

# University of Cuernavaca

## April 4, 2014

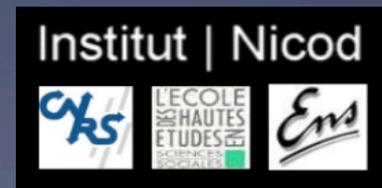
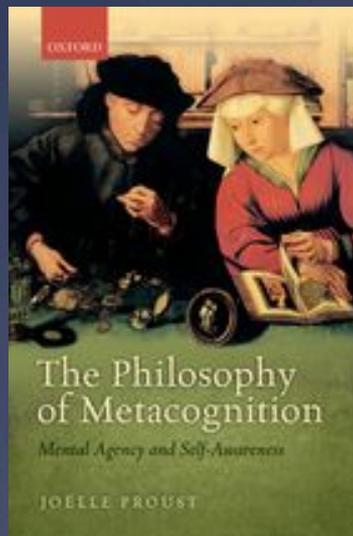
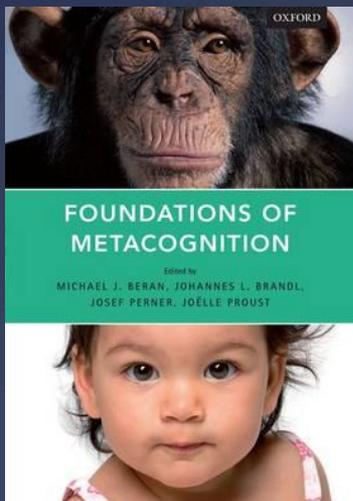
### To what extent are metacognition and mindreading interdependent?

Joëlle Proust

Institut Jean-Nicod, ENS, Paris

<http://dividnorm.ens.fr>

<http://joelleproust.org>



# What does the term “metacognition” refer to?

- In cognitive science, “metacognition” refers to the capacity through which a subject can evaluate the feasibility or completion of a given mental goal (such as learning a maze, or discriminating a signal) in a given episode (Koriat et al., 2006).

→ « **Self-evaluative** » view

- Mindreading specialists take metacognition to refer to first-person metarepresentation of one's own mental states (Carruthers 2009, 2011).

→ « **Self-attributive** » view

# Central examples of metacognition

- **Retrospective monitoring** (judging the adequacy of a response)
- **Prospective monitoring** (evaluating one's ability to carry out a cognitive task)
- **Ease of learning judgments** (reducing uncertainty on time needed to learn)
- **Knowing judgments** (reducing uncertainty about belief accuracy)
- **Monitoring emotions & motivations** (social purposes).

# Not a nominal matter

- The issue is evolutionary, developmental and functional:
  - Is metacognition made possible by, or a part, of mindreading?
  - Is it an independent ability ?
- It has philosophical relevance:
  - is self-knowledge primarily of a “theoretical” kind ?
  - Can a mental state be expressed rather than reported?
  - Does self-knowledge depend in part on non-conceptual contents?

## Origins of the self-evaluative view

- \* The self-evaluative view results from the pioneering work about cognitive control by Miller et al., and on metamemory by Hart, Nelson & Narens, and Koriat.

# Miller, Galanter and Pribram : *Plans and the structure of behavior*, (1960)

- \* The mind controls its own activity & behavior, through feedback loops, called "test-operate-test-exit" (TOTE) units.
- \* The first test phase "involves the specification of whatever knowledge is necessary for the comparison that is to be made" : ie an "incongruity-sensitive mechanism".
- \* The feedback from this test guides action: in case a *discrepancy* between present and desired end state is detected, the operation meant to suppress the discrepancy will be triggered.

# Miller, Galanter and Pribram : *Plans and the structure of behavior*, (1960)

Miller's et al. view of cognitive control has influenced

- \* Nelson & Narens' view of metacognition as control and monitoring
- \* Dan Wolpert's theory of motor control
- \* Tim Shallice's Chris Frith's views about action control and monitoring.
- \* Fleming's, Dolan's, Frith's view of metacognition as « behavior about behavior ».

# Origins of the self-attributive view

## Flavell, 1979

- \* in the next decade, child psychologist John H. Flavell, "in analogy with metalanguage", coined the words "metamemory" (1971) and "metacognition" (1979).
- \* "Metamemory", he says, refers to "the individual's knowledge and awareness of memory".
- \* "Metacognition", by analogy, refers to "knowledge and cognition about cognitive phenomena", including "attention, memory, problem solving, social cognition, and various types of self- control and self-instruction".

# Meta, as in « metalanguage »

- \* Flavell and Wellman ask in 1975 what is common to the various "metas".
- \* Their answer is that they all involve "generalizations about people and their actions vis-à-vis objects" through a reflective abstraction-like process.
- \* In their opinion, metamemory, is, like metacognition in general, "of course a form of social cognition".

# Meta, as in « metalanguage »

- \* Flavell influenced developmental psychologists and the philosopher P. Carruthers to view metacognition as based on metarepresentations of one's own mental activity (although he maintained that metacognition can be experience-based, mental concepts need to be mastered)
  - \* Josef Perner
  - \* Wolfgang Schneider
  - \* Beate Sodian
  - \* Janet Metcalfe
  - \* Peter Carruthers

# A neutral definition

- \* If one wants to adjudicate between the two definitions of metacognition on offer, one needs to use a neutral definition that encompasses both types of phenomena.

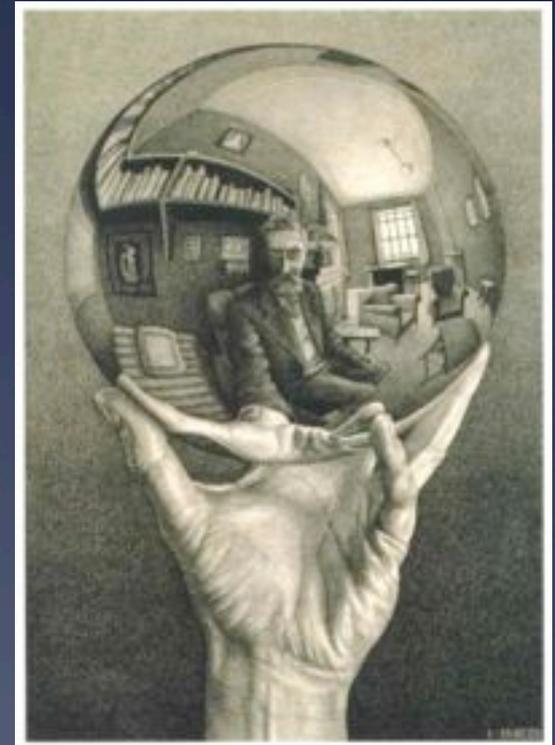
Proposal:

- \* Def. Metacognition is the set of capacities through which an operating cognitive subsystem is evaluated or represented by another subsystem in a context-sensitive way.

# Outline

1. The self-attributive view of metacognition
2. Empirical evidence against the self-attributive view
3. The self-evaluative view of metacognition
4. Metacognition & mental agency
5. Do metacognition and mindreading interact?

# The self-attributive view of metacognition



# 4 main claims

(1) The evaluating subsystem and the evaluated subsystem need not belong to the same organism.

(2) Evaluation is performed propositionally, through metarepresentational or mindreading processes.

(3) Evaluation always requires an ability to represent mental attitudes as such in every form of metacognitive control and monitoring.

(4) Metarepresentation plays a crucial role in higher forms of agency, but is not functionally dependent on agency.

# (1) The evaluating subsystem and the evaluated subsystem need not belong to the same organism.

- \* It is claimed, after Flavell, that metacognition belongs to social cognition.
- \* The latter, however, can be turned inwards, in a Vygotskian type of internalization.
- \* Evaluation is derived from a conceptual, “theoretical” categorizing of the attitudes involved in one’s mental activity (also applying to others).

## (2) Evaluation/representation is performed through metarepresentations

Metacognition coincides with the acquisition, or possession, of second-order propositional attitudes such as "I believe that I believe that  $P$ ", "I believe that I intend to  $F$  etc".

E.g.

Self-attributing or evaluating a belief requires recognizing a first-order occurrent belief *as a belief*.

## (2) Evaluation/representation is performed through metarepresentations

### **Metarepresentations are formed by a mindreading device**

A specialized representational device takes an occurrent thought content  $P$  as input, and produces the embedding representation "I believe" (or "perceive", or "imagine", etc.) that  $P$  as output.

## (2) Evaluation/representation is performed through metarepresentations

Theoretical variants: the device can be

- Neutral as to self-or other usage (Dienes & Perner, 2001, Carruthers, 2009)
- Start with a simulation in self (Goldman, 2006)
- Associated with an executive capacity for decoupling representations (Russell, 1996).

# Carruthers, BBS, 2009



“Our access to our own propositional attitudes is always interpretative” (rather than introspectable), even though “the evidence base for self-interpretation is somewhat wider than we normally have available when interpreting other people”

(p. 124)

## Carruthers, BBS, 2009

Peculiar access to one's mental contents does not make it special or privileged access: **inferences are always needed.**

- \* One may access one's thought contents on the basis of one's motor and linguistic behavior, on the basis of inner speech and rule application, or on the joint basis of inner speech, patterns of attention and emotion, and self-interpretation (Carruthers, 2009).

### (3) Evaluation requires an ability to represent mental attitudes as such

- \* In order to have a critical attitude toward their own cognitive states as well as to others, children must be able to first represent the world-representation link and recognize that, in certain circumstances, it is false or illusory.
- \* *Evaluating whether a remembered name is correct* requires recognizing a first-order occurrent belief *as a memory* (Perner & Ruffman, 1995)
- \* Children tested on various forms of cognitive control, self-evaluation, and source monitoring have trouble distinguishing the perceptual appearance from the real nature of objects (such as a sponge that looks like a rock) before they reach 4-5 years of age (Flavell, 1979)

### (3) Evaluation requires an ability to represent mental attitudes as such

- \* *Attributivism* might take the first step in self-evaluation as a simulation, or a re-representation of one's first-order cognitive state. One looks at the world, and forms a belief, for example, about whether *P*.
- \* The second step, self-evaluation through a circumstance shift, can be performed by tagging the first-order content through attitude concepts. Thanks to them, the thinker is now able to discriminate the epistemic demands of perceiving versus imagining, remembering, desiring.

### (3) Evaluation requires an ability to represent mental attitudes as such

- \* The is justified to claim that it is only when one conceptually understands what is involved in, e.g., knowing versus merely believing, that one becomes able to evaluate the correctness of an epistemic self-evaluation such as "I am certain that I know that p".

## (4) Metarepresentations play a crucial role in higher forms of agency, but is not functionally dependent on agency.

Mindreading theorists generally do not think that acting contributes to action-understanding in a way that merely observing others' actions would not.

In other terms, no procedural information is gained in action that a theoretical belief could not capture.

For mindreading theorists, action does not provide predictive or self-evaluative cues that can be directly introspected by the agent.

Peter Carruthers (2009b), claims that all the cues involved in self-understanding and self-evaluation are, rather, of an inferential kind (where inference is taken to be concept-based)

## (4) Metarepresentation plays a crucial role in higher forms of agency.

**Metarepresenting ones' attitudes is necessary to the control and monitoring of one's own thinking because of the representational structure of higher agency**

Knowledge about cognitive fallibility is seen as providing agents with the proper ability, and rational motivation, needed to monitor and correct their cognitive states.

**Thought control belongs to higher forms of agency.**

**In other terms: thought control has to be mediated by attitude concepts.**

Metarepresenting intentions and beliefs allows an agent to resist interference from the environment and to pursue endogeneous goals. (Shallice, 1988, Perner & Lang, 1999)

## (4) Metarepresentation makes higher forms of agency possible: Shallice 2008

- \* . The contention scheduling system (CSS), a low-level system, activates effectors on the basis of environmental affordances.
  - \* A higher-level form of control, called the Supervisory Attentional System (SAS) triggers non-routine actions, on the basis of a metarepresentation of the organism's intentions and cognitive capacities.
- a metarepresentational capacity the key to the ability to control and inhibit routine actions in a context-sensitive way.

**Against SAV: evidence for non  
metarepresentational  
metacognition in non-humans  
and in human children**

## 2 types of evidence in favor of non-metarepresentational, i.e. “procedural” metacognition

1. Evidence for procedural metacognition in rhesus monkeys.
2. 3 year-old children present the same procedural metacognition as rhesus monkeys.

# A – comparative psychology



# The strategic importance of comparative studies

Allow dissociating

- \* A **metarepresentational** capacity (assumed not present in non-humans)
- \* A **metacognitive** capacity (presumed present)
- \* A **motor control** capacity (present in most animals)

# Central methodological problem

The difficulty consists in showing that monkeys and apes

- \* monitor their uncertainty (internal cues)
- \* **rather than** learn how the world is.

## Two main types of tasks eliciting Judgments of uncertainty

- \* Seeking for information (SI) tasks:

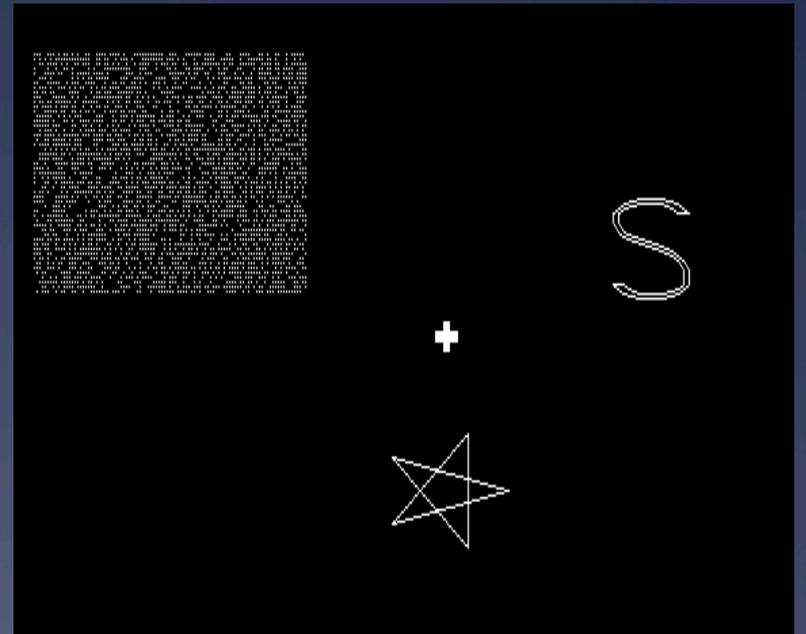
Will an animal ask for information only when needed ? (Call & Carpenter, 2001)

- \* Discrimination judgments in a choose-or-decline-to-respond (« opt out ») paradigm

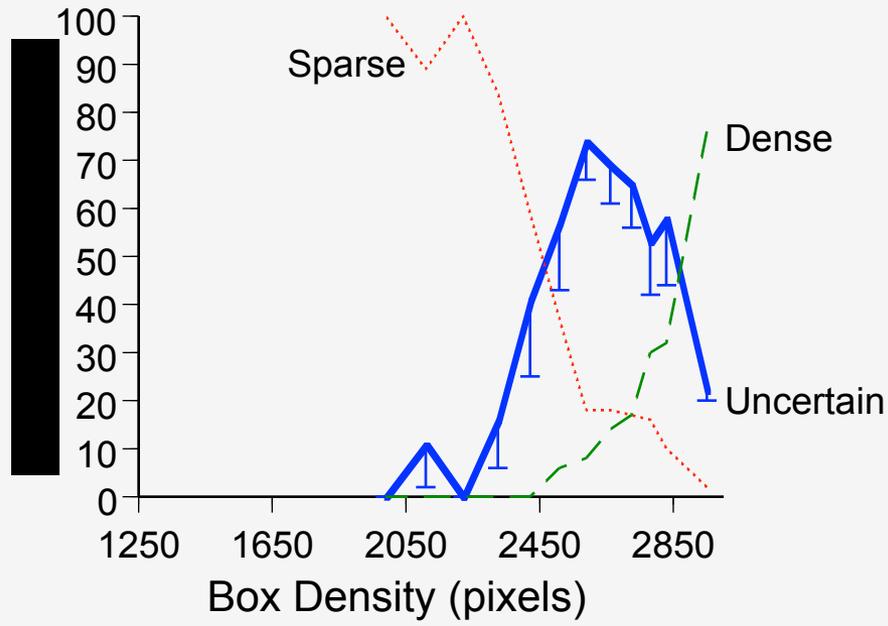
- \* Will an animal choose to decline mostly for difficult stimuli ?

# Smith and/or coll. on metacognition in monkeys

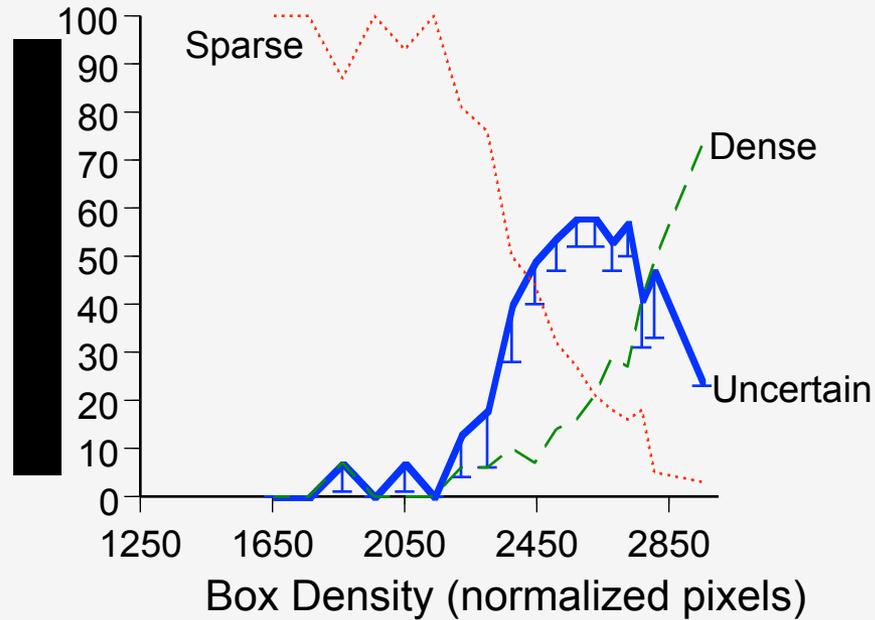
- \* Rhesus monkeys decline most the most difficult trials in **visual discrimination tasks** (Shield, Smith & Washburn, 1997) and in **memory tasks** (Hampton, 2001).
- \* They **generalize** their U- responses to new tasks. (Washburn, Smith & Shields, 2006)
- \* Macaques also use U-responses with **blocked feedback** (Beran, Smith, Redford & Washburn, 2006)



## Monkey



## Humans



# Metacognition in Phylogeny:

## Yes

- \* **Primates:**
  - \* Apes: chimps and orangutans search for info (SI)
- Rhesus macaques (SI & U-R)
- \* **Marine mammals:**
  - \* Bottle-nosed dolphins U-R
- \* **Rats:** Foote & Crystal (2007) U-R in auditory discrimination task

## No

- \* **Pigeons** no U-R (Sutton & Shettleworth, 2008)
- \* **Capuchin monkeys:** no SI, no U-R (Beran et al. 2006)
- \* **Rats:** Smith & Scholl (unpub.), Smith et al. 2007 (no U-R)

# Methodological problems:

What does the U-response exactly mean in an opt-out paradigm ?

- \* - the animal prefers to evade difficult trials, involving time-out penalty for each failure?
- \* (when provided direct feedback): the animal is conditioned to offer a U-response for a class of stimuli ?
- \* the animal rationally evaluates how well it perceives or remembers given the task's constraints?

# 4 possible ways of interpreting a response in a context of uncertainty:

Property of the stimulus relative to the frequency range of the stimulus class:

1. Middle range is **objectively uncertain**
2. middle range responses are directly rewarded (+ cond.)
3. middle range responses are directly punished (- cond)
4. Middle range is **subjectively uncertain**: ie not bound to stimulus or to R-conditioning

# In favor of interpretation 4

The properties of the observed responses

- \* Are not **just cognitive**, ie. not stimulus-bound
- \* They **generalize** to new stimuli and new tasks **without** new learning (Kornell & al, 2007)
- \* **Distinctive pattern**: « fragile & changeable », also in humans.
- \* They suppose access to a metacognitive feeling - e.g. **a feeling of uncertainty**

# Main recent findings

- \* New World monkeys (capuchins) learn middle responses when selectively rewarded but **don't produce metacognitive responses** (ie, don't use the "?" response) when given no feedback, in contrast with Old World monkeys (rhesus monkeys).
- \* There is a **dissociation** between the U-responses and the middle responses; they differ in motivational strength.

# Conclusion

- \* Animal studies show that procedural metacognition exists in the absence of mindreading.
- \* The source of information used to monitor perception or memory cannot consist in conceptual knowledge about own attitudes.

# B – Developmental psychology



## **Dominant view among mindreading theorists**

- \* 3 yr olds don't form judgments of uncertainty because they can only attribute true beliefs to self and others.**

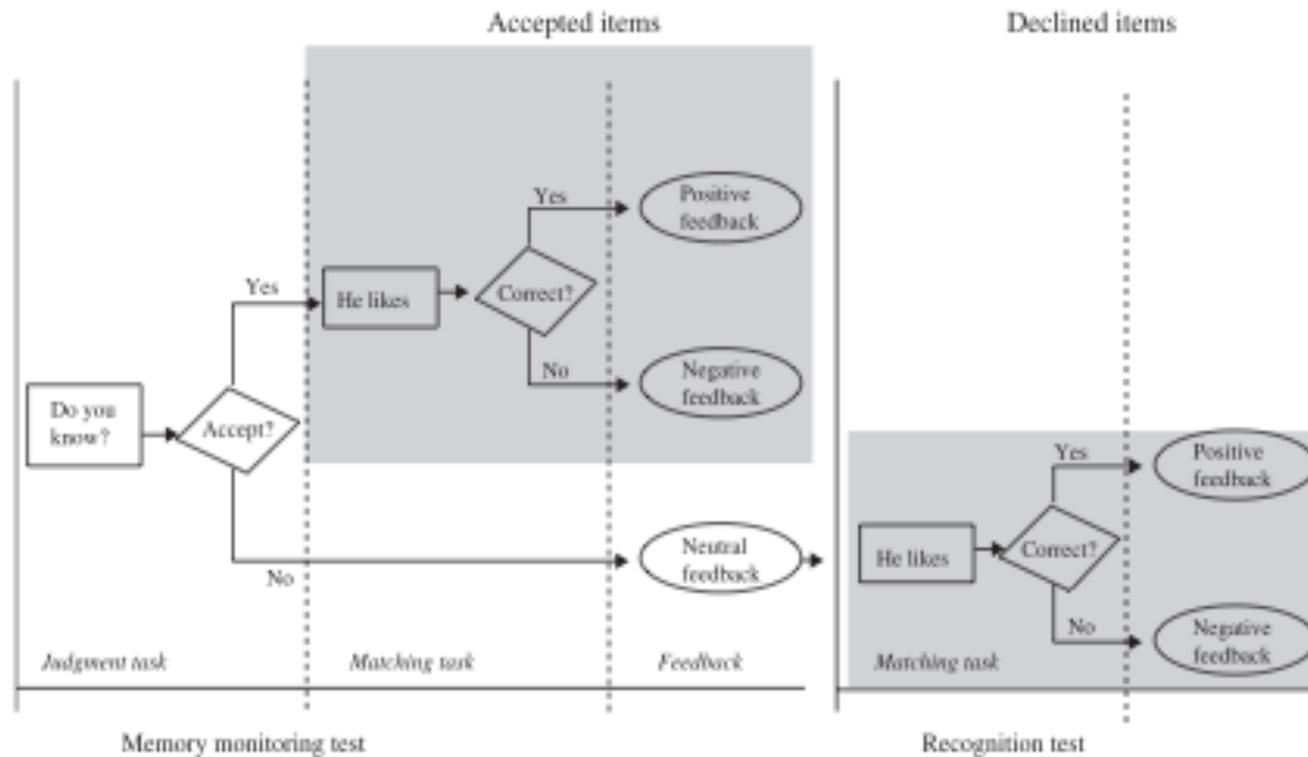
## 3 year-old children present the same procedural metacognition as rhesus monkeys.

- One study used an opt out paradigm similar to Smith et al.'s (2003), to test children aged 3:5
- 3 yr-olds typically fail to form correct self-attribution of knowledge when tested verbally
- Can they still have an implicit access to their own knowledge states?



(Balcomb & Gerken, 2008).

# Opt out paradigm



**Figure 1** Test procedures for Experiments 1 and 2. Although children took all trials on the recognition task, data for accepted items were taken from the memory-monitoring test, and data for declined items were taken from the recognition test (indicated in grey).

(Balcomb & Gerken, 2008)

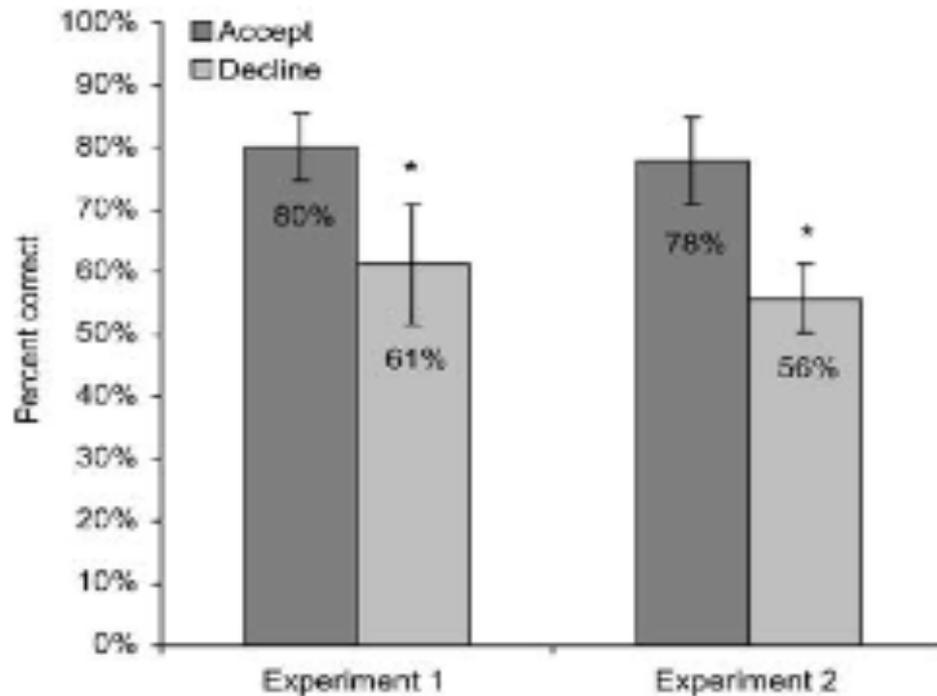
# Results: 3 year-old children present the same procedural metacognition as rhesus monkeys.

- \* Children were able to predict reliably how well they would remember a given item
- \* They presented higher performances when they chose to answer rather than in forced-choice trials.



(Balcomb & Gerken, 2008).

# Higher performances when allowed to opt out



**Figure 2** *Percent correct for accept type trials vs. decline on Experiments 1 and 2.*

- \* Exp 1: children asked to form recognition judgments in presence of target, match and distracter
- \* Exp 2: same, except that judgments are formed in presence of the match only.

# Children's and monkeys' success in opt-out tasks

raises the question: what is the informational source that is used to make an appropriate metacognitive decision?

→ hypothesis: it is an activity-dependent, subpersonal type of information

An alternative concept of  
metacognition

The self-evaluative view

# The self-evaluative view of metacognition



# Main claim of the Self-evaluative view

The information used in metacognition does not need to be of a theoretical nature

It may simply consist in activity-dependent feedback

# Metacognitive monitoring and Epistemic norms.

Sensitivity to epistemic norms is to be found in metacognitive activity:

- \* when **predicting** one's cognitive dispositions (in order to control one's cognition) (Can I recall X?)
- \* When **retrospectively evaluating** the epistemic outcome of one's cognitive performance (Am I confident of being correct?)

# Self-probing

Before trying to act mentally, one needs to know whether, e.g.,

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)

# Post-evaluation

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

# Self-evaluative view : 4 claims

- (1) The operating subsystem and the evaluative subsystem belong to the same organism.
- (2) Evaluation is performed dynamically, through adaptive control, i.e. monitoring- based control.
- (3) Evaluation, i.e., dynamic control, does not need to include an ability to represent mental states as such, but does include it in higher forms of control.
- (4) Metacognition is an ingredient of cognitive, or mental, agency.

# (1) The operating subsystem and the evaluative subsystem belong to the same organism.

- \* Cognitive systems can orient their attention to many objects in order to learn or retrieve relevant facts
- \* They do so, however, with **limited resources**
- \* A subsystem designed to assess uncertainty about its own cognitive dispositions or outputs allows it to allocate its informational resources in an optimal way.

## (2) Evaluation is performed dynamically, through adaptive control, i.e. monitoring- based control.

Control systems involve a loop in which information has a two-way flow.

- \* top-down flow: a command is selected and sent to an effector.
- \* bottom-up flow: reafferences (i.e. feedback generated by the former command) inform the control level of the adequacy of the activated command.

→ What is crucial in any control system is the fact that observed feedback can be *compared* with expected feedback.

### (3) Evaluation only needs to represent mental states as such in higher forms of control.

- In the activity-dependent control of memory and perception: evaluation is conducted on the basis of feelings
  - Metacognitive feelings may, in humans, result in a *judgment that one feels confident to a degree in one's perception.*
- There are higher forms of metacognition (“analytic metacognition” based on metarepresentations of one’s cognitive dispositions or outcomes and on conceptual inferences, rather than on feelings.

## (4) Metacognition is an ingredient of mental agency

Two related features of metacognition are essential, from an evaluative viewpoint:

- Context sensitivity
- Activity-dependence:
  - engaging in a first-order task with the concern of performing well allows agents to attend to activity-dependent cues that would not otherwise be available

# Examples of mental agency

## Purely Epistemic

Perceptual attending

Directed reasoning

Directed memory  
retrieval

Directed visualizing

Directed imagining

## Non purely epistemic

Planning

Reflective deciding

Controlling emotion

Preference management

**How can animals and human  
children perform mental  
actions?**

# Feelings of confidence are used in monitoring mental actions

## **predictive**

- mental effort
- Familiarity
- knowing
- Uncertainty about competence
- Tip of the tongue

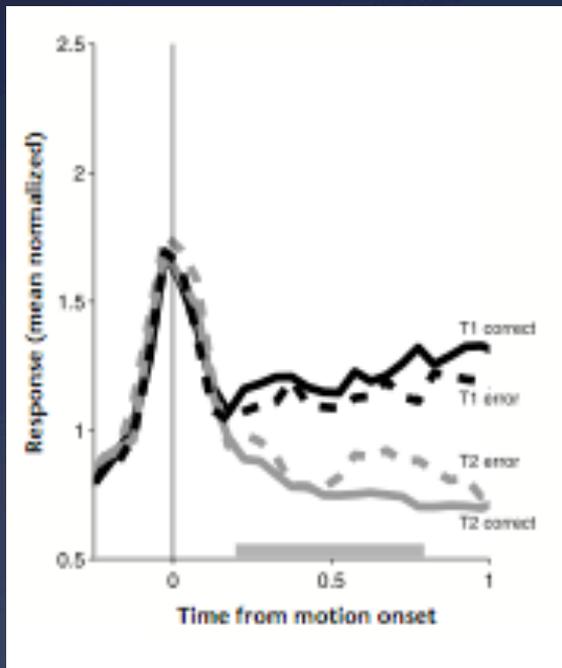
## **retrospective**

- Coherence, incoherence
- Informativeness
- Being right
- Beauty or harmony
- Uncertainty about performance

# What is the informational basis of noetic feelings?

- \* Their basis is the neural vehicle of the first-order task:
- \* Processing onset, latency, intensity and increased coherence of cognitive activity over time together **predict cognitive success.**

# How to implicitly access one's own uncertainty?



- \* Neurons predicting a saccade in a given direction respond less strongly in the erroneous decisions (dashed lines).
- \* This difference could be the basis for feelings of confidence about decision in a task.

Kim & Shadlen, *Nature Neuroscience*, 1999



# How to implicitly access one's own uncertainty?

## The accumulator model

. Evidence for the two alternatives is accumulated in parallel, until one of the evidence totals reaches a criterion value, and the associated response is emitted.

Vickers & Lee, 1998

# The neural correlates of procedural metacognition in rhesus monkeys.



were studied in an opt-out task, where monkeys must

- \* discriminate whether a shortly presented stimulus is moving left or right.
- \* respond, after a delay, with an eye movement.
- \* “Sure bet” option available in some trials



(Kiani & Shadlen, *Science*, 2009)

# Kiani & Shadlen, *Nature Neuroscience*, 2009



- \* They found that the firing rate of neurons in the lateral intraparietal cortex (LIP) correlates with the accumulation of evidence, and the degree of certainty underlying the decision to opt out.
- \* This result fits nicely with an accumulator model of judgments of self-confidence.

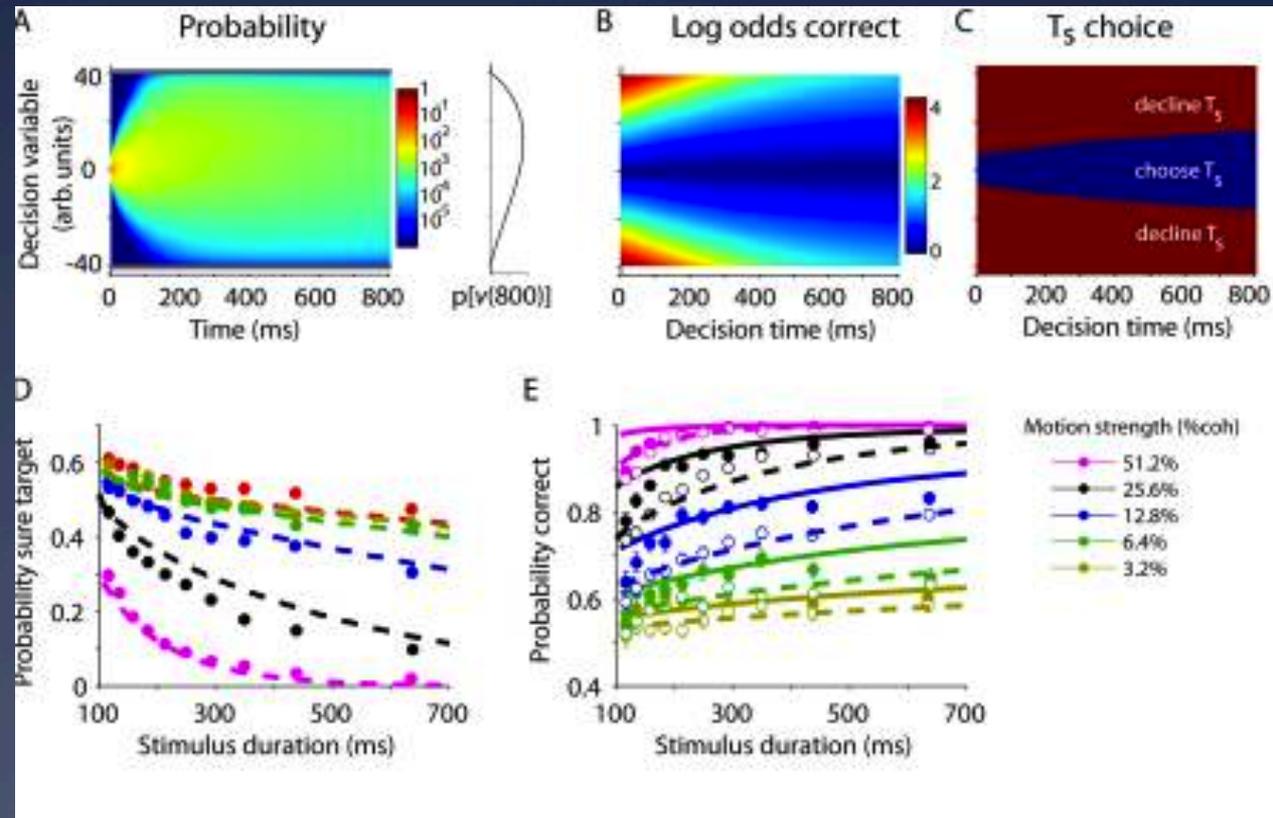


# Behavior reflects appropriate confidence in judgment

\* Monkeys opt for sure target when the chance of making a correct decision is small (short stimulus durations) (Fig D)

\* Better accuracy when the monkeys waived the opt-out option than in trials when no option was offered (dashed line in fig E)

\* Kiani & Shadlen, *Nature Neuroscience*, 1999



# Summary

- \* There is evidence that rhesus monkeys and human 3 yr-olds can reliably and flexibly decide when to perform a cognitive task.
- \* Rhesus monkeys do not present mindreading abilities, and 3 yr-olds have only incipient forms of mental reasoning.
- \* An accumulator model might be used by the brain to extract predictive, probabilistic cues about performance success, **which generate feelings motivating the proper course of action.**

# Metacognition and mindreading From dissociation to interaction

# A dissociation between procedural metacognition and mindreading in human adults

Accuracy of a judgment of learning\* (JOL) about self or other crucially depends on observers' having been allowed to perform the first-order task *before* they form a confidence judgment for this trial, whether concerning their own, or another's performance

(Koriat & Ackermann, 2010).

\* A judgment of learning is one that predicts how well the learner will be able to remember a particular studied item after a delay.



# A remarkable dissociation

## Off-line evaluation

### Mindreading

- \* participants rely on the naïve, **incorrect theory** that longer study time predicts better performance
- \* in a self-paced learning task, devoting more time to a pair of words is taken to predict **better retrieval** for that pair.

## On-line evaluation

### Procedural metacognition

- \* participants **judge correctly** that longer study time predicts poorer performance
- \* "memorizing effort heuristic", based on dynamic cues such as time spent and rate of accumulation of evidence.

(Koriat & Ackermann, 2010)

# A striking contrast in ways of appraisal

1. Vehicle-sensitive noetic feelings (sensitivity structured by the neural properties underlying fluency) influence a range of epistemic decisions.
2. Content-sensitive epistemic decisions do not seem to elicit sui-generis feelings.

# Does mindreading influence metacognition?

- \* YES: by allowing **analytic metacognition**, based on beliefs about perception, belief acquisition, learning, cognitive competences of self, to develop its own strategies and extract its own regularities.

# Does procedural metacognition influence mindreading?

YES

- by orienting social reasoning to shared (consensual) fluent and relevant targets.
- In conversation: by restricting the number of the inferences and implicatures to those that are easy to process and salient.

# Epistemic norms are distributed in two mc systems

## System 1: procedural

### \* Fluency (familiarity)

- Consensus
- Intelligibility
- Coherence
- Informativeness ?

## System 2: analytic

- \* Accuracy (memory, reasoning)
- \* Comprehensiveness or exhaustivity (memory, reasoning)
- \* Coherence (fiction, demonstrative reasoning)
- \* Consensus (negociation, deference to authority )
- \* Relevance (conversation),
- \* Informativeness (conversation)
- \* Plausibility

# A contrast between norms.

- Fluency is a norm that is inherent to processing, and that gives rise to specific feelings (familiarity, confidence in perception or in memory)
  - Other norms, such as truth or plausibility, are inherent to evaluating a cognitive content.
- Contrast between two forms of metacognitive norms: experience-based (procedural) and concept-based (analytic).

# The 2-system view revised

- \* Granting that System 1 generates nonconceptual contents in a featural format, the contrast with System 2 is one between two ways of forming and using representations.

# System 1

- \* Vehicle-based
- \* Inflexible
- \* Economical
- \* Nonconceptual
- \* Gradient structure
- \* Modular
- \* Non inferential

# System 2

- Content-based
- Flexible
- Costly
- Conceptual
- Componential structure
- Non-modular
- Inferential

# Inflexibility

- \* Inflexibility has nothing to do with the fact that feelings are « generated by subpersonal processes ». All our flexible thoughts are also generated subpersonally.
- \* System1 inflexibility derives, rather, from the *nonconceptual format of representation that is used to drive decision.*

# What kind of binding is there between S1 and S2?

- \* The binding between the two systems is the same as that studied in the philosophy of perception between nonconceptual protopropositional content, and propositional content.

# What kind of binding is there between S1 and S2?

- \* The nonconceptual content of perception is inserted within a propositional format including terms for concepts and objects.
- \* Analogously, children's NFs are redescribed in conceptual terms.

# What kind of binding is there between S1 and S2?

- \* When a System 2 is present, agents have access to propositional representations of their cognitive goals, and can assess their cognitive resources under new types of norms.
- \* Although this assessment may take advantage of NFs (e.g.: intelligibility), it is not mainly based on cognitive emotions.

Noetic feelings motivate an agent to launch a given command,

they can, however, be **overridden by conceptual representations** of the task and/or of self-competence.

Noetic feelings (**e.g. fluency → accuracy**) can be normatively **reinterpreted**, (**e.g. fluency → inaccuracy**),

They cannot be suppressed. (Nussinson & Koriat, 2008)

**THANK YOU FOR YOUR ATTENTION !**

This presentation is available for download on :

<http://www.joelleproust.org>

# Metarepresentation

- \* Essentially Reflexive
- \* Engaged processing (simulation)
- \* Poorly recursive
- \* No decoupling
- \* Representational promiscuity
- \* No inferential promiscuity
- \* Predictive-evaluative function

# Metacognition

- \* No essential reflexivity
- \* Disengaged processing (shallowness possible)
- \* Fully recursive
- \* Decoupling involved
- \* No representational promiscuity
- \* Inferential promiscuity
- \* Predictive-attributive function