

Smith, J.D., Hields, W.E. and Washburn, D.A.

The Comparative Psychology of Uncertainty Monitoring and Metacognition

Does metacognition necessarily involve metarepresentation ?

Joëlle Proust

Institut Jean-Nicod (CNRS, (Ecole des Hautes Etudes en Sciences Sociales, Ecole Normale Supérieure), 1bis avenue de Lowendal- 75007 Paris

Max-Planck Institut Für Psychologische Forschung, München

tel : in France : 33 1 53 59 32 87 and 33 3 44 75 38 13.

in Germany : 49-89-38602-245

In Behavior & Brain Sciences, 26, 03, 2003, p. 352

jproust@ehess.fr

<http://joelle.proust.free.fr>

Abstract

Against the view that metacognition is a capacity that *parallels* theory of mind, it is argued that metacognition need not involve metarepresentation, nor semantic forms of reflexivity, but only process-reflexivity, through which a task-specific system monitors its own internal feedback by using quantitative cues. Metacognitive activities, however, may be redescribed in metarepresentational, mentalistic terms in species endowed with a theory of mind.

An important conceptual issue raised by the target article consists in the sense of « self-knowledge » engaged in confidence judgments that monkeys and dolphins seem to be able to form, in contrast with other species such as rats and pigeons. The authors tend to consider that the same notion of « self-reflexivity » applies in the realms of higher mental-states attribution and of metacognitive monitoring of the system's epistemic states. While they accept the view that cognitive self-awareness may be different from self-recognition, they suggest that metacognition is a capacity that *parallels* theory of mind : the latter asks « whether animals know and monitor the other's mental states and states of knowing », the former « whether animals know and monitor their own mental states and states of knowing » (Section 3). This parallel may be misleading, however, in important ways. Given monkeys' lack of theory of mind and absence of self-recognition (Cheney and Seyfarth, 1990, Anderson & Gallup, 1997), it is highly implausible that they have any mentalistic understanding of their ability to evaluate their own epistemic dispositions. But maybe Smith et al. rather suggest that metacognitive abilities constitute a precursor for the mentalistic abilities as found in humans. In this case, however, a clear distinction between mentalizing and metacognitive capacities is still needed.

Whereas by definition an animal endowed with a « theory of mind » capacity is able to

monitor and predict the behavior of others in a mentalistic way, that is, by attributing mental states to others, rather than on the basis of behavioral cues, metacognition does not seem to require any mentalistic attribution, still less so « to oneself ». Such a capacity presupposes that a control system i) has access to information concerning its present epistemic states (information that « self-monitoring » provides), and ii) uses it to select and complete a particular course of action. As the authors convincingly show in the specific case of SDT ideal strategies, such a hierarchical organization optimizes the benefit/cost ratio by applying hard-wired heuristics to the endogenous feedback. It clearly is a *procedural* form of metacognition, a « know-how to decide », that is not based on mental concepts and does not *need* to be made explicit (even if one grants Shiffrin and Schneider, 1977, that consciousness may be favorable to control, the proof of its being necessary is not made yet : see Reder & Schunn, 1996, Bargh, 1997, Spehn and Reder, 2000).

On the background of the uncontroversial control/monitoring model, two claims made by Smith et al. deserve discussion: that metacognition is metarepresentational, and that it is « about the self ».

A) Any control system involves a form of reflexivity at the task-level : there must be, as the authors write, a connection between the judgment of certainty and « the primary discriminatory process » in which it originates. (14.4). But they add the following comment on the relation between the two « .. The uncertain response (..) *is about* the status of the primary discriminatory process and about its probable failure. It stands structurally *outside* the primary discrimination and *intrinsically meta-* to it. » (14.4). This observation, however, conflates « being about » and « being meta » : « being about » involves mental reference to an object, an event, or a property. « Being meta » just involves hierarchical control between processes. The latter could qualify as metarepresentational, if control processes modelled not only the current epistemic states of the system, but also the attitudinal contents of the latter. But why should the control system need do this ? It is much more economical to have a mechanism that simply correlates the feasibility (probable success) of a task with preselected types of cues (like the quantity or intensity of the feedback), rather than one relying on the semantic processing of the first-order content of its epistemic states. Evidence of fractionation of the control system in task-specific frontal lobe modules (Shallice and Burgess, 1991) is compatible with the view that no metarepresentation is taking place. The various primary processes present invariant properties that are predicting reliably success or failure in performing the corresponding task. This correlation becomes exploited when the corresponding mechanism is established by evolutionary selection and fine-tuned by learning.

B) The notion of reflexivity at work is thus not necessarily intentional (i.e.

representational) or referential ; it may more plausibly be considered executive, architecture-bound or structural : the output of the control process depends in major part on the feedback it receives from the courses of action in their first-order « simulated » or « attempted » runs. But this dependence does not *need* to be semantic. An interesting « accessibility model » of how the search process is reflexively used in control is offered in Koriat, (1993). Here too, accessibility heuristics does not rely on content, but on properties of the content *vehicles* – for example, trace strength. Thus metacognition necessarily involves neither self-reflexivity, (in the sense of using an integrated representation of one's own mental, social and physical dispositions), nor even mental-state reflexivity, but process-reflexivity.

This leaves us with the question of how such an epistemically implicit control system can be a step towards theory of mind (and to consciousness). Having a procedural form of metacognition puts an organism in a position to gain the corresponding form of declarative knowledge if the conditions for de-modularization are met. Karmiloff-Smith, (1992) hypothesizes a mechanism of « representational redescription » making knowledge contained *in* the mind accessible *to* the mind. Such ideas have since then been explored in the evolutionary history of theory of mind (Povinelli, 2000). In this perspective, a metacognitive control system is a phylogenetic precursor for mentalizing ability not only because it offers procedural knowledge to a potential redescription mechanism, but also because the resulting enhancement of executive capacities offers the control structure that decoupling requires. Inhibiting one's own view of how things look in appreciating another person's perspective, depends on adequate mental control (Perner, 1998). The difference between an implicit, non-mentalistic form of metacognition and its « redescribed » or explicit form, is that reflexivity occurs not only at the process level, but also at the semantic-intentional level (Proust, 2002). Metacognition now can be accessed by metarepresentations, and through language becomes available to self- and other-report, to training, and, here we are, to theorizing.

References

- Anderson, J.R. & Gallup, G.G. Jr. (1997). Self-recognition in *Sanguinus* ? A critical essay. *Animal Behaviour*, 54, 1563-7.
- Bargh, J.A. (1997). The automaticity of everyday life. In R.S. Wyer, Jr. (Ed.), *Advances in social cognition*, vol. 10, 1-61, Mahwah, NJ : Erlbaum.
- Cheney, D.L. & Seyfarth, R.M., (1990), *How monkeys see the world*, Chicago, University of Chicago Press.
- Perner, J., (1998) The metaintentional nature of executive functions and theory of mind, in P. Carruthers & J. Boucher (eds.), *Language and thought*, Cambridge, Cambridge University Press, 270-283
- Karmiloff-Smith, A., (1992), *Beyond Modularity : A developmental Perspective on cognitive Science*, Cambridge, MIT Press.
- Povinelli, D. (2000), *Folk Physics for Apes*, Oxford, Oxford University Press.

Proust , J. (2001). A plea for mental acts, *Synthese*, 2001, 129, 105-128.

Koriat, A. (1993), How do we know that we know ? The accessibility model of the feeling of knowing. *Psychological Review*, 100, 609-639.

Reder, L.M. & Shunn, C.D. (1996). Metacognition does not imply awareness : Strategy choice is governed by implicit learning and memory. In L.M. Reder (Ed.), *Implicit memory and metacognition*, Hillsdale, NJ : Erlbaum, 45-78.

Shallice, T. & Burgess, P. (1991). Higher-Order Cognitive Impairments and Frontal Lobe Lesions in Man, in H.S. Levin, H.M. Eisenberg & A.L. Benton, *Frontal Lobe Function and Dysfunction*, Oxford, Oxford University Press, 125-138.

Shiffrin, R.M. & Schneider, W, Controlled and automatic human information processing : Perceptual learning, automatic attending and a general theory, *Psychological Review*, 1977, 84, 127-190.

Spehn, M.K. & Reder, L.M. (2000). The unconscious feeling of knowing : a commentary on Koriat's paper. *Consciousness and Cognition*, 9, 187-192.
