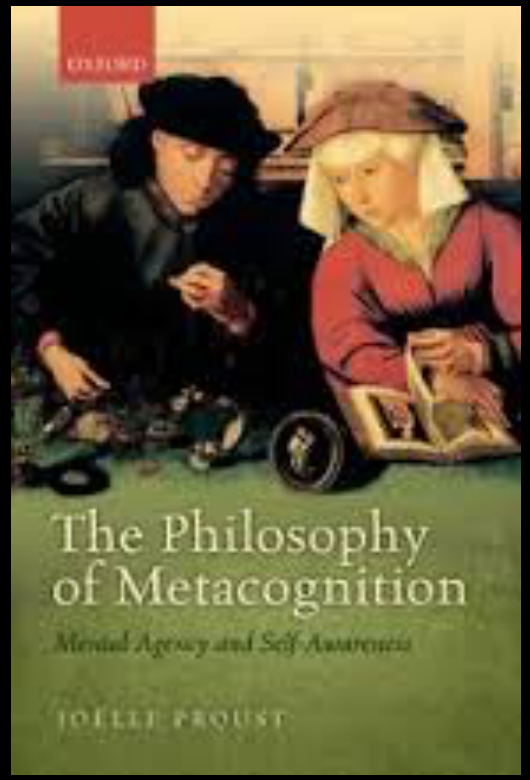
AIPS CONFERENCE

"Mechanistic explanation, Cumputability, and Complex Systems"

October 28, 2016



Metacognition and the Role of Mechanistic Explanations in the Philosophy of Mind

Joëlle Proust

http://joelleproust.org

Institut | Nicod

Our aim

- Analyse a test case of a philosophical controversy about the phylogenetic scope and nature of metacognition

- through the lenses of the mechanistic constraints in modeling informational processes.

Outline

1. Functionalism and the devolution of mechanistic explanations
2. Mechanistic explanations & the controversy metacognition/theory of mind: what is metacognition?
3. Using Marr's trichotomy to compare views:
   - Program level
   - Representation level
   - Mechanical level
4. Conclusion

1 - Functionalist philosophy of mind and the devolution of mechanistic explanations

David Marr's trichotomy of analysis of information processing systems

- Computational or program level: **what** does the system do (e.g.: what problems does it solve or overcome) and similarly, **why** does it do these things
- algorithmic/representational level: **how** does the system do what it does, specifically, **what representations** does it use and what **processes** does it employ to build and manipulate the representations
- implementational/physical level: how is the system **physically realised** (in the case of biological vision, what neural structures and neuronal activities implement the visual system)

Functionalism

- Functionalism in the philosophy of mind individuates mental states in terms of their causes and effects, described at the program level
- Machine state functionalism is the view that mental contents are computational states.

→ Program level (belief box/desire box)

→ Computational level: propositions in a "language of thought" (LOT)

Ned Block and Jerry Fodor (1972) defend the multiple realizability of mental states on physical substrates, and claim that any physicalist type-identity hypothesis will fail to be sufficiently abstract.

- This line of thinking discouraged the search for "reductive" mechanistic explanations.

Objection to functionalism by neuroscientists Changeux & Dehaene (1989): Marr's trichotomy works within layers

A single neuron is already performing a computational task (the program level); it is following an algorithmic process, and does so according to specific physical properties (molecular properties of the synapse and of the membrane).

A second anatomical layer encompasses "circuits", i.e. neuronal assemblies of thousands of cells organized in well-defined structures, i.e. presenting task-dependent synchronous firings.

A third layer is constituted by the "metacircuits", i.e. relations of neuronal assemblies. Finally the traditional mental faculties are taken to roughly correspond to various of these metacircuits.

Objection to functionalism by neuroscientists Changeux & Dehaene (1989)

→ No causal-explanatory autonomy of any one task-level, but rather a relation of “co-dependence” among levels.

- The constraints of the synapse and the membrane determine, in part, which computations can be performed, as well as which kind of goal they can serve.

- Reciprocally, serving a goal modulates both the computational and the physical levels, and helps stabilize the physical properties of the cell.

Consequences of a functionalist viewpoint on exploring the mechanistic aspects of mental functions

- Once it is accepted that mental states are multirealizable,

- the requirements of a mechanistic explanation become much less relevant than higher level considerations (at the program level or at the computation level).

Requirements for a mechanistic explanation: Kaplan & Craver (2011)

(**3M**) In successful explanatory models in cognitive and systems neuroscience,

(*a*) the variables in the model **correspond to** components, activities, properties, and organizational features of the target mechanism that produces, maintains, or underlies the phenomenon, and

(*b*) the (perhaps mathematical) dependencies posited among these variables in the model correspond to the (perhaps quantifiable) causal relations among the components of the target mechanism.

Most philosophical explanations of the mind do not fulfill these requirements

For example, Jerry Fodor's Language of Thought hypothesis* merely presupposes that a physical syntax should account for the ability of beliefs and desires to interact causally.

* The language of thought hypothesis (LOTH) is the hypothesis that mental representation has a linguistic structure, or in other words, that thought takes place within a mental language

Most philosophical models of the mind do not fulfill these requirements

- It is proposed that the syntax of the system of representations can organize causal relations between thought contents that parallel their semantic relations.
- EG: logical rules of inference such as modus ponens are defined over the syntax of the representations.

No evidence is provided, however, that the target causal mechanism has the syntactic structure of LOT, nor that syntactic derivations are mapped in the causal mechanisms of thought.

2. Mechanistic explanations & the controversy metacognition/theory of mind

Why is it worth analysing the controversy metacognition/theory of mind as a test case for the relevance of mechanistic explanations in the cognitive philosophy of mind?

It enables to show that most functionalist philosophical theories

- Do not consider the mechanistic constraints on the representations apt to perform given cognitive tasks.
- ignore the dynamical constraints on computation
- Have a restrictive notion of mental content, leaving no constitutive role to emotions and embodiment in cognition.

Metacognition

Metacognition refers to the set of processes through which agents contextually control and monitor their first-order cognitive activity (such as perceiving, remembering, learning, or problem solving) by assessing its feasibility or likely success.

Central examples of metacognition

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

Predictive monitoring: 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)

Retrospective monitoring: 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

Noetic Feelings

Predictive

- cognitive effortfulness
- Familiarity
- knowing
- Tip of the tongue

Retrodictive

- Feeling of being right
- Feeling uncertain about one's own performance

Controversy: What does the term “metacognition” refer to?

- In cognitive science, “metacognition” refers to the capacity of evaluating the feasibility or completion of a given cognitive goal (such as learning a maze, or discriminating a signal) and controlling cognitive performance accordingly.

→ **« Self-evaluative » view (procedural metacognition)**

- Mindreading specialists take metacognition to refer to first-person metarepresentation of one's own mental states (Perner, 1991, Carruthers 2009).
- → **« Self-attributive » view (belief-based metacognition)**
-

3. Using Marr's trichotomy

Program level: does metacognition require self-attribution?

- From the self-attrbutive viewpoint, metacognition requires representing one's own mental states in propositional terms, i.e. "as mental states"
- Hence metacognition requires forming metarepresentations such as:
  - I **believe (or I seem to remember)** with degree of certainty i
  - That [there is beer in the fridge]

- Granting that only humans have the ability to attribute beliefs to themselves, then only humans are able to have metacognition.

3 arguments against the Self-attributive view of metacognition

1. Program level: non mindreaders have metacognition

   → self-attribution of attitudes not required by self-evaluation

2. Computational level: representations used to perform evaluations cannot have a propositional structure

3. "Implementation level": Graded valence and intensity are expressed in the neural activity through dynamic signatures and in subjective feelings.

Program level: non mindreaders *have* metacognition

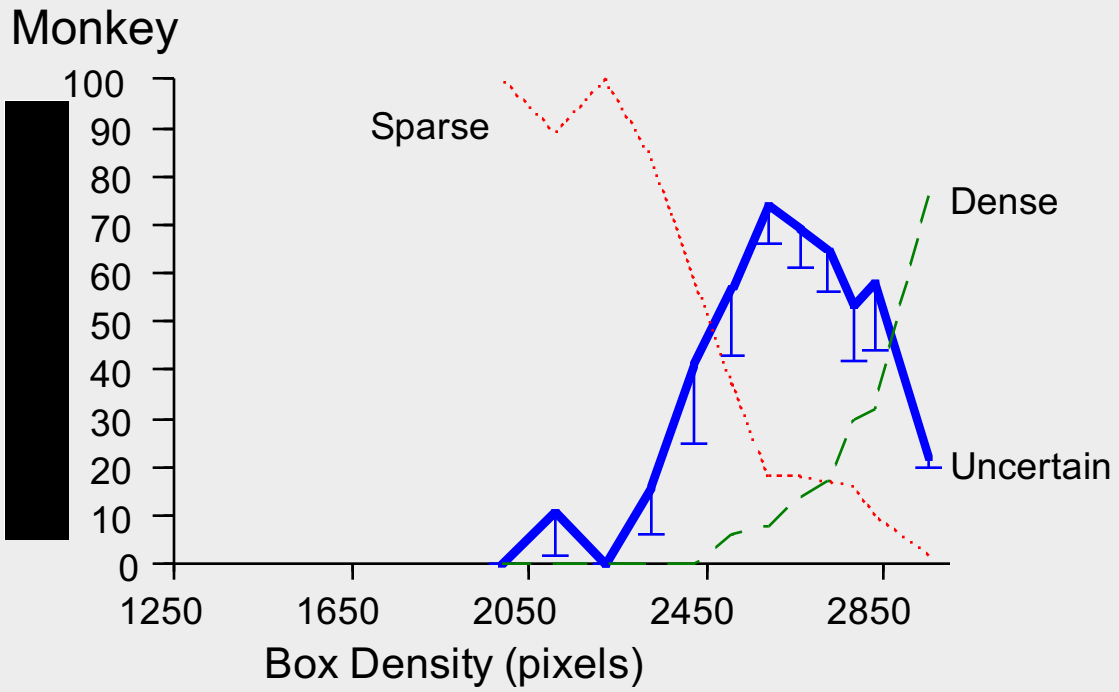
Experimental evidence for non-human metacognition

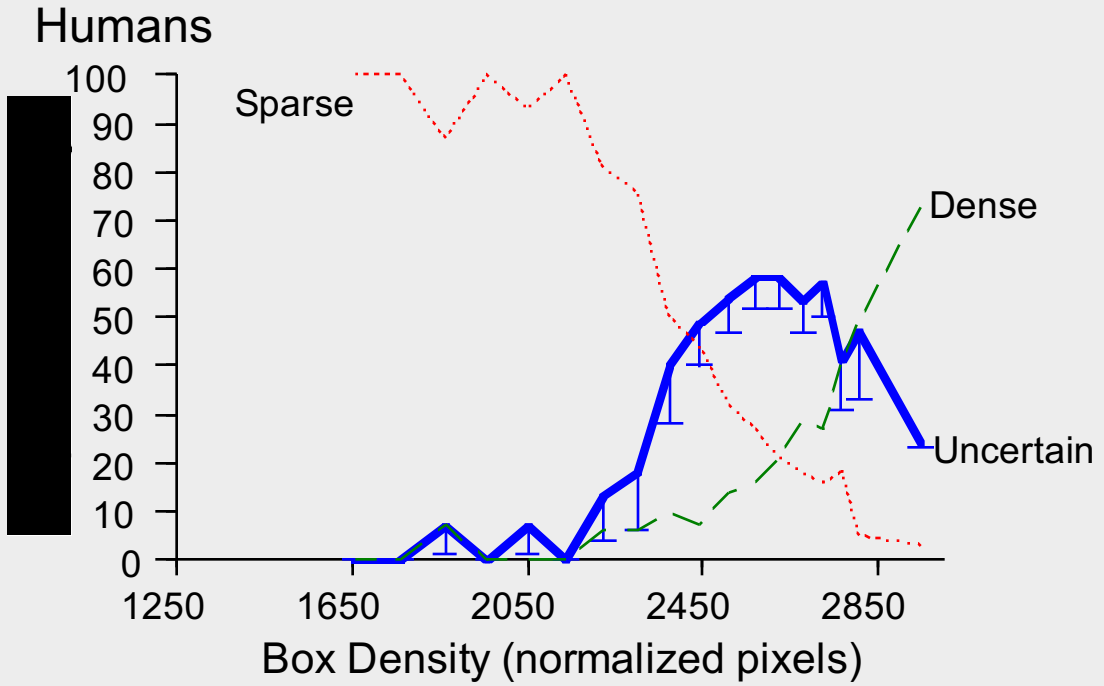
3 main experimental paradigms (behavior/brain)

1. Seek information before acting? (Call 2010) or obtain it from a helper at a cost? (Hampton, 2009)
2. Choose/decline to perform a task of various difficulty?
   - Smith et al 2008: visual discrimination
   - Kepecs et al. 2008, 2012): olfactory discrimination
   - Hampton 2001: memory retrieval of paired items
3. Wager on previous cognitive decision? (Kornell et al. 2007).

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
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)

Metacognition in Phylogeny:

Yes

- Pigeons U-R opt out (Adams & Santi 2011)
- Rats: Foote & Crystal (2007); Kepecs et al (2008) U-R
- Capuchin monkeys: U-R (Fujita 2009)
- Rhesus macaques (SI & U-R) (Smith et al, Kornell, Hampton))
- Bottle-nosed dolphins U-R (Smith)
- Chimps and orangutans (SI) and UR (Suda-King 2008)

No

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

If rhesus monkeys and rodents have metacognition

Then metacognition does not require mindreading

(even though mindreading may enrich metacognition when available or be used to communicate about metacognition)

Computational level: representations of gradient for evaluation *cannot have* a propositional structure

Propositions versus graded evaluations

1. Propositions are detached and objective (Strawson, 1959)
2. **Propositions express facts**
3. Propositional structure permits combinatorial thinking
4. Propositions involve an inferential network.
5. Propositions are based on concepts
6. Propositions are involved in strategic reasoning

1. Evaluations are relational and subjective
2. **Evaluations express gradient**
3. Evaluative structure permits recalibration and fusion of evaluations.
4. Evaluations involve an associative network
5. Evaluations mainly rely on nonconceptual information
6. Evaluations are involved in reactive action guidance

Proposal: procedural metacognition is based on affordance-sensings, a specific type of evaluation for informational reliability

- Evaluations have a non propositional semantics, but they do have semantic structure.
- They have no truth-conditions, **but still have conditions of felicity. (Proust, 2015, 2016)**
- They have specific dynamics as a function of the time constraints inherent to the task which they control (Proust, 2014)
- They are accessed through **pattern matching**

The semantic structure of evaluative attitudes (affordance sensings)

- An AS is affectively indexing an *occurrent* (relational) opportunity, rather than an individual event or object.
- Affordance$_a$ [Place=here],[Time= Now/soon],
- [Valence$_a$], (on a scale 0 to 1)
- [Intensity$_a$ (on a scale 0 to 1)],
- [motivation of degree$_d$ to act according to action program$_a$].

→All the constituents are associatively related to perceptual cues in the affordance sensing

→ A subset may activate the full representation and thus predict an opportunity

Strawson (1959), Proust (2013, 2015, 2016)

Noetic feelings are Affordance sensings

- Express a relation, not a state of affairs
  - Indicate a subjectively relevant condition and motivate an action
  - Are evaluative and graded

→ Nonpropositional

→ Do not conceptualize, but categorize "informational affordances" by mere associative pattern matching

Computational properties

- The various cues associated in an affordance sensing, taken together, constitute **heuristic devices** (each cue has its own weight in the overall cognitive affordance sensing).
- **Gradient** is expressed on a continuous scale
- Several evaluations (from different affordances) **can be fused** and impact a single decision making.
- The relevant computations are **activity-dependent**, i.e they use feedback from the present task, whether bodily, environmental, or neural.

"Implementation level": lesional studies vs dynamic signatures

How are the requirements for a mechanistic explanation met? : Kaplan & Craver (2011)

(*a*) variables in the model → components of the target mechanism

(*b*) Variable dependencies → causal relations among the components of the target mechanism?

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 metacognitio 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

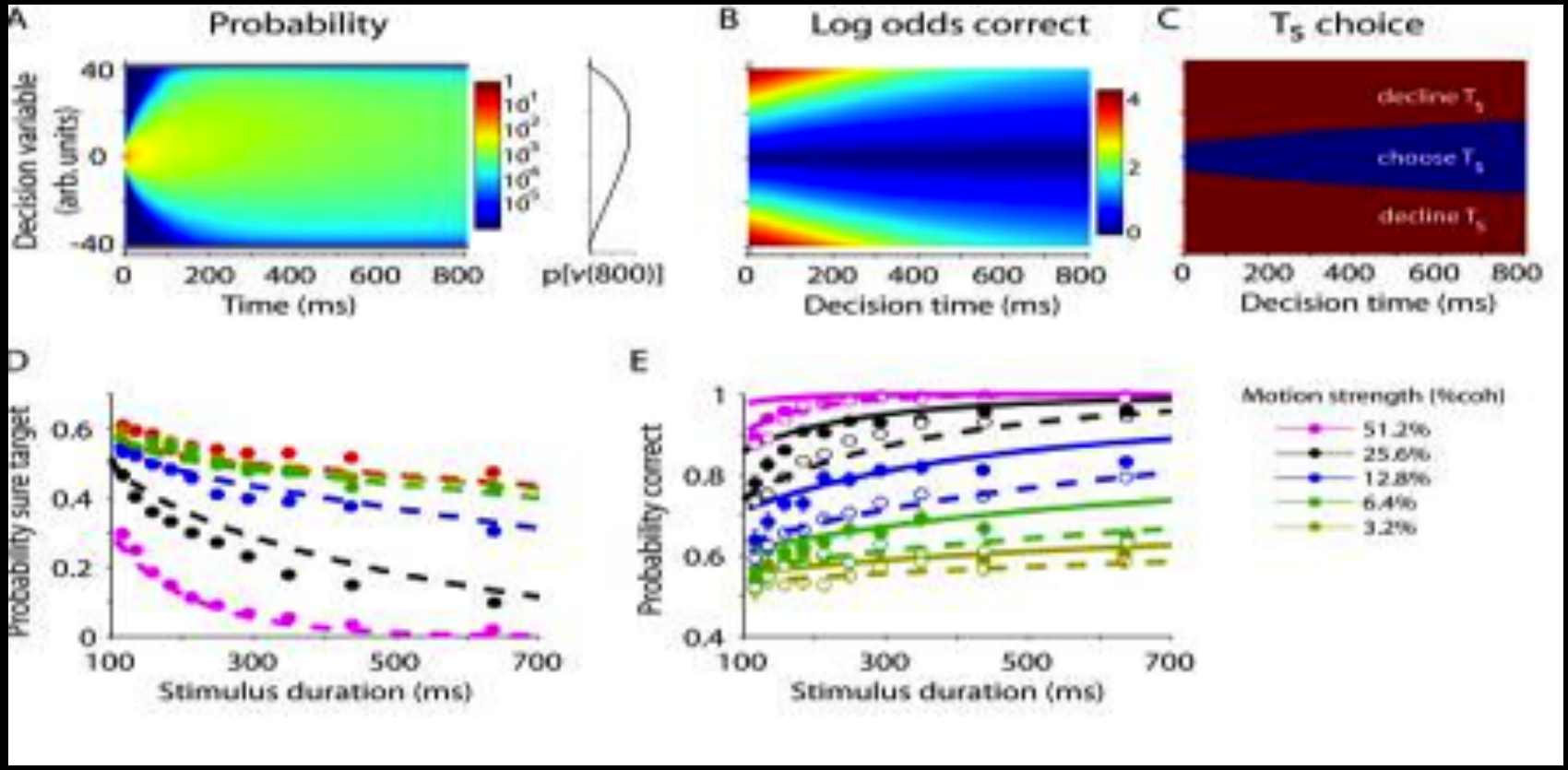
(Kiani & Shadlen, *Science,* 2009)

Kiani & Shadlen, *Nature Neuroscienc* 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



How does a cognitive affordance sensing of certainty work?

Animals and humans extract predictive information from the "neural signature" of the activity elicited by a cognitive task

- Processing onset,
- intensity (amplitude of activation)
- coherence of cognitive activity over time
- Latency to reach threshold (fluency)

These cues are part of a heuristic predicting likely cognitive success of a given epistemic decision.

(Kiani & Shadlen, 2009,Kepecs & Mainen, 2012).

How does mindreading work?

No consensus about it.

- Innate TOMM module stepping in (some say by 7 seven months, others by 4 and half)?
- Theorizing on perceived regularities with the proper conceptual understanding of representation?
- Adult fMRI data (TPJ) + lesional studies & autistic deficit do not allow validation of one specific account.

Conclusion
Metacognition and mechanistic explanation

A recapitulation of mechanistic explanations

	Self-attributive view of metacognition	**Self-evaluative view of metacognition**	
Model variables → components in the mechanism	Metarepresentations →executive abilities + ??	Reactive intensity and valence of a-sensings → Dynamic signature in activity-dependent neural activations	
Model dependencies → Causal relations	FBT → unsolved in frontal patients and in autistic disorder	Uncertainty responses→ Dynamic signatures relative to calibrated threshold	
Global assessment	No computational method available, because heuristics are still unknown	Computational methods available to predict observed outcomes from neural signatures.	

Metacognition and mindreading

are different abilities

- Metacognition guides cognitive action
- Mindreading explains in mental terms why people behave as they do

Different functions

- Metacognition specializing in evaluation of cognitive adequacy in own cognition
- Mindreading specializing in verbal report on self and other for communicational purposes.

→ Mindreading can resdescribe, and sometimes enrich or disrupt metacognition through background beliefs and inferences

Procedural metacognition

- evaluation of cognitive adequacy in own cognition
- Requires task engagement
- Based on activity-dependent cues
- Heuristics involved are implicit
- Evaluations are only made conscious through feelings

Mindreading

- verbal report on self and other for communicational purposes.
- Detached attribution
- Based on background beliefs about the mind
- Beliefs involved are explicit.
- Metarepresentations are verbally conveyed

Thanks for your attention!

Papers and presentations downloadable at:

http://joelleproust.org