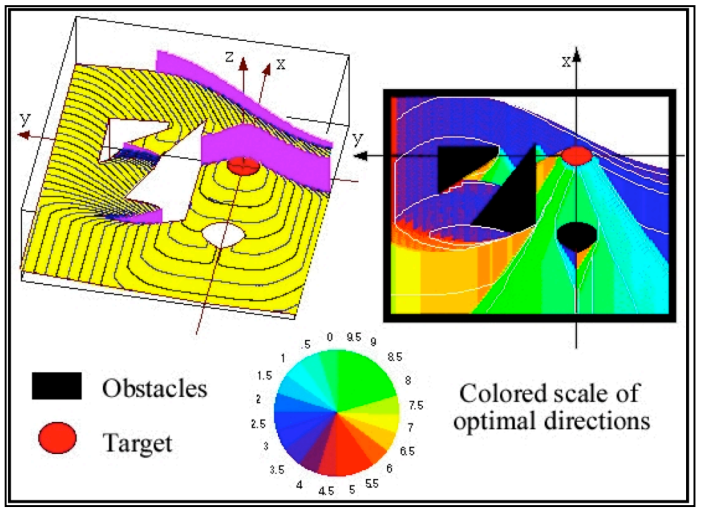
Dynamic coupling

Two forms of regularities determine the way in which a given organism can interact with the environment in a controlled way.

- **regulation laws** determine which affordances/viability/capturability correlate with specific **commands** in specific environments.
- **feedback laws** determine what portion of the regulation space is accessible to an organism with a given history.

Regulation laws

- Determine regulation mappings between commands and relevant properties (viability and capturability)
- The respective sizes of the mappings measure robustness or resilience of viability for a system in an environment



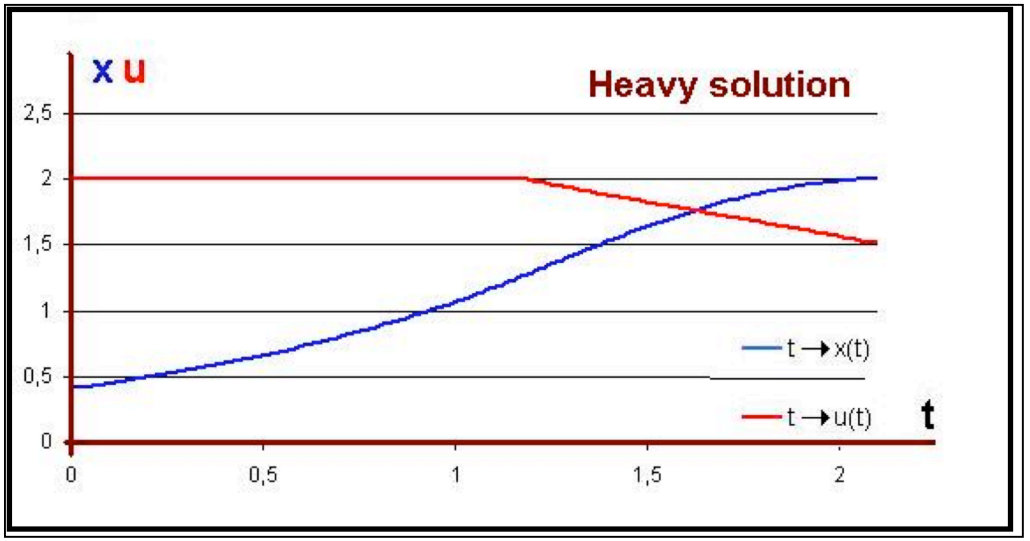
Regulation laws

- Can predict/describe viability crises and the kinds of transitions that can restore viability (impulsions applied to the state variables, changing velocities etc.)

Regulation laws: principles

- **Principle of inertia**
  Controls evolve only when viability is at stake.

  Most evolutions are governed by this principle: for example those that select the controls with the **smallest velocity** ("heavy solutions").



Feedback laws

- Feedback laws allow agents to predict on the basis of observed outputs at a given time the regions of interest (in action space) for further regulation.
- By sampling network propagation velocity, in associative networks, **a mind/brain can predict its own learning curve for a task.**

- **How does it work (roughly) ?**

Feedback loop and metaloop

- **Control and monitoring of action:**

  represent and minimize discrepancy between observed feedback and expected feedback (« reference value »)

- **Meta-monitoring and control of action:**

Represent **the rate** of discrepancy reduction in the monitoring system over time (**how rapidly** discrepancies are diminishing).

Carver & Scheier, 1998

Different functions

- Simple action loop: deals with ‘distance’ from the goal (an executive function)
- Metaloop: deals with **velocity,** ie the first derivative of distance over time (or **acceleration,** the second derivative).
  - **Function: control the rate of discrepancy reduction and use it in rational decision to pursue goals.**
- Metacognition of type 1 relies on metaloops, while metacognition of type 2 mainly uses simple loops.

Self- evaluating requires successive trials :

- comparing feedback to the system's reference value -- **at the loop level** (was the goal/subgoal met?)

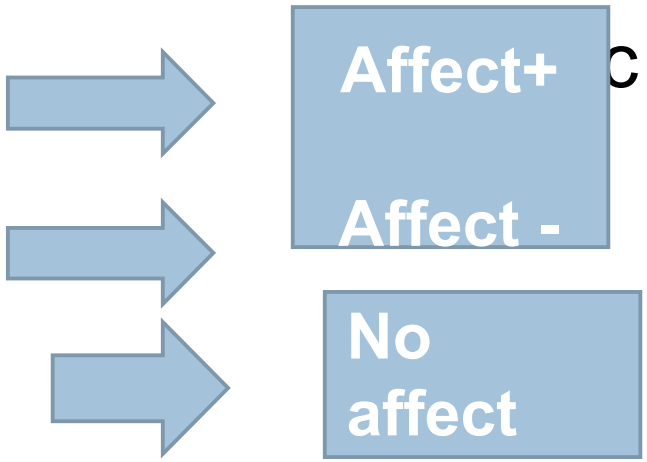
And

comparing trial outcomes over time to the meta-system reference value at the metaloop level

above the reference value

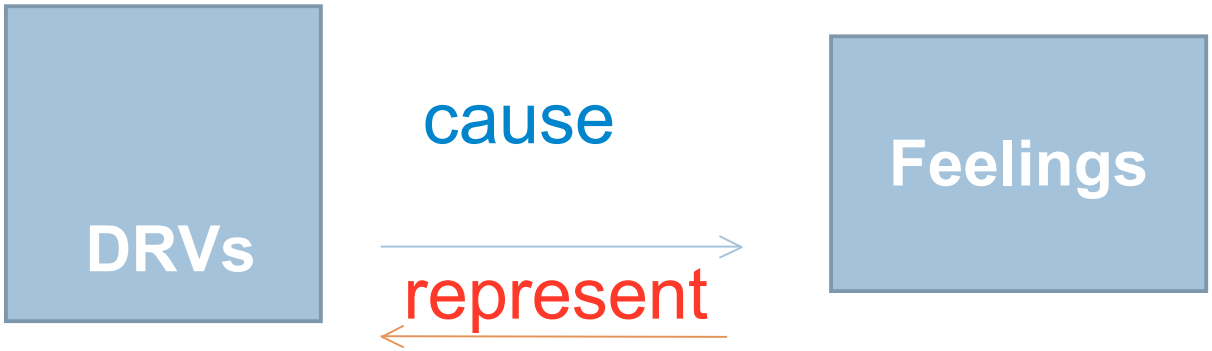
below the reference value

At the reference value



Epistemic feelings and externalism

- Observed or predicted discrepancy reduction velocity (DRV, ie gradient comparator) is the informational source and the causal origin of feelings of confidence.



- Agents are unaware of feedback rate and probabilities of success; they are only aware of their feelings (not necessarily of having them).

Two possible sources of comparisons & associated feelings

- Koriat et al., 2006:
- Running can make you feel afraid: « control-based monitoring »
- Seing that running increases the distance from danger lessens your fear: « output-based monitoring ».
- The two forms of monitoring play a role in modulating control of one's memory in self-paced learning.

2 types of metacognitively significant dynamics

- How **the task, (what is to be controlled)** is organized (classroom aspect, timing of meetings, organization of cursus and stimuli, difficulty over time etc.)

- How the **output monitoring** is organized (grading system, ranking system, adult or self-regulation, etc.)

Conclusion of section 1

- Feelings are the subjective counterpart of dynamic facts,
- Dynamic facts determine the evolution over time of the action space
- Monitoring and meta-monitoring of action allow agents to guide their actions according to stored norms and contextual reference values.
- Agents can also use non dynamic facts, eg folk- theories of cognition.

Conclusion of section 1

→ Externalism in :

- Conceptual content
- Nonconceptual, evaluative content
- active externalism: role of external dynamic stabilizers and extractors (attention-structuring artefacts and spatial lay-outs)

A metaperceptual experiment:

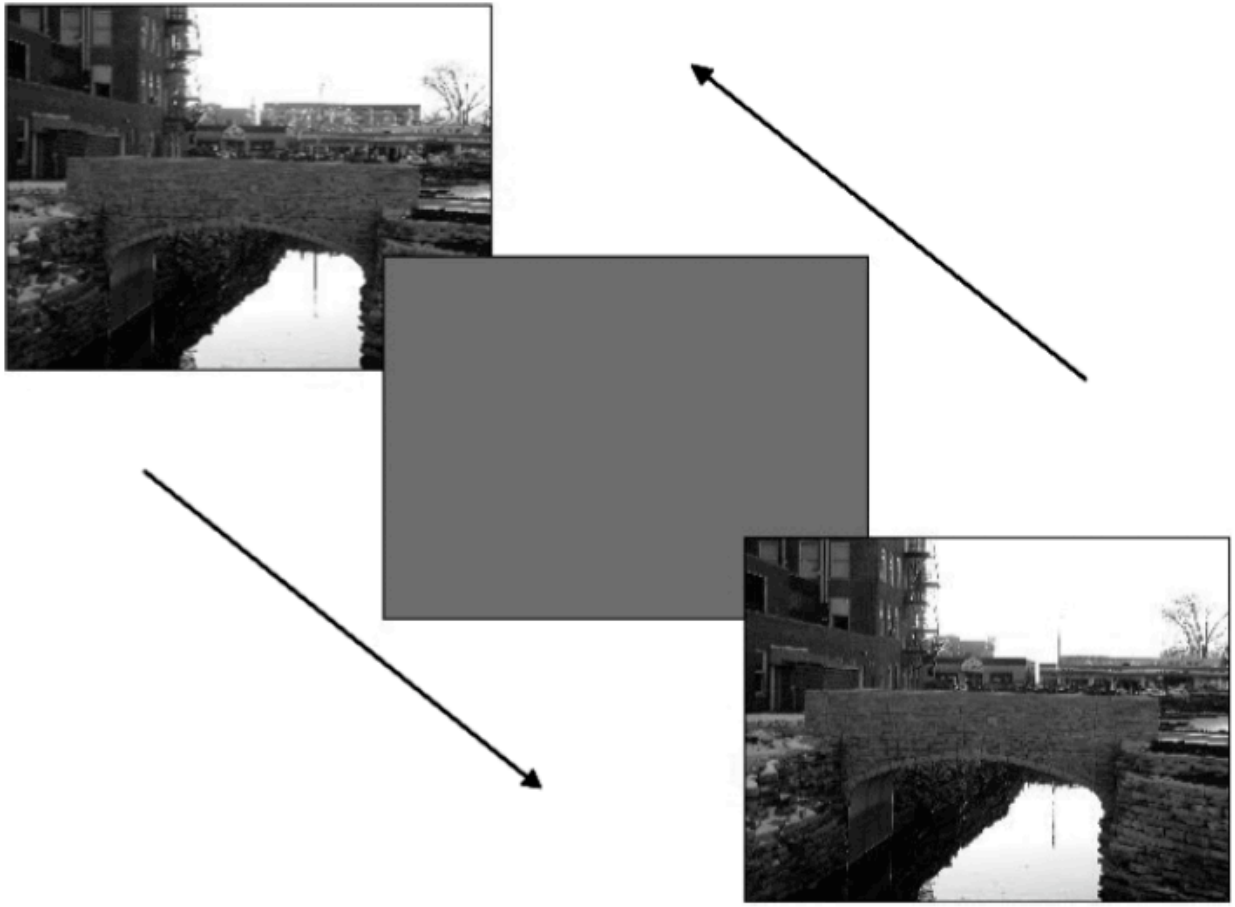
Change blindness blindness

What is change blindness ?

- Observers are surprisingly poor in detecting salient differences in visual images viewed successively, in the absence of perceptual cues directing attention to the locations of the changes

Rensink & al., 1997, Simons 2000, 2005

Flickering paradigm (Rensink, 1997)



Simons & Levin, 1998

Change blindness blindness

A **metacognitive** error, manifested by:

- A failure to **predict** CB (Levin, 2000, 2002)
- A failure to **post-evaluate** CB (Scholl, 2000, 2004):
  - Flatly denying CB, (Scholl, 2000)
  - Under-evaluating their CB (Scholl, 2004)

Hypothesis

- Change blindness blindness is generally considered to be explained by folk theorizing (creating the illusion of richness).

- If it is right to think that the reference value and the comparator VDR also determine the way subjects feel about a task, then there should exist no unconditional CBB.

HYP:2 sources of metaperceptual evaluation

- Implicit source: control-based monitoring increment/decrement in task difficulty.

- Explicit source: output-based monitoring incl. feedback from others, folk theorizing.

Opposition paradigm

- In order to have an indication of the specific contributions of each dimension engaged in self-evaluation, we have to put implicit and social feedback **in opposition to each other.**
- **Control-based monitoring: task organization (difficulty gradient)**
- **Output-based monitoring: positively or negatively evaluated outcomes**

Control-based monitoring

- An experimental variable manipulates **gradient of difficulty** for change detection

- The changes are introduced at successive delays of
- 4 secs, 8 secs and 12 seconds
- Or
- 12 secs, 8 secs, and 4 secs.

Output-based monitoring

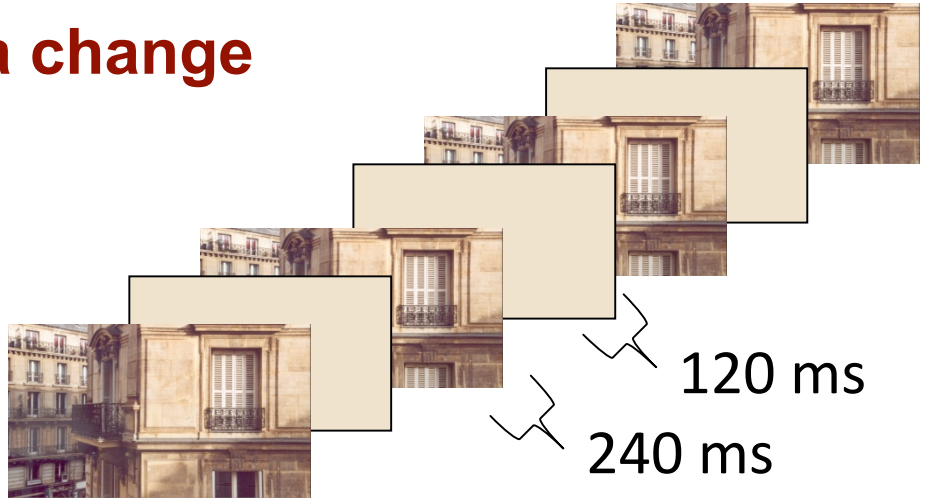
Measuring the explicit appreciation of one's performance:

An experimental variable manipulates **the observed delay in detecting change.**

Description of the task

- Basic task (common to all conditions)

1. Searching a change

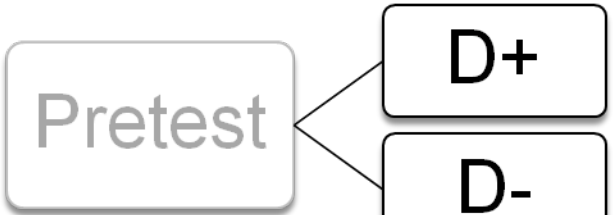


2. Localizing a change

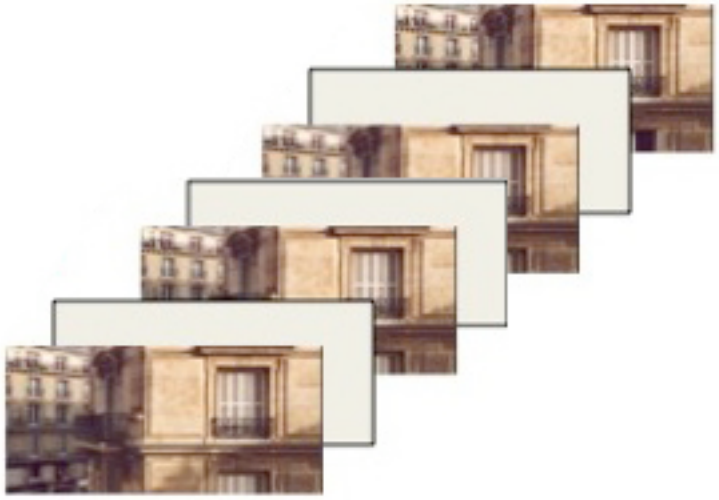
Time bar stops running

3. Estimating when change was introduced

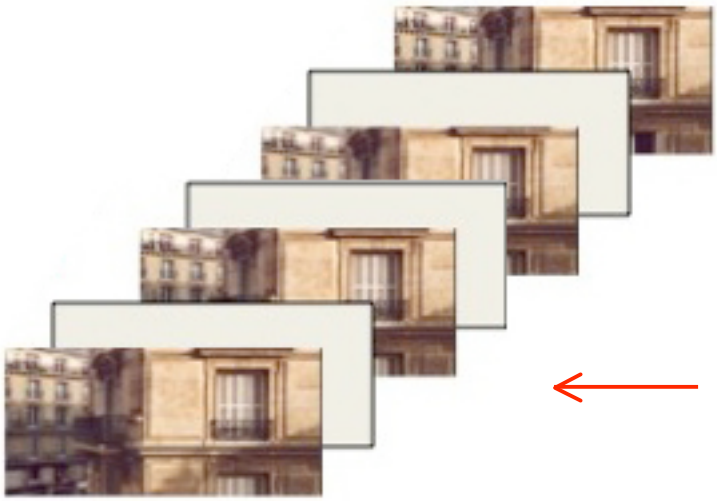
1st session: two groups, with inverse gradients of difficulty



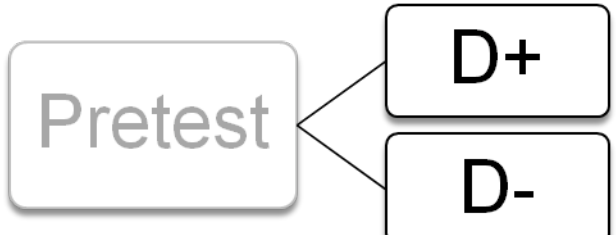
CONDITION IMPLICIT +



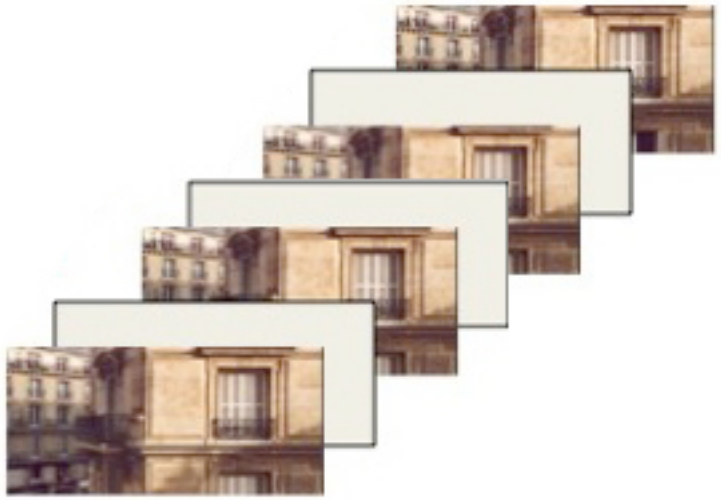
CONDITION IMPLICIT -



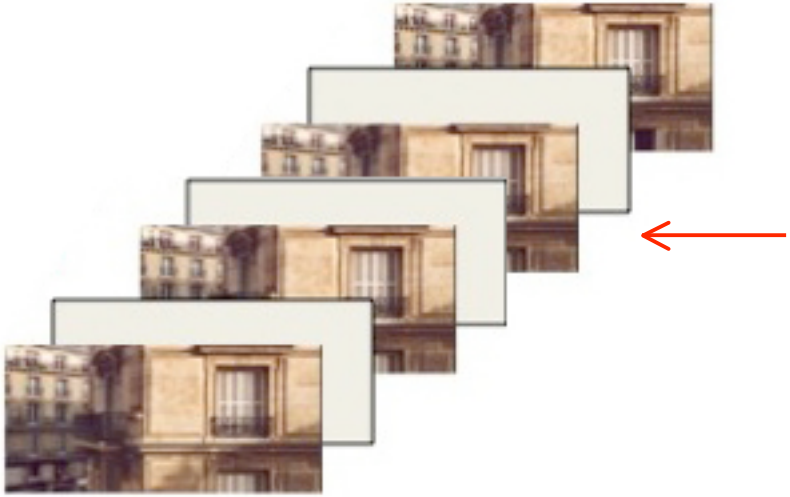
1st session: two groups, with inverse gradients of difficulty



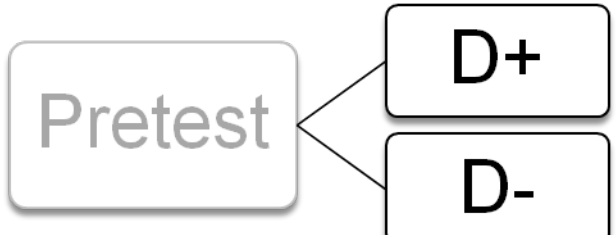
CONDITION IMPLICIT +



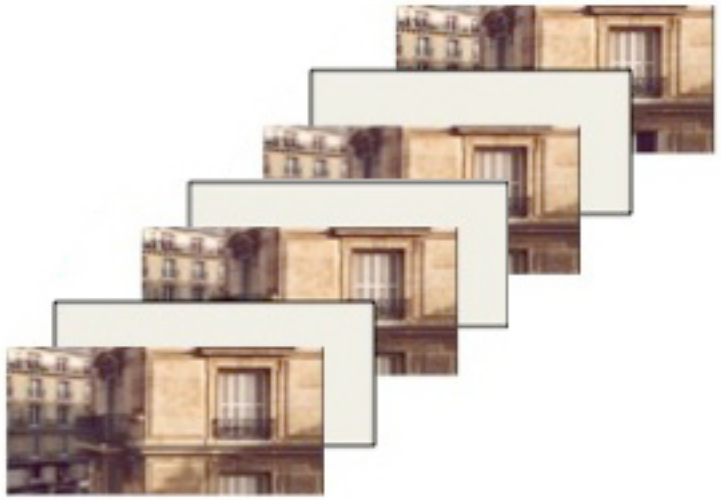
CONDITION IMPLICIT -



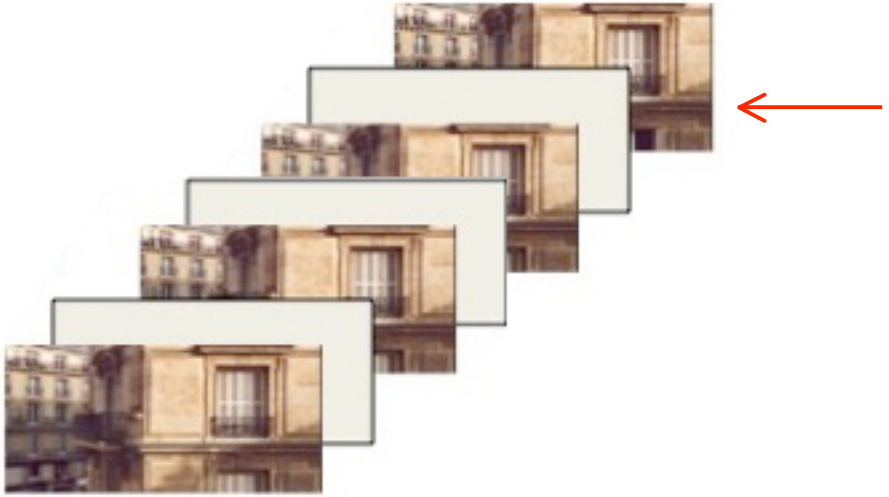
1st session: two groups, with inverse gradients of difficulty



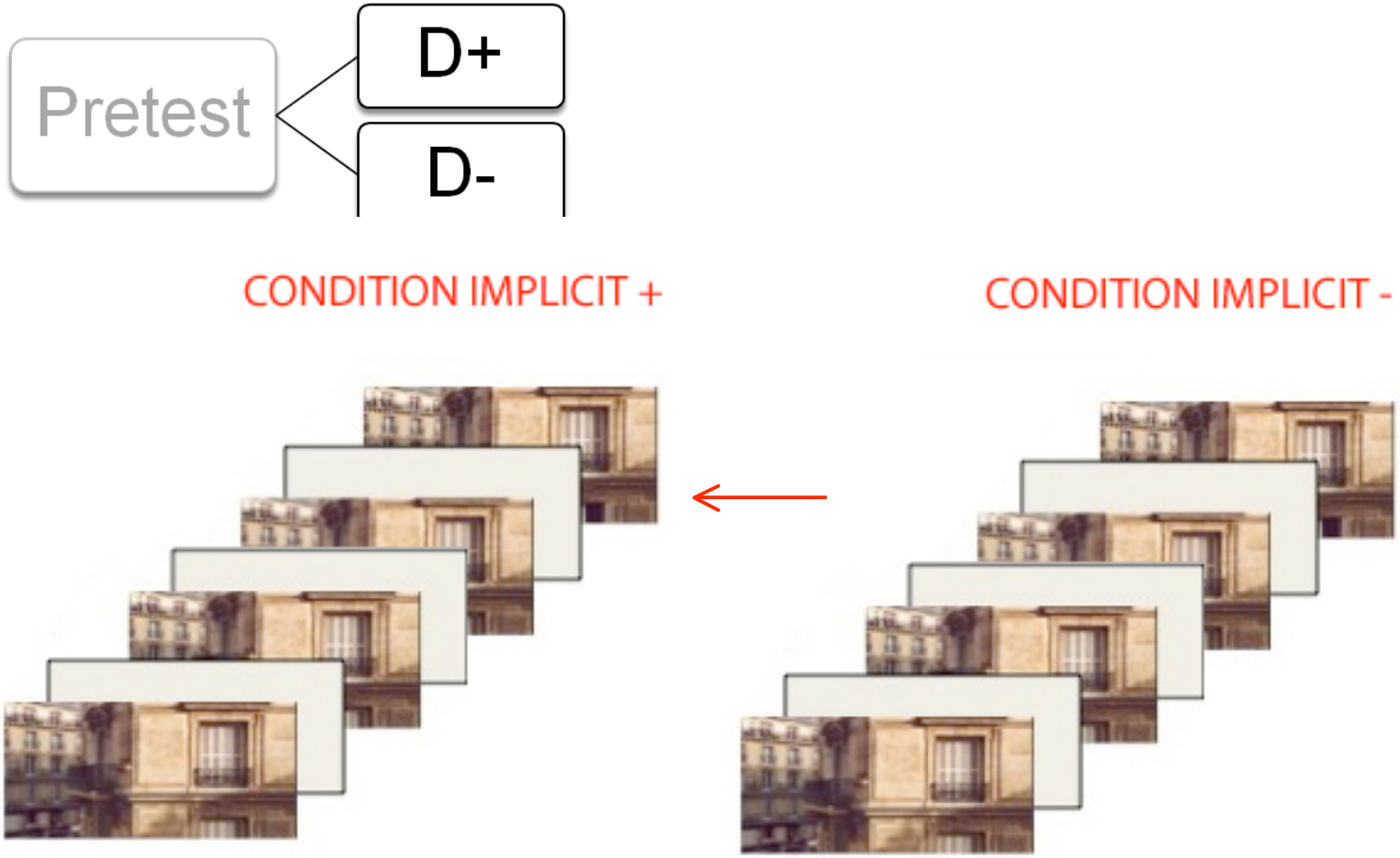
CONDITION IMPLICIT +



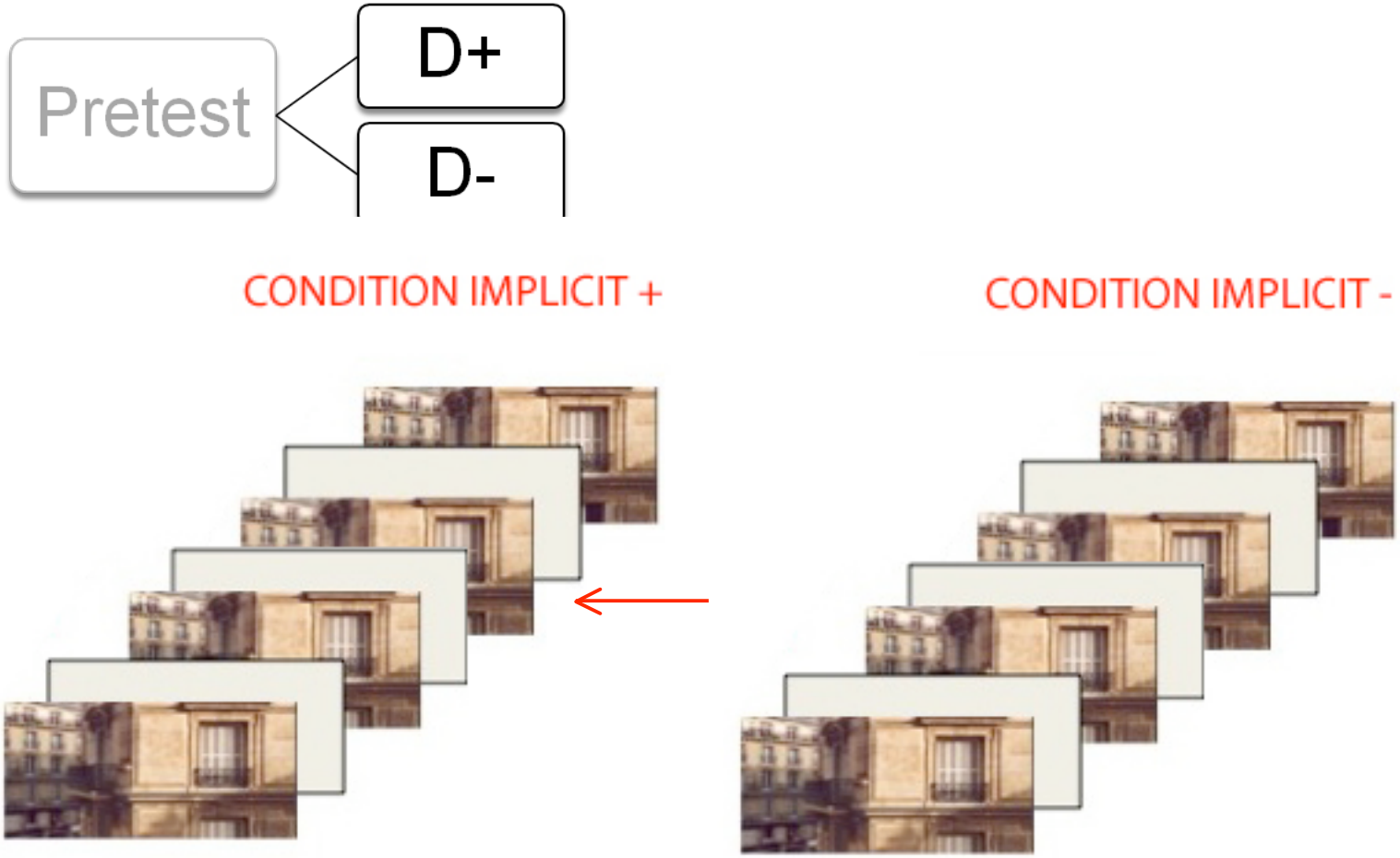
CONDITION IMPLICIT -



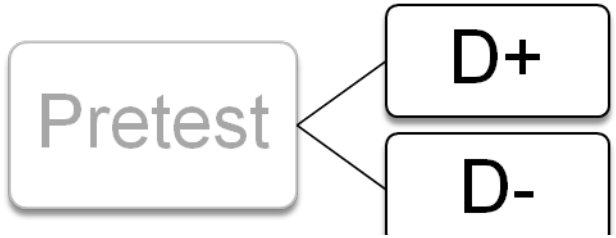
1st session: two groups, with inverse gradients of difficulty



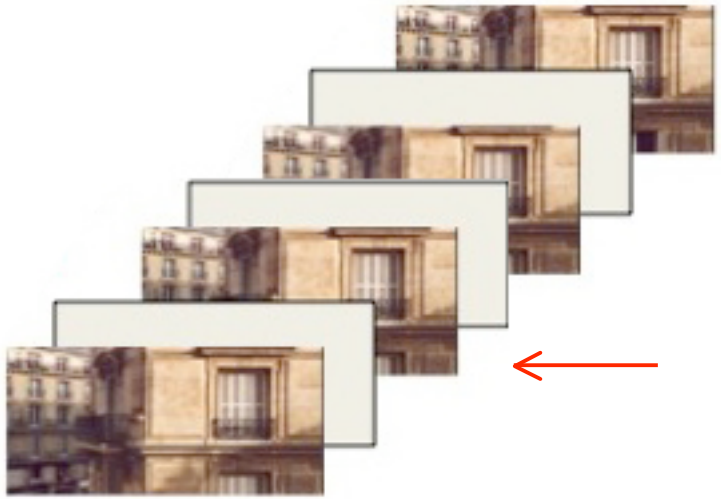
1st session: two groups, with inverse gradients of difficulty



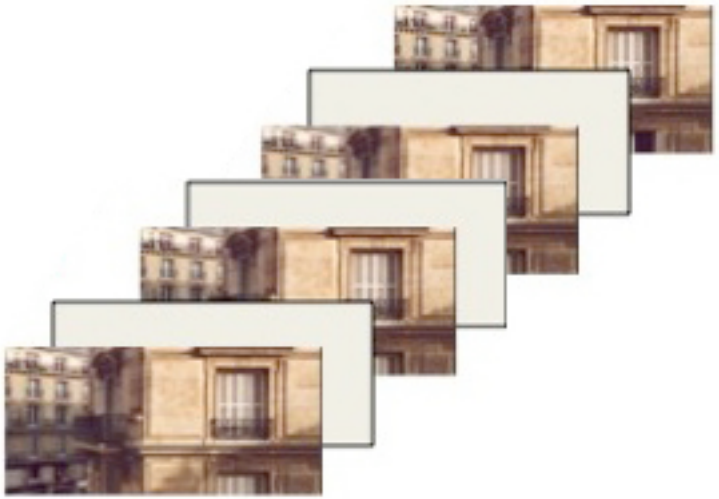
1st session: two groups, with inverse gradients of difficulty



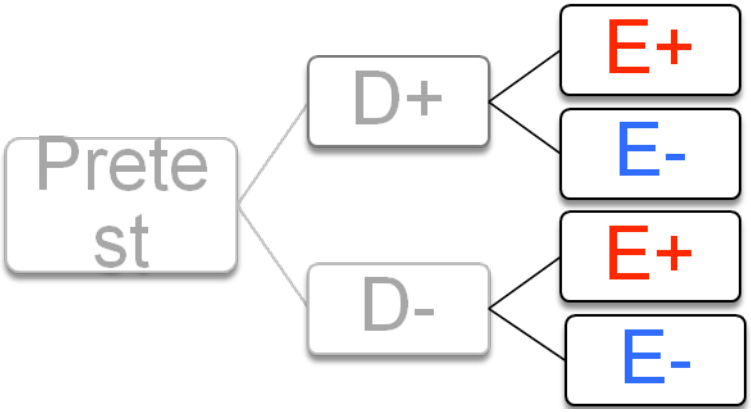
CONDITION IMPLICIT +



CONDITION IMPLICIT -

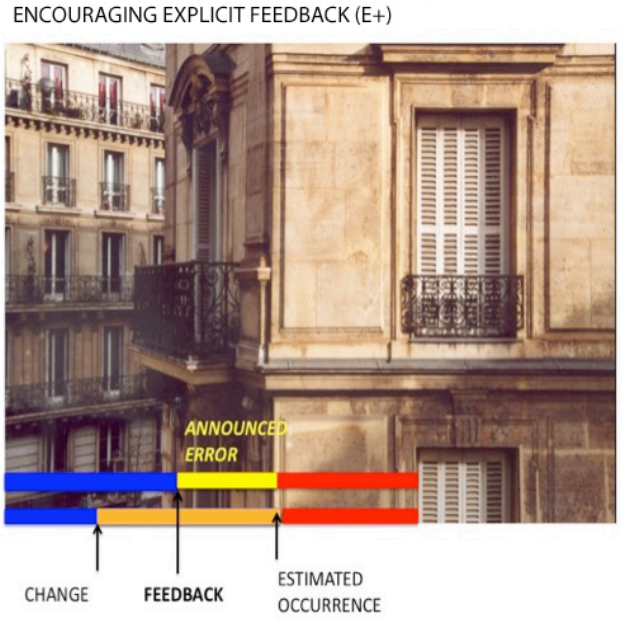


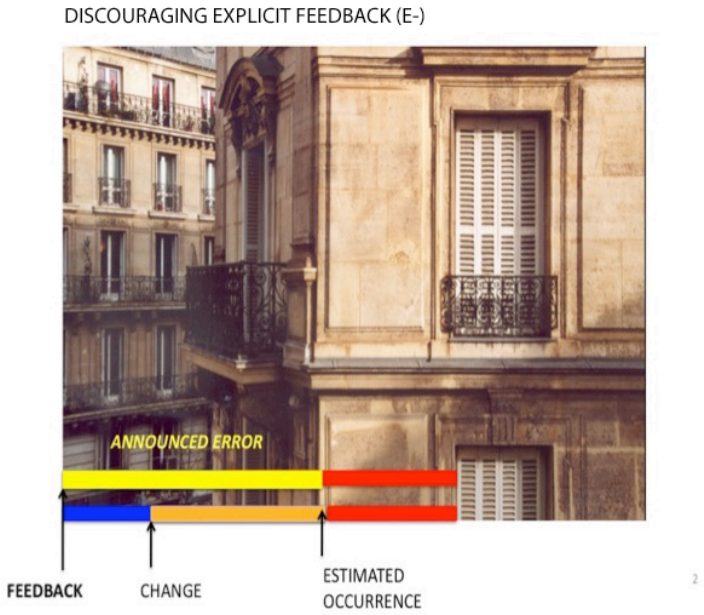
2nd session: two groups, with biased over/underestimation of performance



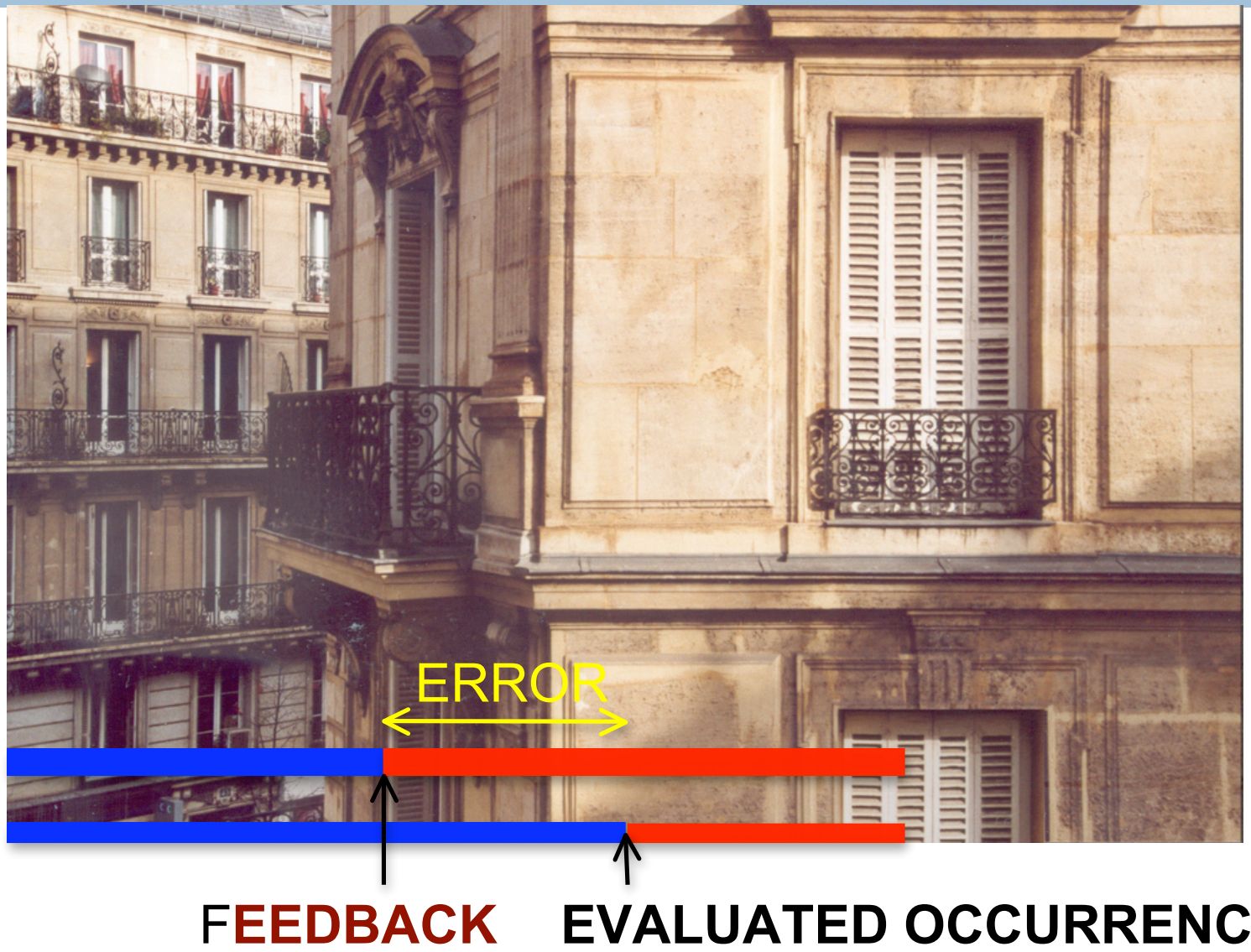
Giving **explicit feedback**

1. About delay of detection
2. About error in self-evaluation (« CBB »)

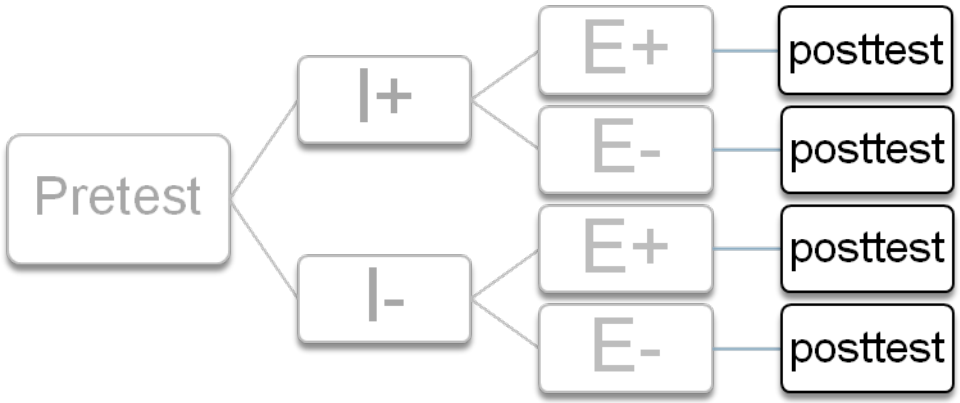




Explicit feedback



Post-test:
How persistent are the effects of control-based and output-based feedback?



1. **No feedback provided, neither implicit nor explicit**

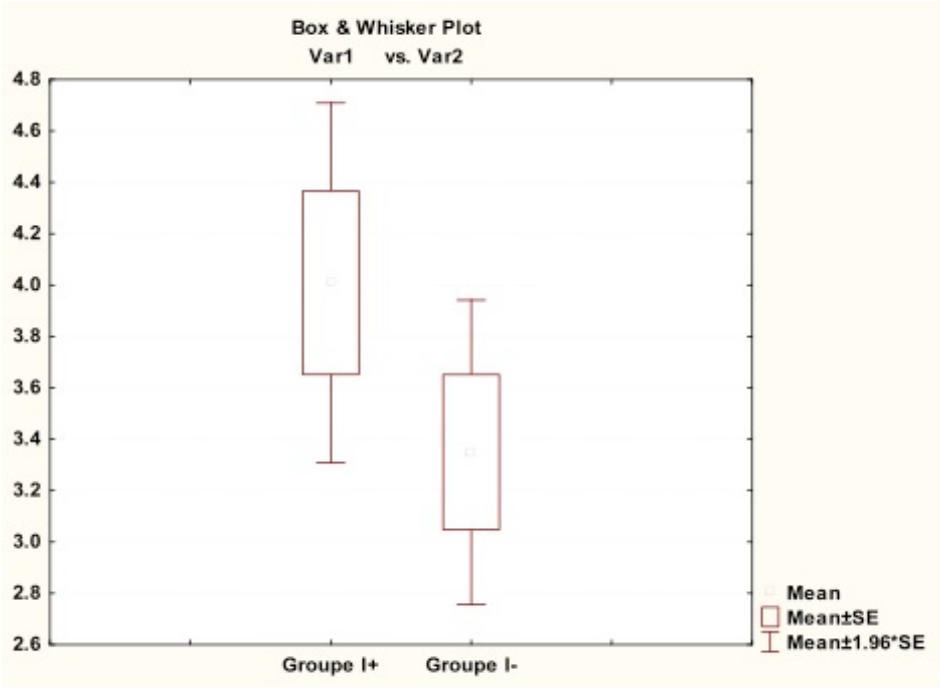
2. **Participants provide own self-evaluation (« basic task »)**

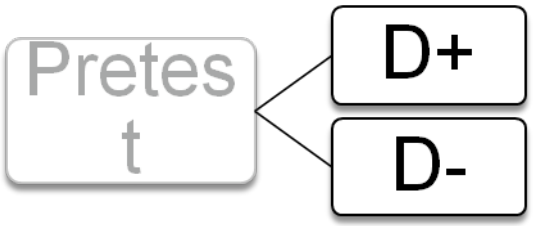
Results

1 - Impact of gradient of difficulty

Impact of gradient of difficulty?

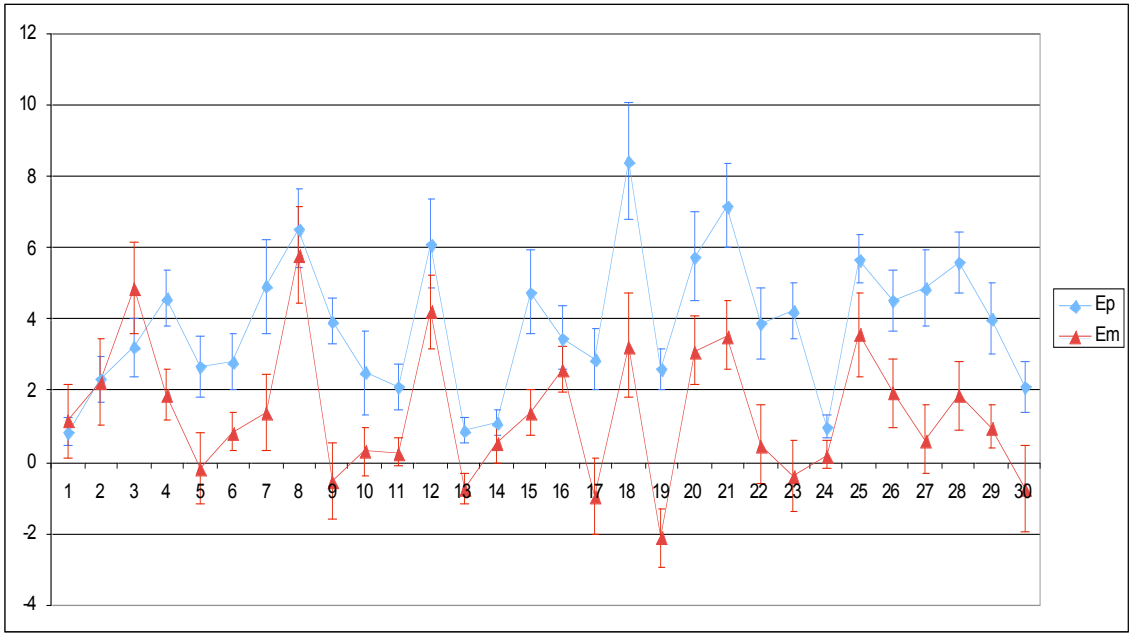
- D+ group tends to have a lesser tendency to overestimate their performance than D- group, **but not significantly** (p=0,7)

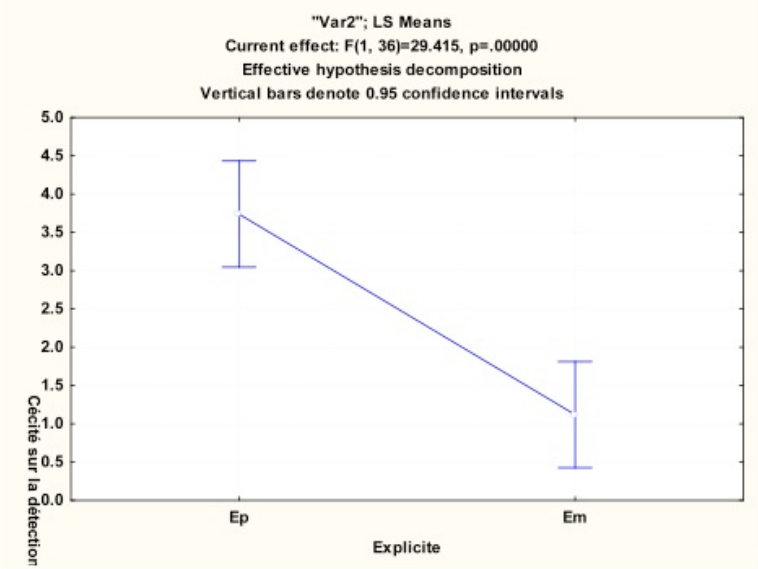




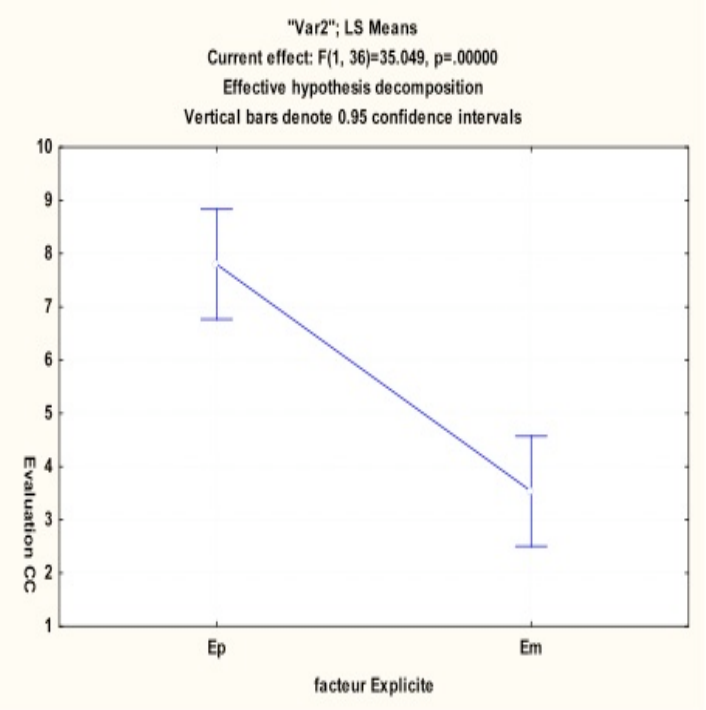
2 - Impact of explicit biased feedback

E- have less overestimation than E+ in block 2





E- have again less overestimation than E+ in post-test



Even when no feedback is delivered anymore, participants maintain self-evaluation from previous explicit feedback.

Summarizing results on self-evaluation

- Our hypothesis was that there are two sources for metaperceptual evaluation:
- Task difficulty (control-based monitoring)

- Explicit output feedback

Summarizing results on self-evaluation

- We found:

A small effect of Task difficulty (control-based monitoring)

A large effect of Explicit output feedback

New question

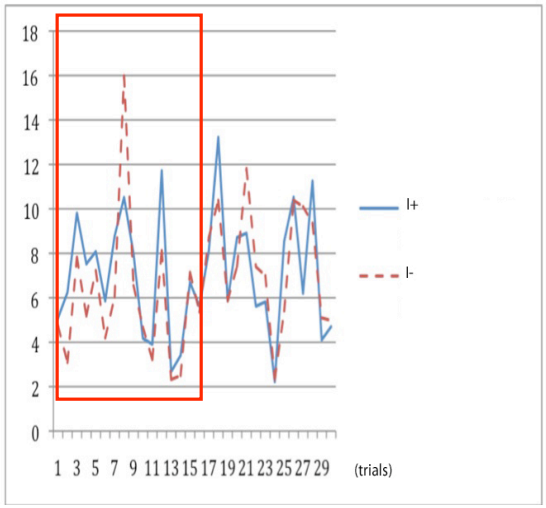
- Does each dimension of biased self-evaluation affect performance in first-order task??

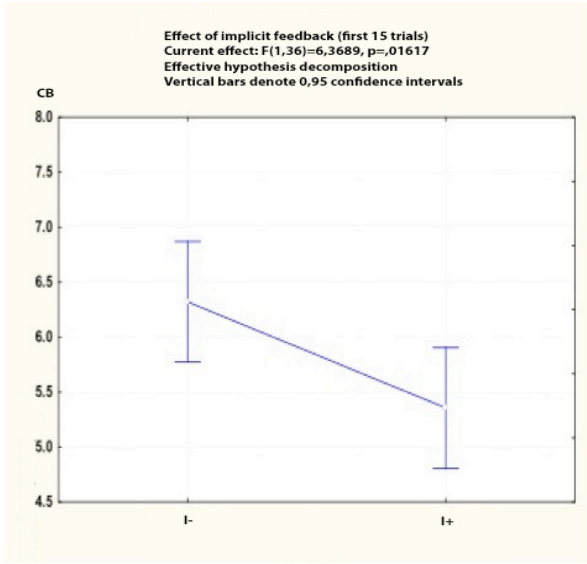
- Hypothesis 1:subjects implicitly encouraged but still not meeting the norm should be more motivated in the task than decouraged subjects, and have better performances. Carver & Scheier (1998)
- Hypothesis 2: subjects explicitly encouraged, but still not meeting the norm, should be more motivated and have better performances

3- impact of Self-evaluation on Change detection

A - Gradient of difficulty and detection

CB is affected by the gradient of difficulty in the task

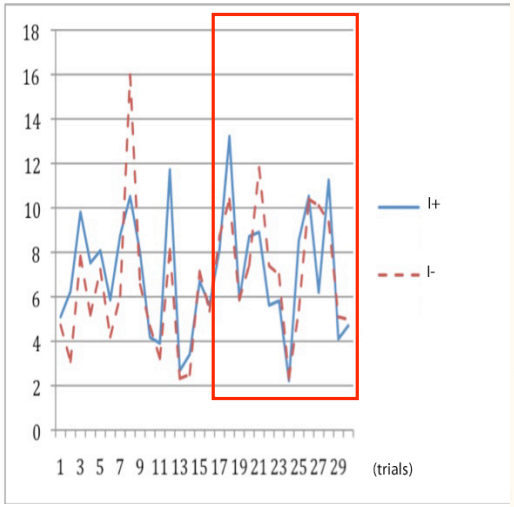


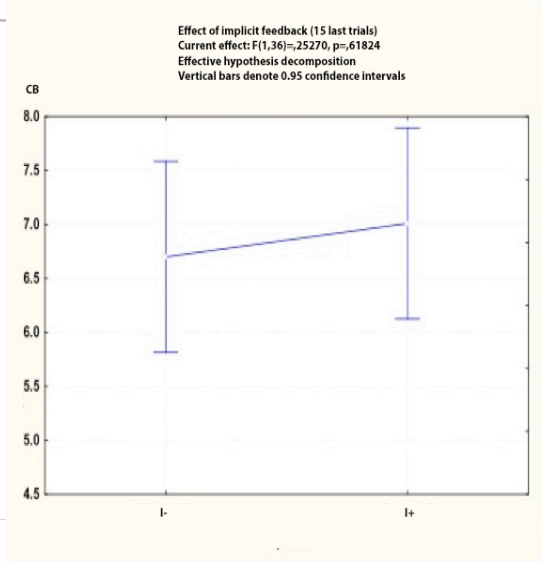


Results

- D- subjects detect better than D+, but only the first part of session 2

CB is affected by the perceived rate of progress to the goal





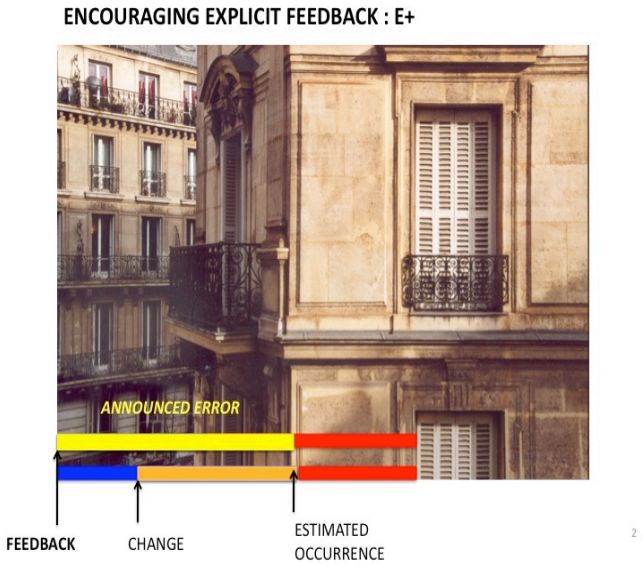
Results

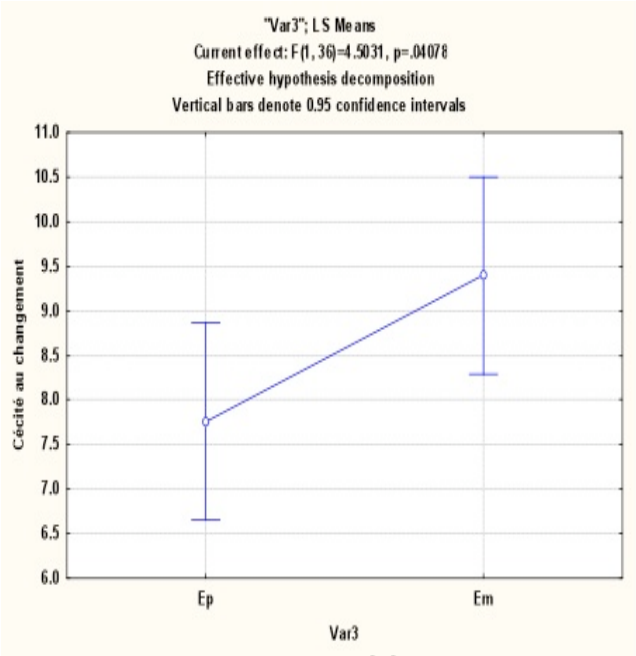
- In the second part of the session, no statistical difference are observed anymore

3- impact of Self-evaluation on Change detection

B- Biased explicit feedback and detection

Explicit feedbacks affects detection only when the task is difficult





Results

- E+ have detect better than E- for the most difficult trials

- Ceiling effect for easy trials

Summarizing results on detection performance

- Hypothesis 1 verified: A better detection is elicited by a negative gradient of difficulty (difference =1,5 sec, p=0,016)

- Hypothesis 2 verified: A better detection is elicited by a positively biased explicit feedback. (difference=2 sec, p=0,04)

Discussion

A dual-process metaperception?

These results seem compatible with a dual-process view on metaperception.

- Feeling-based (dynamics sensitive) monitoring
- Analytic monitoring (based on social feedback)

- Koriat (2000) Stanovic (2009)

Why does gradient of task difficulty influence performance?

- We assumed, with Carver & Scheier (1998), that higher motivation is associated with perceived progress (implicit decrement of difficulty level).
- Higher performance might be driven, however, by better, more engaged attentional focus in the D- condition.
- An experiment dissociating attentional and motivational factors could address this issue.

What is the mechanism for the causal influence of explicit feedback on self-evaluation?

- Is so-called « explicit evaluation » merely an effect of perceptual learning?

  (subjects merely learn the spatial interval in which the CB is likely to have taken place) ?

- This objection deserves to be discussed in the wider context of what language itself would, in contrast, allow to express.

Caveats & future research

1- We only studied biased explicit self-evaluation, in order to allow a contrast between two forms of biased dynamics.

Further research should explore the learning curve of accurate self-evaluations based on accurate vs. Inaccurate explicit feedback.

2- We studied the impact of individual explicit feedback. Future research should explore the impact of comparative social explicit feedback

3 - The effect of implicit task difficulty on explicit self-evaluation is only a tendency.

(Its stronger effect is on detection)

Philosophical conclusion: externalizing metacognition

- « Looping out of the neural domain » (A. Clark) seems to include objective dynamic characteristics of stimuli (gradient of difficulty of task), as well as contextual conditions for extracting such dynamics.
- A proposal: active externalism expresses the necessary extraction and stabilization of dynamic information over time.
- Sensorimotor embodiment amounts to just that.

THANKS FOR YOUR ATTENTION !