Amsterdam Graduate Philosophy Conference
December 17, 2011

Is metacognition a form of self-interpretation?

Joëlle Proust

Institut Jean-Nicod, Paris

jproust@ehess.fr

Subject of this talk: 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 (Perner, Carruthers 2009).

→ **« Self-attributive » view**

Not a nominal matter

- The issue is evolutionary, developmental and functional:
  - Is metacognition made possible, or a part, of mindreading?
  - Is it an independent ability ?

- It has philosophical relevance:
  - is self-knowledge primarily of a “theoretical” kind ?
  - Does self-knowledge primarily depend on non-conceptual contents?

Outline

1. The self-evaluative view about metacognition
2. The self-attributive view about metacognition
3. Three arguments in favor of recognizing two independent functions
4. Has a dual-process view about metacognition epistemological relevance ?

The self-evaluative view about metacognition

Metacognition: a self-evaluative definition

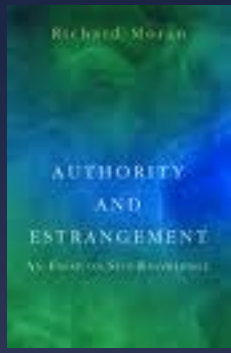
Metacognition is the set of abilities that allow humans (and some non-humans)

- to evaluate the cognitive adequacy of their dispositions to obtain a given cognitive output and
- to control their cognition accordingly.

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).

Main features of the self-evaluative view



1 - Contrast between a descriptive and engaged view about one's mental states

Self-evaluation is associated with the normative awareness that one is committed to the truth, coherence, relevance, etc. of one's beliefs, to the rationality of one's intentional actions, etc. (Moran, 2001)

Main features of the self-evaluative view

2- Metacognition is part of mental agency

- A bodily action is a controlled behavior which one intends to result in some physical change in the world or in the self
- A mental action is a controlled sequence of mental operations which one intends to result in some cognitive change.
- When acting, one needs to proportionate one's goals to one's mental as well as bodily resources.

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

Main features of the self-evaluative view

3 – Metacognitive evaluations decompose in two steps, before and after performing a mental action.

The function of metacognitive monitoring is, in conjunction with instrumental considerations, to drive metacognitive control

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

Main features of the self-evaluative view

4 – Metacognitive self-evaluations necessarily involve comparing an observed with an expected value.

- Self-probing: they predict how uncertain it is that the action is feasible.
- Post-evaluating: they report how uncertain it is that the action is successful.

Main features of the self-evaluative view

5 - Metacognitive control consists in deciding whether, given the monitoring status (how divergent the predicted/observed values are):

- ✓ the envisaged mental action should be taken (self-probing level)
- ✓ its outcome should be used in acting on the world (post-evaluating level).

Main features of the self-evaluative view

6 - Monitoring status and the ensuing control decision are respectively expressed by non-conceptual representations.

- The output of the comparator is made available to the agent in the form of dedicated feeling.
- These feelings may, but don’t need to be redescribed in metarepresentational terms, in order to motivate a given decision.

Feelings of ..

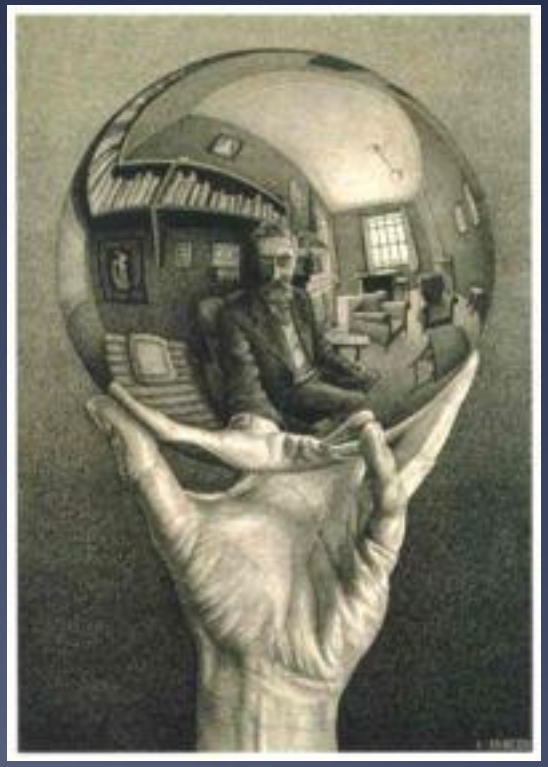
- mental effort
- Familiarity
- knowing
- Tip of the tongue
- Coherence, incoherence
- Being right
- Beauty or harmony
- Uncertainty about performance
- Uncertainty about competence

Main features of the self-evaluative view

7 - Although the feelings motivate the agent to a select a given command, they can be overriden by conceptual representations of the task and/or of self-competence.

The feelings can also be normatively reinterpreted (e.g. fluency / accuracy), but cannot be suppressed.

The self-attributive view about metacognition



Main features of the self-attributive view

1- **The semantic condition**

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 a belief requires recognizing a first-order occurrent belief as a belief.

Main features of the self-attributive view

2- Metarepresentations are formed by a mindreading device or by a theory of mind

- 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.

Main features of the self-attributive view

2- Metarepresentations are formed by a mindreading device or by a theory of mind

- 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).

Main features of the self-attributive view

3- Metarepresentations are used in the same detached sense when attributing thoughts to self or to other.

Detachment is a feature derived from a conceptual, “theoretical” categorizing of attitudes being involved in both cases (“generality principle”).

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"

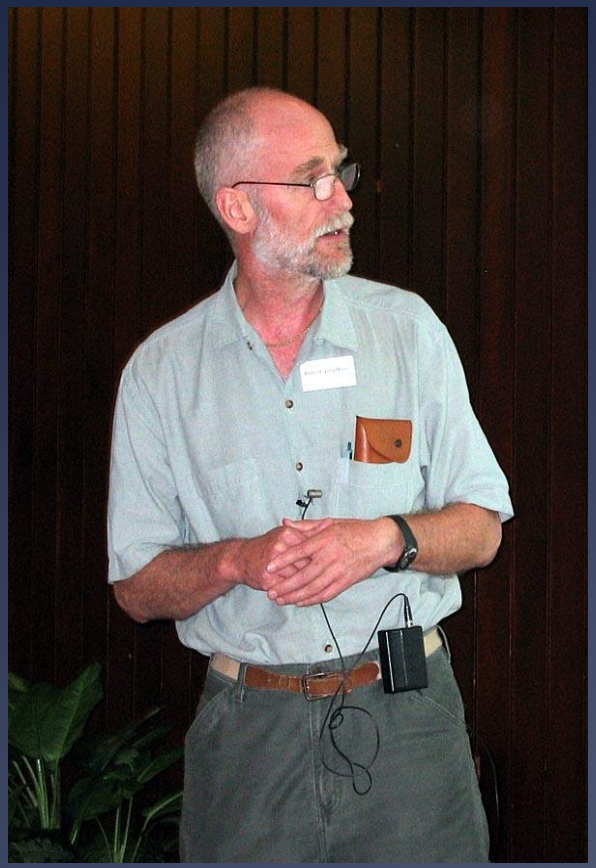
(p. 124)

Main features of the self-attributive view

4 – 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).

Carruthers, BBS, 2009



Perceptual Judgements **are directly** introspectable: “we see an object as a man or as bending over”.

Introspectibility derives from “being received as input by the mindreading system(by virtue of being globally broadcast”) (p.125)

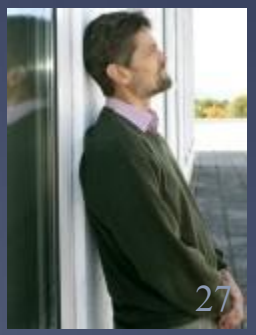
Main features of the self-attributive view

5 - Metarepresenting ones' attitudes necessary to the control and monitoring of one's own thinking.

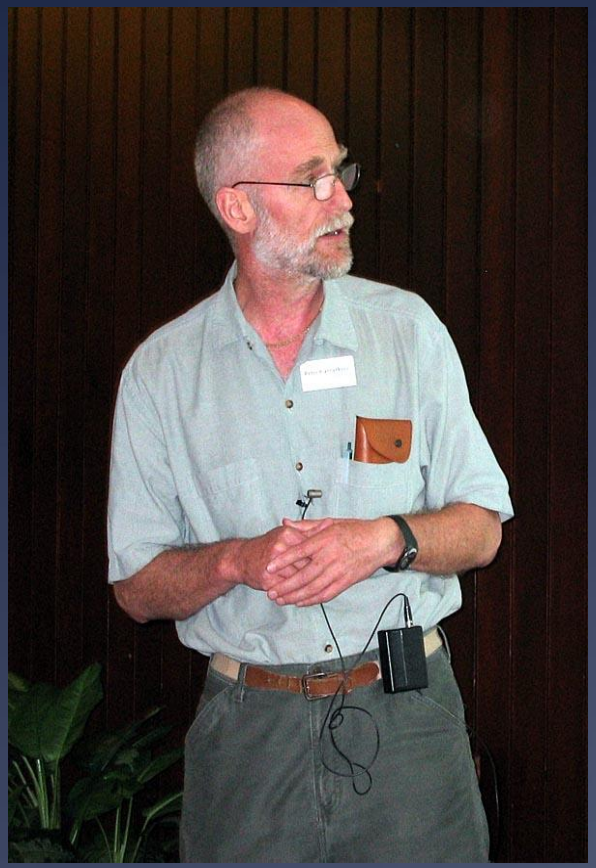
In other terms: thought control has to be mediated by concepts.

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

.



Carruthers, BBS, 2009



« Our metacognitive interventions don't require introspection; they have no direct impact on cognitive processing that would be predicted if metacognition had, indeed, evolved for the purpose » (p.129)

« For ex., most metamemory capacities only require an ability to initiate or to intervene *in behavior ».*

Discussion

If the self-attributive view is right, then:

- Either metacognition is not « meta » at all.
  - metacognitive processes are confined to the control of behavior.
- Or metacognition is indeed « meta »: it operates over cognitive rather than motor states, and should be metarepresentational.
  - Then non-humans, having no metarepresentations, should have no flexible cognitive control of their own cognitive dispositions

Discussion

If the self-attributive view is right, furthermore,

- There should be no substantial change in the evaluative processes when applied to self or to others.

- There should be no specialized neural activity associated with activity-dependent cognitive control and monitoring.

In summary: 4 arguments in favor of a 2-function view

1. Metacognition does not control behavior, but epistemic decisions
2. Metacognition is present in organisms unable to metarepresent (« procedural » metacognition).
3. In humans, procedural and concept-based metacognition lead to different epistemic evaluations and decisions.
4. The neural correlates for self-evaluation and self/other-attribution are different.

1 - Metacognition does not control **behavior, but epistemic decisions**

Are metacognitive processes confined to the control of behavior ? As was shown before, they are rather involved in the control of *cognitive actions*.

Phylogenetic difference:

- Most animals know how to control their actions on the environment, thanks to sensorimotor feedback.
- Few know how to control their cognitive actions (regulate their cognition). Another type of feedback needs to be extracted and used.

2 - Metacognition is present in organisms unable to metarepresent

- Non-humans, having no metarepresentations, should have no flexible cognitive control of their own cognitive dispositions
- There is now strong converging evidence that they do.

Robert Hampton (2009): Objective markers for metacognitive behavior:

1. There must be a primary behavior that can be scored for its *accuracy*.
2. *Variation* in performance (i.e. uncertainty about outcome) must be present.
3. A secondary behavior, whose goal is to *regulate* the primary behavior, must be elicited in the animal.
4. This secondary behavior must be shown to benefit performance in the primary task (for example, animals must decline tests that they would otherwise have failed).

Paradigms able to test procedural metacognition

- The *retrospective gambling paradigm* (also called “wagering”)
- Forms of the *prospective opt-out test*, where animals are asked to decide whether or not to perform the task without simultaneously perceiving the test stimuli (Hampton, 2009).

Paradigms unable to test procedural metacognition

* Seeking for information:

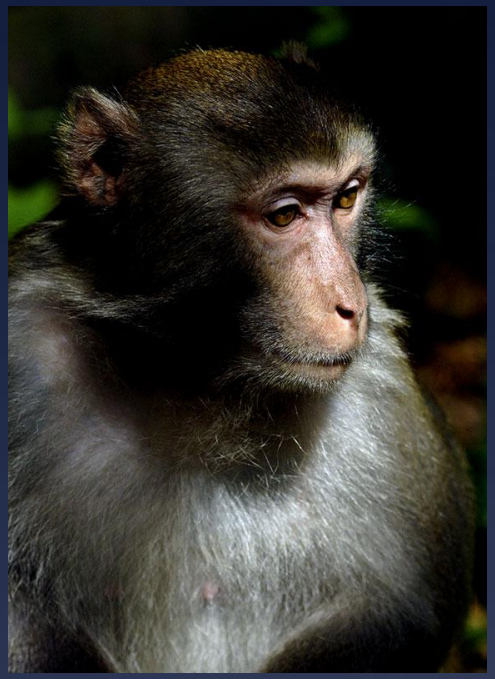
Will an animal ask for information only when needed ? (Call & Carpenter, 2001)

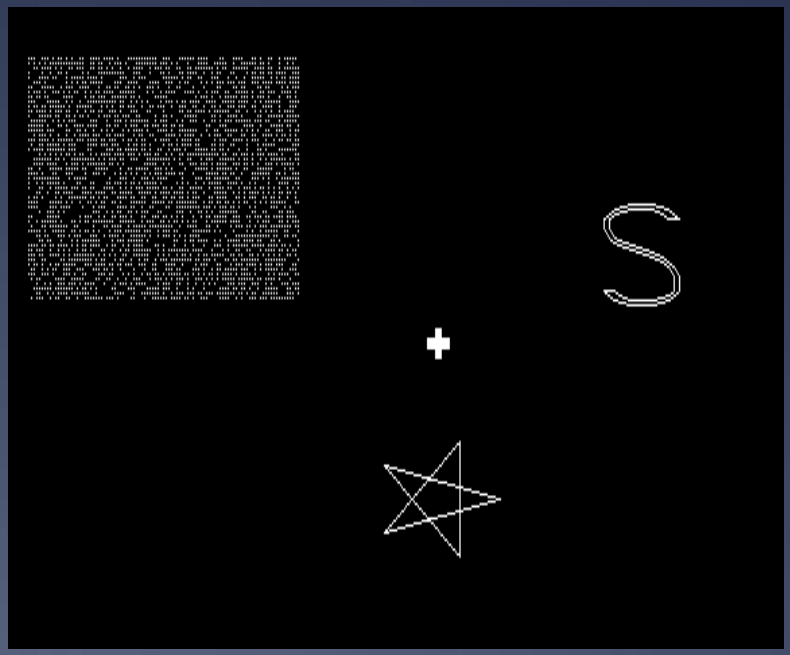
Methodological problem: basic associative learning may be sufficient

Contrast with: « Buying » hints for information (Washburn et al, 2006, Kornell et al, 2007)

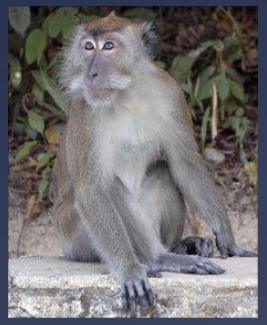
Example of an opt out paradigm

- Discrimination judgments in a choose-or-decline-to-respond (« opt out ») paradigm
- Will an animal choose to decline mostly for difficult stimuli ?
- perceptual task: Shield, Smith & Washburn, 1997 and memory tasks (Hampton, 2001)



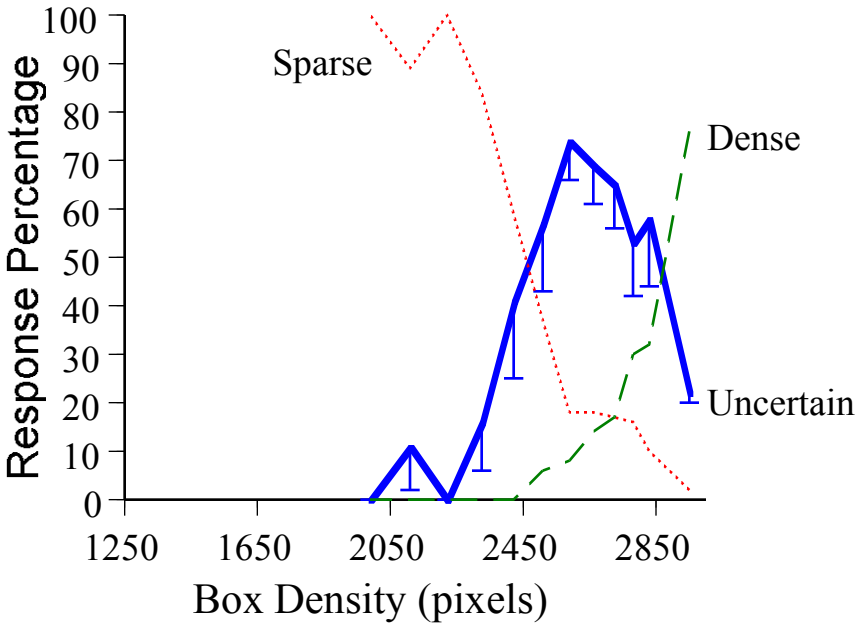

S

Smith and Hampton on declining in monkeys

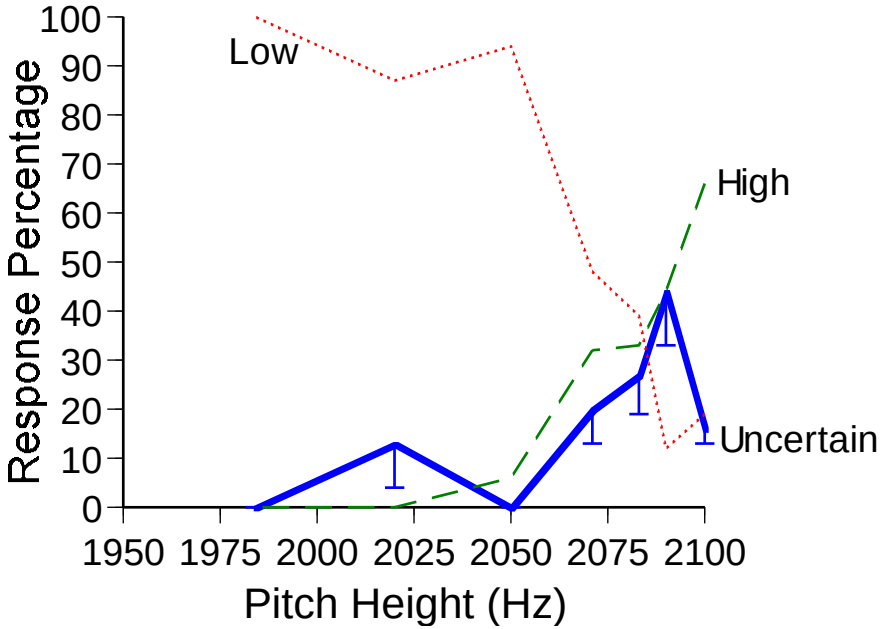


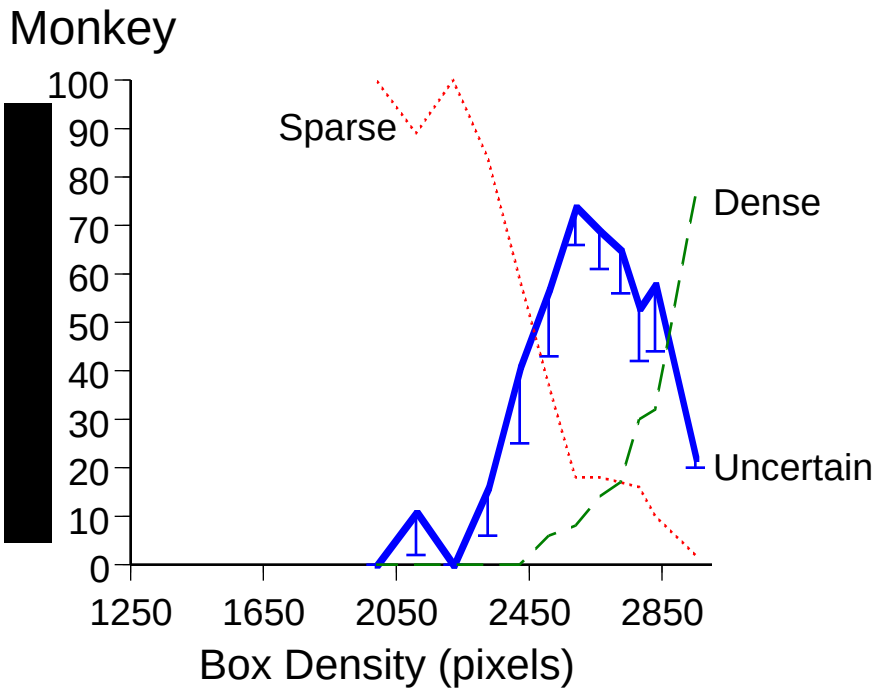
- Rhesus monkeys decline most the most difficult trials in visual discrimination tasks (Shield, Smith & Washburn, 1997) and in memory tasks (Hampton, 2001).
- They generalize their U- responses to new tasks. (Washburn, Smith & Shields, 2006)
- Bottle-nosed dolphins have similar results.

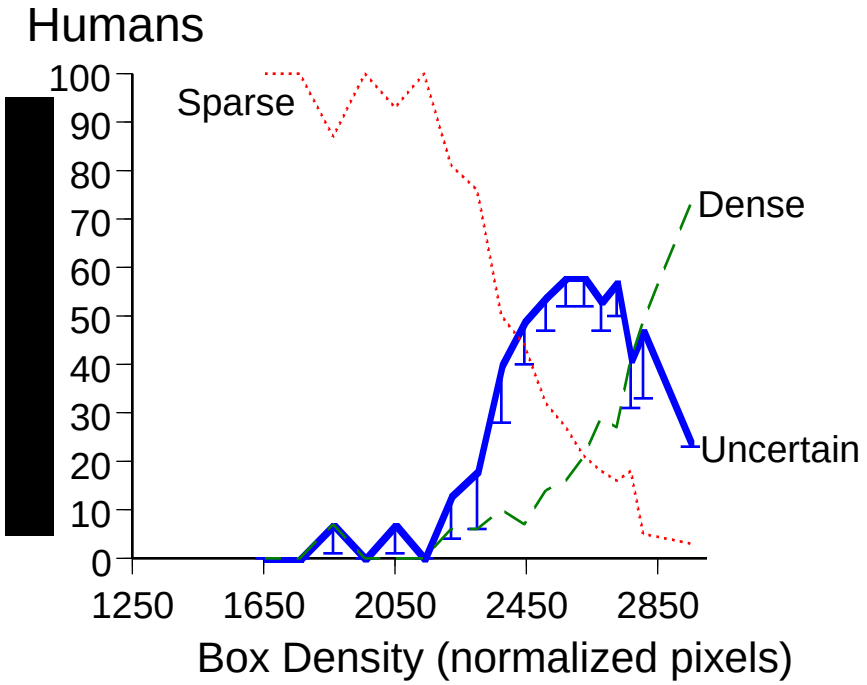
Macaque



Dolphin




Monkey
100
90
80
70
60
50
40
30
20
10
0
Sparse
Dense
Uncertain
1250
1650
2050
2450
2850
Box Density (pixels)


Humans
100
90
80
70
60
50
40
30
20
10
0
Sparse
Dense
Uncertain
1250
1650
2050
2450
2850
Box Density (normalized pixels)

Are MC judgments only based on perceptual cues ? Transfer tasks

* After a first set of declining trials, animals have to opt out, or to express their confidence in new, unrelated tasks (Kornell et al., 2007).

* Transfer shows that animals do not respond to features of the stimuli (their associative force with a given key) , but rather than to their own uncertainty. (Couchman et al., 2010)

Objection: reward rather than metacognition

* Is it reward, rather than an animal's judgments of confidence, that guides its decisions?

  (Carruthers, 2008, 2009, Crystal & Foote, 2009).

* To address this problem, animals were denied any access to reinforcement scheduling by providing blockwise, rather than trial-by-trial reinforcement.

  Macaques also use U-responses with blocked feedback (Beran, Smith, Redford & Washburn, 2006, Smith et al., 2006, Couchman et al., 2010)

Uncertainty responses by these animals are

- flexible (transfer to new tasks)
- used nimbly (trial 1 of new tasks)
- In the absence of direct reinforcement

3 – Differences in reliability of attribution and self-evaluation in human adults

If introspective forms of metacognition were based on mindreading, there should be a substantial similarity in epistemic evaluations

- concerning self or others,
- made on-line or off-line

This was tested by 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.

* Subjects are asked to provide such judgments either
  * after having performed the task or not
  * concerning their own or others' performances

A remarkable dissociation

Off-line evaluation

- 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

- 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)

Replicated in part by Loussouarn et al. for metaperception

Off-line evaluation

Mindreading

- participants rely on the naïve, incorrect theory that they can accurately and rapidly detect changes in a perceptual lay-out: "change blindness blindness"

(Levin et al., 2000)

- Plausible folk belief: A longer search time predicts a more accurate perceptual judgment (not tested)

On-line evaluation

Procedural metacognition

- participants judge correctly that a longer search time predicts poorer detection performance
- "perceptual effort heuristic", based on dynamic cues such as time spent and rate of accumulation of evidence.

(Loussouarn , Gabriel & Proust, *Consciousness and Cognition, 2011*)

The converse case: Schwarz (2004)

- How do you rate your memory?
- If invited to retrieve 12 childhood events, subjects rate their memory as less reliable than when having to retrieve six events → effort heuristics does not correlate with memorial ability
- A yoked participant attributes a higher memorial ability to subjects retrieving more events.

How are procedural metacognition and mindreading related in human adults ?

The metacognitive literature shows that, in humans, self-evaluations can occur either

- Based on the experience of the task
- Based on a conceptual, or analytic, representation of the task. The latter may rely on naïve theories about perception, memory, etc. and on mindreading-like explanations of cognitive processes.

Possible objection and response

- Subjects engaged in the task have access to introspective data that they do not have when merely predicting learning in others,: assymmetrical outcomes are compatible with an interpretive view.
- Response: if self-directed mindreading occurs, requiring as it does a conscious access to introspected evidence (Carruthers, 2009, 2011), why is it that subjects engaged in the task are unable to express the basis of their predictions ?

Metacognition & mindreading: “a complex message”

1- different processes mediate mindreading from those that mediate metacognition (theory vs experience)

2- metacognition can inform mindreading: participants can derive insight from observing their own mental processes, and can apply that insight in making predictions for others (although they are unaware of the heuristics involved)

3 - social interaction and mindreading may also affect metacognition.

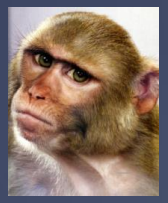
Koriat and Ackerman 2010

4 - Difference in Neural mechanisms associated with self-evaluation and self-attribution

The neural correlates of procedural metacognition in rhesus monkeys.

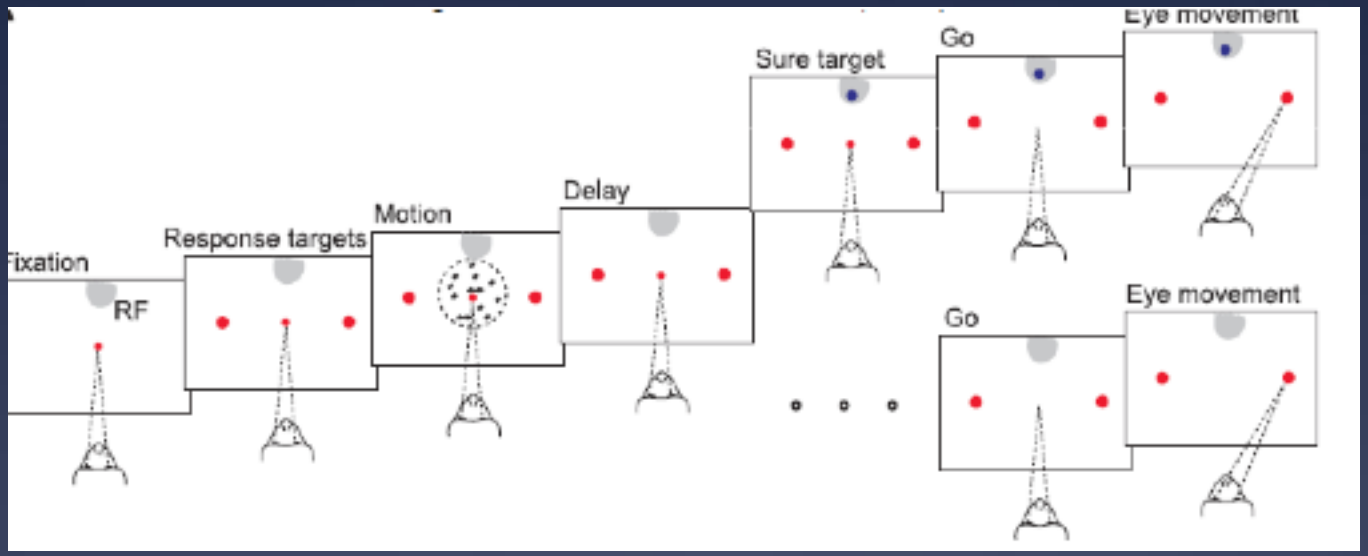
were studied in an opt-out task, where monkeys must

- discriminate whether a shortly presented stimulus is moving left or right.
- respond, after a delay, with an eye movement (Newsome paradigm)
- "Sure bet" option available in some trials



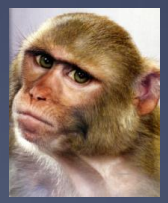
(Kiani & Shadlen, *Science*, 2009)

Opt-out paradigm in monkeys: neural correlates



Kiani & Shadlen, *Nature Neuroscience*, 2009

- 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.



Discussion: The double accumulator model

- Kiani & Shadlen, as well as other similar studies, can be modelled by the concept of adaptive accumulator modules.

Vickers and Lee (1998) and (2000)

- Evaluating one's uncertainty depends not on actual behavior, but only on the informational characteristics of brain activity, in particular the comparative rate of accumulation of evidence favoring one against competing decisions.

Neural correlates of

Metacognition

- The sensory areas (primary accumulators),
- the dorsolateral prefrontal cortex
- the ventro-medial prefontal cortex, in particular area 10 (control accumulators)
- The anterior cingulate cortex.
- the right medial prefrontal cortex: feelings-of-knowing judgments.
- prefrontal cortex: metacognitive visual awareness (Schnyer et al. (2004), Del Cul et al. (2009).

Mindreading

- the right temporal-parietal junction,
- the prefrontal antero-medial cortex,
- anterior temporal cortex.

(Perner & Aichorn, 2008)

Neural correlates of reward

- Ventral prefrontal cortex
- Ventral striatum – Basal ganglia

A dual process view of metacognition: epistemological significance

Different functions ?

- Metacognition specializing in evaluation of cognitive adequacy in own cognition
- Metarepresentation specializing in verbal report on self and other for communicational justificatory purposes.
- Can be performed jointly (redescription), but can also be dissociated :
  - Metacognition can be implicit
  - Metarepresentation can be shallow

Metacognition

- EssentiallyReflexive
- Engaged processing (simulation)
- Poorly recursive
- No decoupling
- Representational promiscuity
- No inferential promiscuity
- Predictive-evaluative function

Metarepresentation

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

Two analyses of “I know that p”

- in metacognitive terms: as a response to a procedural, self-addressed query such as: “do I have the answer whether p in my memory ?”

- In metarepresentational terms: as a statement involving reference to the concept of knowledge.

Two types of justification

- Gained on the basis of one' s epistemic experience of a task

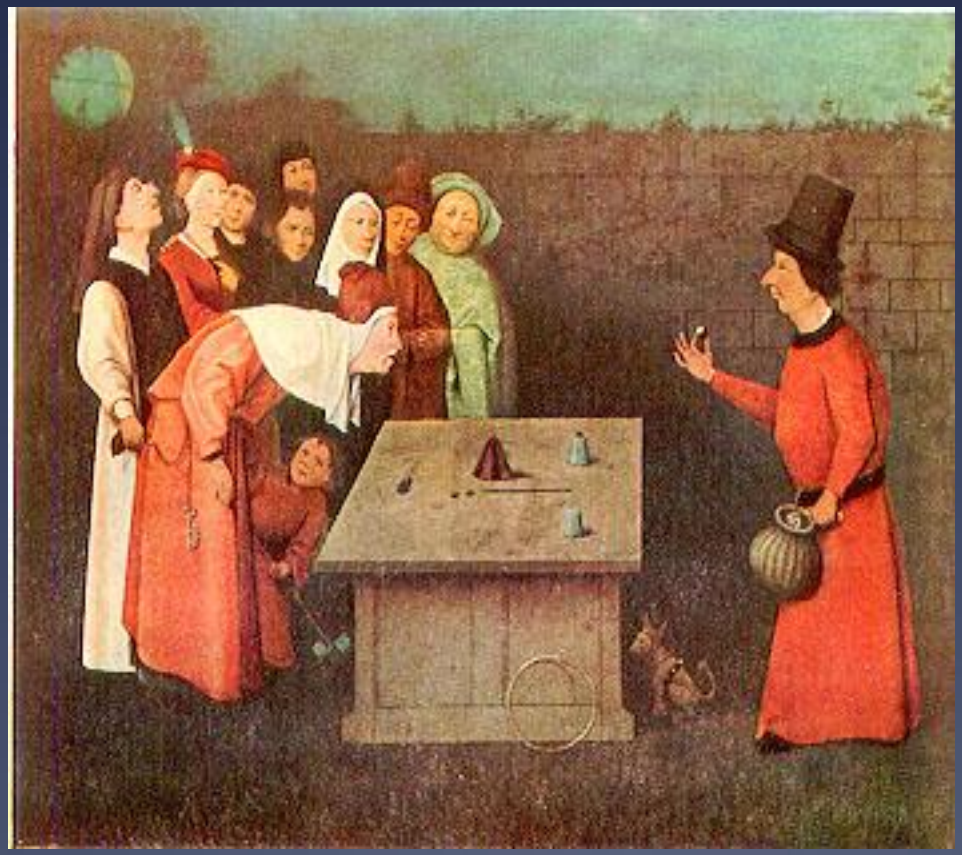
- Gained on the basis of one's epistemic beliefs about one's competence and success in the task.

Does this justify an internalist reading of metacognition?

- Procedural metacognition offers direct, privileged access to one's epistemic feelings, which predict success and motivate a decision to perform a cognitive action.

- But these feelings are themselves dependent upon prior experience in similar cognitive tasks:
  - If prior experience is too restricted, or misleading (biased feedback), confidence is miscalibrated
  - Subjects may categorize incorrectly a given task, and thereby have illusory feelings of ease of processing.

Metarepresenting metacognitive failures (Hieronymus Bosch, *The conjuror*)



The end

This ppt is on-line on:

http://joelleproust.hautetfort.com

Experience of fluency

Influences processing when

- Experience is seen as meaningful for task.
- task motivation is low
- task is not personally relevant.
- In elated mood conditions
- In divided attention, ie in low cognitive resources condition

Does not influence processing when

- A naive theory questions relevance of experience (discounting), eg. attributes it to environmental infuence
- task motivation is high
- task is personally relevant.
- In bad mood conditions
- In undivided attention, ie in high cognitive resources condition