

Abstract word count: 48
Main text word count: 975
References: 230
Entire text (total): 1301

Title: Overlooking metacognitive experience

Joëlle Proust

Institut Jean-Nicod, EHESS & ENS, Department of Cognitive Studies, Ecole Normale Supérieure, 29 rue d'Ulm, 75005 Paris, France.

Ph: + 33 1 44 32 26 68

Email address: jproust@ehess.fr

<http://joelleproust.hautetfort.com>

Abstract

Peter Carruthers correctly claims that metacognition in humans may involve self-directed interpretations ie. may use the conceptual interpretative resources of mindreading. He fails to show, however, that metacognition cannot rely exclusively on subjective experience. Focussing on self-directed mindreading can only bypass evolutionary considerations and obscure important functional differences.

Peter Carruthers' main goal is to show that metacognition is a form of self-directed interpretation, akin to other-directed mindreading. Introspection, he claims, defined as "any reliable method for forming beliefs about one's own mental states that is not self-interpretative and that differs in kind from the ways in which we form beliefs about the mental states of others", is not needed to have access to one's own mental attitudes. One can agree with the author that metacognition in humans *may* involve self-directed interpretations ie. *may* use the conceptual interpretative resources of mindreading, without accepting the stronger claim that metacognition can *never* be based on "introspection".

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). In Peter Carruthers' use, however, metacognition *refers to* first-person metarepresentation of one's own mental states; as a result, the theoretical possibility that metacognition might operate in a *different* representational format cannot be raised. (Proust, in print_a). Revising the meaning of a functional term such as "metacognition" is a bold strategy. It generally seems more adequate to leave it an open empirical matter whether a capacity of type A (reading one's own mind) or B (evaluating one's cognitive dispositions) is engaged in a particular task. A revision is deemed necessary, according to Peter Carruthers, because B-capacities in fact always involve

self-directed mindreading; therefore apparent contrary cases (self-evaluation in non-mindreading animals) either i) are simply instances of first-order types of learning, and/or ii) are capacities 'too weak to be of any interest' (5.1 & 9).

Two methodological problems, however, hamper the discussion so conceived. First, it is quite plausible that, in *human* forms of metacognition (as instantiated in speech production, metamemory, etc.) judgments of self-attribution *redescribe* elements of metacognitive experience. Metacognitive feelings might, on this view, represent subjective uncertainty and guide noetic decision-making, without needing to involve a conceptual interpretative process. What needs to be discussed, in order to establish the superiority of model 4, is whether or not subjects can rely on dedicated feelings alone to monitor their on-going cognitive activity.

A second, related problem is that Carruthers' discussion conflates two domains of self-control, namely the control of one's physical actions through perceptual feedback and the control of one's mental actions through metacognitive feedback (see §§5.1 & 9). Meta-action however is only functionally similar to metacognition when a metarepresentational reading is imposed on both, in spite of their different evolutionary profiles (Metcalf, 2008, Proust, in print_b). If extracting, from a given task context, an evaluation of the mental resources available to complete the task were just another case of first-order action control, then one might agree that B-metacognition is nothing other than executive function. But metacognitive and executive abilities can be dissociated in schizophrenia (Koren et al. 2006). Mental action control is thus distinct both from executive memory as usually understood and from physical action control.

These methodological problems strongly bias the discussion against models 1 and 3. Here are three examples.

i) "Our metacognitive interventions don't require introspection; they have no direct impact on cognitive processing" (section 5.1.)

From a B-sense viewpoint, prediction and evaluation of one's mental states and events presuppose appreciating one's subjective uncertainty regarding correction, adequacy, etc. of first-order decisions or judgments; this evaluation does not require that the target states are *represented qua mental*. For example, a child chooses to perform one memorization task rather than another by relying not on *what she knows about memory*, as the author claims, but on the *feeling* she has that one task is easier than another (Koriat, 2000, Proust, 2007). The impact on decision is quite direct, and independent of mindreading.

ii) A combination of first-order attitudes is sufficient to explain how animals select the uncertainty key in Smith et al.'s metaperceptual paradigm (section 5.2.)

If this is correct, how can monkeys rationally decide to opt out when no reinforcement of the uncertainty key is offered, and when, in addition, novel test stimuli are used? Why should there be transfer of the degree of belief associated with first-order items to novel tasks where these items are no longer included? A second rule must apply, as Carruthers (2008) himself admits: having conflicting impulses to act or not to act on a given stimulus, the subject becomes uncertain of its ability, say, to categorize. So the decision to act depends, after all, on subjective – not objective – features. Can these subjective features influence behavior *only* through their being metarepresented? This is the crucial question that fails to be raised.

iii) "Evidence suggests that if mindreading is damaged, then so also will metacognition" (section 10).

Clinical research on autism and on schizophrenia suggests rather a dissociation of metacognitive and mindreading skills as predicted by model 1 (cf. Helwig & Principe, 1999, Koren et al., 2004, Bacon et al. 2001). However, its relevance for the present discussion is

downplayed as a smart behaviorist effect; introspection in patients with autism is rejected "because it is not metacognitive in the right sort of way". Negative results in meta-action from patients with autism are presented as evidence for impaired metacognition. Such appraisals implicitly appeal to the preferred metarepresentational interpretation of metacognition under discussion. Similarly, rejecting the relevance of metacognitive capacities which are 'too weak to be of any interest' presupposes recognizing the superiority of model 4 over models 1 and 3.

A fair examination of the contribution of "introspection" in metacognition in models 1 and 3 would require studying the respective roles of control and monitoring in non-human and human epistemic decisions, in experimental tasks engaging metaperception, metamemory and planning (Koriat et al. 2006, Proust, in print_b). Focussing on self-directed mindreading can only bypass evolutionary considerations and obscure important functional differences.

Acknowledgments

I am grateful to Dick Carter, Chris Viger and Anna Loussouarn for their remarks and linguistic help. The preparation of this comment was supported by the European Science Foundation EUROCORES Programme CNCC, with funds from CNRS and the EC Sixth Framework Programme under Contract no. ERAS-CT-2003-980409.

References

- Bacon E., Izaute M. et Danion J.M. 2007. Preserved memory monitoring but impaired memory control during episodic encoding in patients with schizophrenia. *Journal of the International Neuropsychological Society*, 13 : 219-227.
- Carruthers, P. (2008). Meta-cognition in Animals: a Skeptical Look, *Mind and Language*, 23, 1, 58-89.
- Helwig C.C. , Prencipe A., (1999), Metamemory in Children with Autism, *Child Development*, Volume 70, 1, January 1999 , pp. 107-131.
- Koren, D. Seidman, L.J., Goldsmith, M & Harvey, Ph. D. (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. (2000). The Feeling of Knowing: some metatheoretical Implications for Consciousness and Control. *Consciousness and Cognition*, 9, 149-171.
- 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.
- Metcalf, J., Greene, M.J., (2007) Metacognition of Agency, - *Journal of Experimental Psychology General*, 2007
- Proust, J. (2007). Metacognition and metarepresentation : is a self-directed theory of mind a precondition for metacognition ? *Synthese*, 2007, 2, 271-295.
- Proust, J. (in print_a) The representational basis of brute metacognition: a proposal, in R. Lurz (ed.), *Animal Cognition*, New York, Cambridge: Cambridge University Press.
- Proust, J. (in print_b), Epistemic Agency and Metacognition: an externalist view, *Proceedings of the Aristotelian Society*, vol CVIII, part 3, 241-268.