Seminar on Metacognition
All Souls College, Oxford

Metacognition and mental agency

Joëlle Proust

Institut | Nicod
CNRS L'ECOLE DES HAUTES ETUDES EN SCIENCES SOCIALES ENS
FONDATION PIERRE-GILLES DE GENNES POUR LA RECHERCHE
l'innovation AUTREMENT
erc

Metacognition: a self-evaluative definition

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

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

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)

Control-based (or « data-driven » monitoring, Koriat, Ma'ayan & Nussinson, 2006

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

« Goal-driven monitoring », Koriat et al, 2006

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).

Three claims

1. Spontaneous metacognitive evaluations are a source of evidence about norm sensitivity in humans.
2. There are several norms to which agents are sensitive when monitoring their cognitive performances and thereby forming acceptances, i.e. epistemic decisions.
3. Epistemic control and monitoring operates before, and independently from, utility heuristics.

Outline

1. Metacognition, mental action and norm sensitivity
2. Complex mental actions
   A. Instrumental selection of epistemic norms
   B. Epistemic evaluation autonomous
   C. Strategic acceptance
3. Application to various puzzles

1 - Metacognition, mental action and norm sensitivity

Mental actions

are ways of controlling one's cognitive activity.

- Examples:
  - controlled memory (versus automatic memory)
  - Perceptual attention (vs passive registering)
  - Def: An acceptance is formed as a result of one's cognitive performance being monitored and controlled for correctness.

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 (and its associated acceptance).

Metacognition is part of mental agency

- When acting, one needs to proportionate one's goals to one's mental as well as bodily resources: self-probing needed in both cases.

- When acting, on needs to check whether the action has been correctly performed: post-evaluation needed in both cases.

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

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 in the correctness of my recall?)

The selected epistemic norm determines which mental action is to be performed

- Thinkers can aim to recall a list **accurately** (no false alarm allowed), or **exhaustively** (no omission allowed).
- They monitor their memory in a different way in each case.

Koriat & Goldsmith (1996), Reber & Schwartz (1999)

Epistemic norms determine which mental action is to be performed

- trying to remember **accurately** who was there at a meeting: correction requires no false positives, but tolerates omissions.
- Trying to remember **exhaustively** who was there at a meeting: correction tolerates false positives, but requires no omission.

→ two distinct cognitive actions, which respond to different norms.

Norms for metacognitive control and monitoring

- Accuracy (memory, reasoning)
- Comprehensiveness or exhaustivity (memory, reasoning)
- Coherence (fiction, demonstrative reasoning, epistemic vigilance)
- Consensus (negotiation, deference to authority)
- Relevance (conversation)
- Intelligibility or fluency (perceptual judgment, epistemic vigilance)

2 - Complex mental actions

A puzzle

- A mental action is selected on the basis of the ultimate goal of a world-directed action
- Therefore it is prima facie unclear whether a given mental action, with its associated monitoring and control, should respond to one or both of the goals involved:
  - Epistemic
  - Instrumental

A - A cognitive action is selected for Instrumental reasons

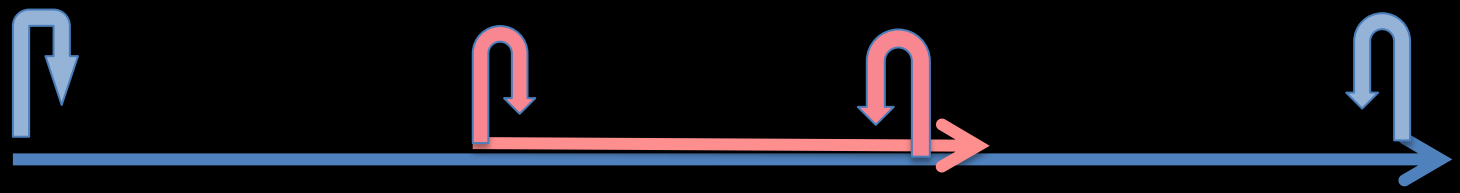
Example:

- The particular strategy of remembering (exhaustivity/accuracy) is selected for instrumental reasons.

A mental action is answerable to two types of norms

Epistemic action:

Epistemic norm(s)



Instrumental action: norm of utility

What are epistemic norms?

- The normative feature of epistemic norms derives from the structure of action being polarized (success vs failure: here, correctness vs incorrectness)

- The computational and semantic properties underlying the various norms derive from the various dimensions of information which can be profitably controlled and monitored in mental agency (through metacognition)

Norms for metacognitive control and monitoring

Epistemic norm

- Fluency
- Accuracy
- Comprehensiveness or exhaustivity
- Coherence
- Consensus
- Relevance
- Plausibility

Informational dimension

- Qualitative conservation of information
- Qualitative conservation of information
- Quantitative conservation of information
- Compatibility of contents
- Shared information
- Appropriateness in inferential power
- Predictaibility from an initial set of beliefs

What are instrumental norms?

- Utility, however, dictates that a given norm will be used to control cognitive activity given one's ultimate goal (e.g.: exhaustiveness)
- Context is determined by selecting a cognitive action as relevant to an ultimate goal. (now trying to remember exhaustively)

How, then, are epistemic and instrumental norms respectively regulating the various types of acceptance?

A solution

Although a simple mental action cannot be subject **both** to epistemic and non-epistemic norms, a complex action can

How do epistemic actions contribute to world-directed action?

- An epistemic action is usually embedded in an instrumental (world-directed) action. For example:
- In order to shop for food, I need to remember the items on the list (which I forgot to bring with me).

→ The epistemic norm guiding a mental action is selected on the basis of the ultimate goal of the world-directed action

B- Epistemic evaluation is autonomous from instrumental evaluation

A cognitive action is successful iff the corresponding epistemic norm is actually satisfied.

Selecting a norm for instrumental reasons does not influence sensitivity to correctness

- Agents' epistemic confidence in accepting$_n$ *P* (accepting *P* under norm *n*) is not influenced by the cost or benefit associated with being wrong or right: the epistemic content is not influenced by utility.

- Thus we don't need to endorse the view that an epistemic acceptance of *P* is yielding to utility considerations.

C - From epistemic to strategic control

Why strategic control ?

- A subject may or not decide to act on his/her epistemic acceptance, depending on the risk and benefit at stake.
  - Utility does not just influence the selection of certain epistemic norms of acceptance.
  - It also influences decision to act in a way that may depart greatly from the cognitive output of epistemic acceptance.

Subjective Expected Utility theory

- Parameters:
  - Value
  - Probability
  - Expected Utility = value x probability
- Each course of action (x_i) should be evaluated by myltiplying a subjective valuation of its consequences (reward) $u(x_i)$ by their probability of occurrence $P(x_i)$

$$\Sigma_i \, u(x_i) \, P(x_i)$$

WHY IS STRATEGIC ACCEPTANCE A SECOND, INDEPENDENT STEP?

Conceptual argument

- The existence of an autonomous level of epistemic acceptance enables agents to have a stable epistemic map that is independent from local and unstable instrumental considerations.

→ It is functionally adaptive to prevent the contents of epistemic evaluation from being affected by utility and risk.

Argument from metacognitive studies (Koriat and Goldsmith, 1996)

- In situations where agents are forced to conduct a cognitive task, strategic regulation is ruled out: agents merely express their epistemic acceptance.

- In contrast, when agents can freely consider how to plan their action, given its stakes, they can refrain from acting on the unique basis of their epistemic regulation and subsequent acceptance.

Argument from metacognitive studies

- A decision mechanism is used to compare the probability for their acceptance being correct and a preset response criterion probability, based on the implicit or explicit payoffs.
- Agents are allowed to strategically withhold or volunteer an answer according to their personal control policy (risk-aversive or risk-seeking), associated with the anticipated costs and benefits (Koriat and Goldsmith, 1996).

Argument from metacognitive studies

- Strategic acceptance can be impaired in patients with schizophrenia, while epistemic acceptance is not **(Koren et al. 2006)**
- This suggests that epistemic and strategic acceptances are cognitively distinct steps.

Strategic regulation < decision criterion

- Such varied factors as monetary incentives, instructions, and hypnosis have all been shown to affect the strictness or liberality of the criterion one uses to decide whether to volunteer or withhold potential responses in free-response testing, without having any effect on forced-response performance.

To prevent confusion between epistemic and instrumental norms

Primatologists David J. Smith and Michael Beran have prevented rhesus monkeys to have access to the reinforcement schedule of trials

→ a particular decision cannot be explained by the expected utility of that decision.

Smith et al. (2006), Couchman et al. (2012)

Context: determined by stakes in 2 ways

Acceptance is context dependent for two reasons:

1. Its norm (constituting this type of accepting) is strategically dependent on the instrumental context of a plan to act.
2. The decision to act on its content (what is finally accepted) is secondarily adjusted to the expected gain/cost of content being correct or incorrect.

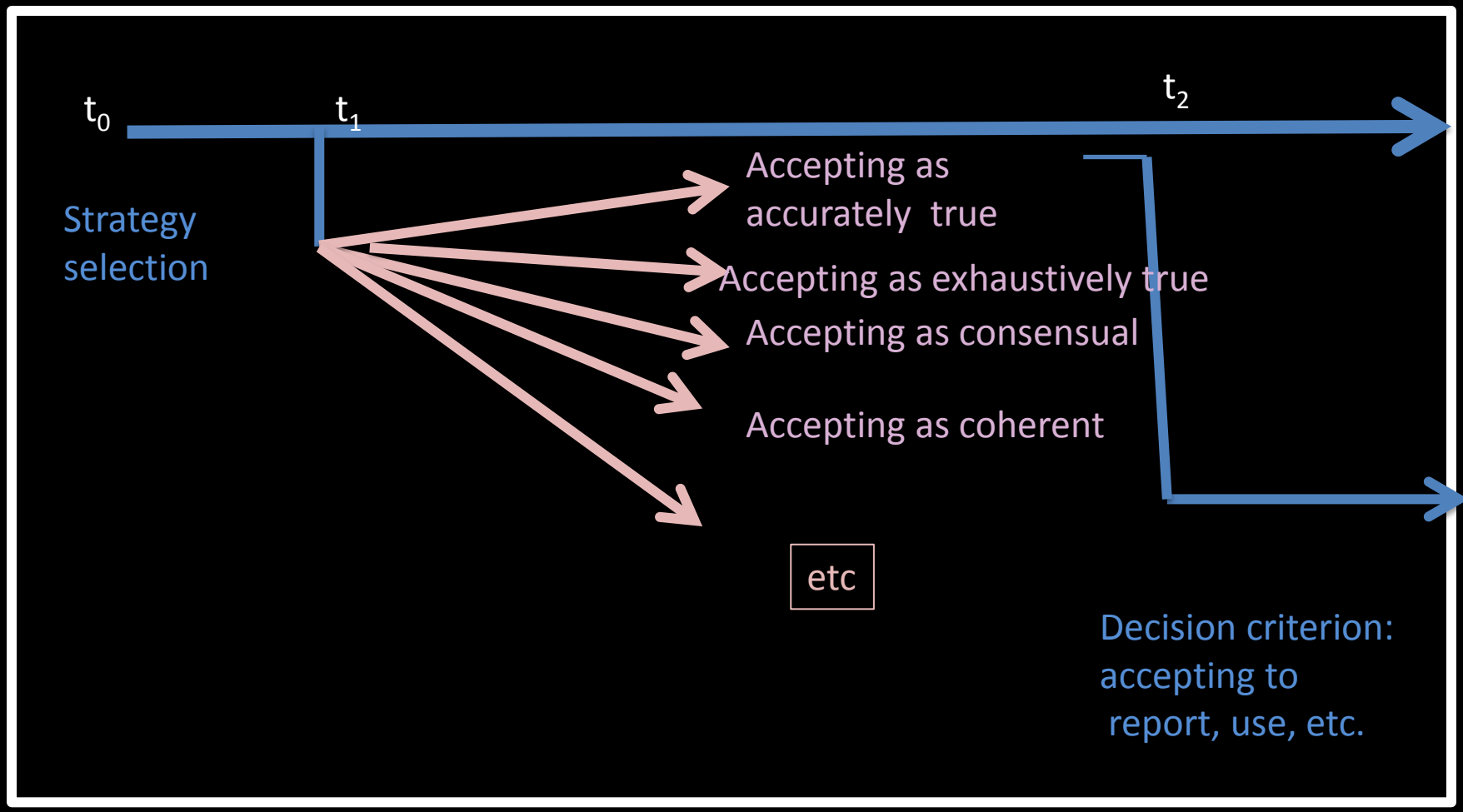
Acceptances thus need to have a layered structure

- Epistemic layer:
  - An epistemic evaluation is made about the subjective certainty associated with a given cognitive performance → $Acceptance_E$
- Instrumental layer
  - A revision of $Acceptance_E$ can be conducted to maximize utility: an evaluation is made of the importance of error and gain in a context against a level of subjective certainty → $Acceptance_I$ (modulated by a decision criterion to act or not on $Acceptance_E$)

Why the distinction of layers is a functional requirement

- The agent must be in a position to revise her acceptances both when the world is changing, and when the stakes are varying.
- If there was only a one-level acceptance, the agent could not tell apart what she ought to think from what she ought to do.
- Epistemic acceptance should always precede instrumental acceptance, and forms the only basis of decision in forced-choice responses.

The sequence of acceptances



4 – Application to various puzzles

Puzzles about acceptance (Stalnaker, 1984)

- The lottery paradox (Kyburg, 1961, p. 197) arises from considering a fair 1000 ticket lottery with one winning ticket. It is rational to accept that some ticket will win, while also accepting that ticket 1, 2 etc. will not win.

  I cannot aggregate my acceptances without being inconsistent.

Solution of lottery paradox

- An agent accepts_{at} (as accurate truth) that there is one winning ticket in the one thousand tickets actually sold
- She does not need to accept_{pl} (as plausible or likely) that the single ticket she wants to buy will be the winning one.
- There is no contradiction between the two acceptances, because they respond to different epistemic norms.

Solution of the preface puzzle

- The author's epistemic goal is one of offering an ideally comprehensive presentation of his subject matter:
- she can $accept_{ct}$ (comprehensive truth) that her book includes all the truths relevant to her subject, while $accepting_{pl}$ (accepting as plausible or likely) that one of her claims is false.
- Hence, a mental act of $acceptance_{ct}$ does not allow aggregation of truth, because its aim is exhaustive (include all the relevant truths) rather than accurate truth (include only truths).

Descriptive power: Rita Astuti's work on reasoning about death

When Vezos accept apparently contradictory accounts of what happens after death, their acceptance is context-dependent in the following senses:

- Epistemically: they select the context-adequate norm (under a norm of accuracy vs under a norm of consensus)
- Strategically: they select the context-adequate overt response.

Astuti & Harris (2008).

Epistemic context

The epistemic norm they are applying in reasoning about death depends on their current goal, e.g., social communication (consensus) vs. expression of knowledge (accuracy):

- Consensus when ritual is salient
- Accuracy when factual knowledge is salient

Strategic context

The instrumental norm (expected utility) they are applying, depends on the current cost-benefit ratio:

- May lead them not to express an acceptance even though they are confident that it is true or consensual.

How the proposal may contribute to the metacognition-as-mindreading debate

- Peter Carruthers, Josef Perner, and others, have claimed that cognitive control and monitoring requires a self-directed mindreading ability.

(Carruthers, 2011, Perner, 2012)

They are (only) partly right

- Sensitivity to norms such as accuracy, plausibility, or consensus, cannot develop in individuals with no mindreading ability.

- Indeed full-fledged mindreading, and the ability to form and understand metarepresentations, might be **constituted** by sensitivity to these norms (Proust, forthcoming)

What they are wrong about

- Sensitivity to fluency and coherence, however can be present in individuals with no mindreading ability. (Proust, 2012)
- Animals and humans can control their cognitive actions by subpersonal mechanisms extracting predictive dynamic cues in the neural vehicle
  - Kiani & Shadlen (2009)
  - Koriat & Ackerman (2010) « activity-dependent cues »,
  - Koriat's self-consistency model (2012)

Thank you for your attention !

This ppt presentation and associated articles are available at

http://www.joelleproust.org.fr

Appendices and bibliography

The varied philosophical uses of the terms « to accept » and « acceptance »

- judging true flat-out
- judging likely,
- Judging plausible,
- Judging reasonably defensible.
- Adopting a premising policy
- As regarding to be made true rather than being really true.

Partial bibliography

Astuti, R. & Harris, P. (2008)

Carruthers, P. (2011). *The Opacity of Mind*. Oxford: Oxford University Press.

Goldsmith, M. & Koriat, A. (2008).The strategic regulation of memory accuracy and informativeness. *The Psychology of Learning and Motivation,* 48, 1-60.

Kaplan, M. (1981) Rational acceptance. *Philosophical Studies*, 129-145.

Kiani , R. & Shadlen, M.N., (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science* 324, 759-764.

Koren et al. (2006) Real-World Cognitive—and Metacognitive—Dysfunction in Schizophrenia: A New Approach for Measuring (and Remediating) More ‘‘Right Stuff’’, *Schizophrenia Bulletin, 32, 2, 310-326.*

Koriat, A., Ma’ayan, H. & Nussinson, R. (2006). The intricate relationships between monitoring and control in Metacognition: lessons for the cause-and- effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General, 135, 1, 36-69.*

Partial bibliography

Proust, J. (2012). Mental acts as natural kinds. In A. Clark, J. Kiverstein& T. Vierkant (eds*), Decomposing the will*, Oxford University Press

Proust, J. (2012). Metacognition and mindreading: one or two functions? In Beran, M., Brandl, J., Perner J. & Proust, J. (eds): *The Foundations of Metacognition.* Oxford: Oxford University Press.

Proust, J. (2012). The norms of acceptance. In: B. Roeber and E. Sosa (eds.) Action Theory, *Philosophical Issues,* 22.

Proust, J. (forthcoming). *The philosophy of metacognition*. Oxford: Oxford University Press.

Reber, R. & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. Consciousness and Cognition, 8, 338-342.

Stalnaker, R. (1984*). Inquiry*. Cambridge: MIT Press.