Metacognition and mindreading: one or two functions?

Joëlle Proust (Institut Jean-Nicod, Paris)

Abstract

Given disagreements about the architecture of the mind, the nature of self-knowledge, and its epistemology, the question of how to understand the function and scope of metacognition – the control of one's cognition - is still a matter of hot debate. A dominant view, the self-ascriptive view (or one-function view), has been that metacognition necessarily requires representing one's own mental states as mental states, and, therefore, necessarily involves an ability to read one's own mind. The self-evaluative view (or two-function view), in contrast, takes metacognition to involve a procedural form of knowledge that is generated by actually engaging in a first-order cognitive task, and monitoring its success. The comparative and developmental arguments supporting, respectively, each of these views are discussed in the light of Hampton's operational definition of metacognition. New arguments are presented in favor of the two-function view. Recent behaviorial and neuroscientific evidence suggests that metacognitive assessment relies on dedicated implicit mechanisms, which are wholly independent, and indeed dissociable, from theory-based self-attribution. The two-function view is claimed to be the best interpretation of these findings.

There is no agreement, in cognitive science or philosophy, about the nature of self-knowledge and its epistemology. No agreement about the functional underpinnings of conscious experience, about the role of emotion in cognition, and about the evolution of the brain. No wonder, then, that at the intersection of all these topics, the function and the scope of metacognition, i.e. cognition about one's cognition, has been hotly debated,[1] and forms the main issue in the present volume. Part of the controversy has to do with the informational processes involved in metacognition. According to a *self-ascriptive view of metacognition* (or SAV), thinkers cannot select, monitor, and control a cognitive activity unless they are also able to reflexively represent *that* they have mental states with specific contents. According to a *self-evaluative view* (or SEV), in contrast, metacognition is one step in a process of active thinking, where agents monitor the available metacognitive feedback in order to adjust their cognitive commands to their cognitive dispositions. In contrast with SAV, however, SEV denies that mindreading is either sufficient or necessary for procedural metacognition. Although procedural metacognition is independent from mindreading, it may be upgraded, when mindreading is available, into analytic, or theory-based metacognition.

The first section summarizes the comparative and developmental arguments supporting, respectively, the existence of one or two different functions associated with "self-knowledge".

[1] See inter alia: Carruthers, (2008, 2009), Proust, (2007, 2010).

Hampton's operational definition of metacognitive behavior is introduced as an important constraint in the discussion. The second section first examines how, in the abstract, such a function might be fulfilled, through a discussion of a theoretical model, called an *adaptive accumulator module (AAM)*. The compatibility of this model with Hampton's definition is discussed, and experimental evidence is presented that AAMs could be major building blocks in procedural metacognition.

In the last section, objections from two angles are addressed. The first claims that procedural metacognition only uses first-order information. The second argues that it engages a form of awareness, which deserves to be classified on a par with analytic, concept-based forms of self-control.

I. Does metacognition have to involve mindreading?

A. The case for a single function.

The theoretical idea behind SAV is the following. Metacognition, by definition, requires from a creature the capacity to represent cognitive activity, in addition to representing a first-order task in which this activity is being exercized (Carruthers, 2008). Nelson and Narens's (1990) two-layered schema for metacognition seems prima facie compatible with this definition, although on a theoretical basis rather than as a definition: any good regulator of a system, they insist, must include a model of that system. This architectural constraint, claimed to form a theorem in the mathematics of adaptive control by Conant and Ashby (1970), has long been taken to entail that a second-order representation of the first-order task, at the control level, is a precondition for adequately monitoring a cognitive task.[2] This inference, however, depends on the assumption that a first-order task can only be modeled by a metarepresentation, i.e. a representation attributing to oneself, for example, the belief of being able (with uncertainty *U*) to correctly perform a cognitive task. This assumption has contributed to shaping the intellectualist stance in metacognitive studies, and inspired mainstream research in educational studies.

Another source of inspiration for SAV, however, has come from developmental psychology. Children tested on various forms of cognitive control, self-evaluation, and source monitoring have been shown to have trouble distinguishing the perceptual appearance from the real nature of objects (such as a sponge that looks like a rock) before they reach 4-5 years

[2] An alternative interpretation of Conant and Ashby's theorem will be offered at the end of this chapter.

of age.[3] Similarly with the control and monitoring of memory: children do not seem to try to retrieve events or names before they have understood that they have a mind able to remember. On the basis of such evidence, Josef Perner has persuasively argued that the development of episodic memory in children derives from to the ability to introspect an ongoing experience and interpret it as representing an actual past event[4]. According to him, children do not possess episodic memory until they are able to understand the representational nature of their own minds. As another SAV theorist, Peter Carruthers, puts it, "It is the same system that underlies our mindreading capacity that gets turned upon ourselves to issue in metacognition".[5]

The development of epistemic evaluations, furthermore, appears to be more or less parallel with that of mindreading. When 3 yr-olds are asked whether they *know* what is inside a box they have never seen before, they, surprisingly, find it difficult to make a reliable judgment. They often answer with a guess, but do not seem to distinguish knowing from guessing before the age of 4 or even later.[6] When asked how long they have known an item of knowledge that was just communicated to them, 3 yr-olds regularly respond that they've always known it.[7] In summary, when asked to verbally report about what they know, what appears to them, what they can remember, etc., children seem unable to offer reliable answers before they are able to read their own minds. However, once they have acquired, through verbal communication, the concepts for the basic mental states, and thereby become able to understand how other agents can be wrong about the world, children learn to attribute errors and misrepresentation to themselves as well.[8] It has seemed, then, that cognitive monitoring relies upon the ability to identify one's mental states as such: understanding, first, that people – as well as oneself - have mental states and mental dispositions, that they may or not be correct, and that such correction depends on the amount and quality of evidence available.

B. The case for two functions

The developmental argument above, however, has been weakened by three types of findings. First, comparative evidence suggests that non-human primates (including monkeys)

[3] Flavell (1979).
[4] Perner & Ruffman (1995).
[5] See Carruthers (2008, 2009, this volume, in press). See also Gopnik (1993).
[6] Sodian et al. (2006).
[7] Gopnik & Astington (1988).
[8] Schneider (2008).

present metacognitive competences comparable to those of humans. Granting this result, primate phylogeny should be reflected in human ontogeny, leading us to expect distinctive developmental patterns for self-evaluation and mindreading. Second, recent data indicate that human 3 yr-olds indeed present the same metacognitive performances as monkeys, even though they do not yet solve "false belief task" problems. Third, a series of studies suggest that mindreading is also a biological, rather than a merely cultural ability, which surfaces in various early implicit forms of social sensitivity to others' intentions and beliefs. This hypothesis makes previous correlations between mindreading and metacognition more difficult to interpret, and suggests an independent role for an executive function capacity in both types of performance. We will briefly examine these findings in turn.

1. Comparative evidence about metacognition as distinct from mindreading

A powerful argument against SAV comes from comparative psychology. Various non-human species that are not adapted to read minds, such as bottlenosed dolphins (*Tursiops truncatus*), and rhesus macaques (*Macaca mulatta*), are able to evaluate whether they are able to discriminate two visual stimuli; they can make a prospective judgment of memory in a serial probe recognition task (Hampton, 2001, Smith et al., 2003). In such tasks, the animals are offered the opportunity to "opt out" from a perceptual or memory task when they feel unable to perform it. The animals' response patterns strikingly resemble those of human subjects. Granting the validity of these experiments, they are compatible with the view that metacognition is a specific adaptation, whose phylogenetic distribution overlaps, but does not coincide, with the ability to read minds. Mindreading is used here to refer to the capacity of identifying beliefs, i.e. what Carruthers & Ritchie, this volume, call "stage 2 mindreading", rather than intentions or spatial perspective. Mindreading so understood is a uniquely human ability). Two preliminary issues must be clarified. First, do the methodological difficulties attached to these experiments threaten the validity of this view? Second, supposing that they don't, how could one operationalize the concept of "metacognition" in non-verbal agents?

A - Methodological concerns

Important methodological concerns have been raised against a hasty metacognitive interpretation of these findings (Carruthers, 2008, 2009, Hampton, 2009, Crystal & Foote, 2009). First, is it not *reward*, rather than the animals' judgments of confidence, that guide decisions? To address this problem, animals were denied any access to reinforcement

scheduling, and offered blockwise, rather than trial-by-trial reinforcement. This modification did not affect their metacognitive performance (Smith et al., 2006, Couchman et al., 2010). Second, are not so-called "metacognitive judgments" actually prompted by *associations between environmental cues*? This worry has been addressed through generalization tests, where the animals need to predict performance in unrelated tasks (Kornell et al., 2007). When the animals immediately transfer their disposition to opt-out, (e.g., from a perceptual to a memory task), it is safe to assume that metacognitive ability is not dependent on the associative strength of the stimuli involved (Hampton, 2009, Couchman et al., 2010). Third, are not difficult trials merely *aversive* ones? In a discrimination test, "middle" stimuli on a continuum might be avoided, not on the basis of a confidence judgment (a judgment of uncertainty), but simply because animals dislike categorizing them (Perner, unpublished communication). Several ways of addressing this question have been considered. First, it was shown that capuchin monkeys are able to sort stimuli into three categories, A, B, and middle. However, they are unable to use uncertainty as a motivation to decline difficult trials on an A-B only task, as rhesus monkeys do, although they thereby incur the cost of long timeouts (Beran et al., 2009). Second, a threshold task, which does not allow a "middle" category to emerge, has elicited adaptive uncertainty responses in rhesus monkeys (Couchman et al., 2010).

B - A step forward: an operational definition of metacognition

The methodological problems above have emphasized the need for determining more carefully what counts as a metacognitive capacity. Influenced by SAV, some theorists have claimed that, by definition, metacognition should be based on a secondary *representation.*[9] It is sometimes also claimed that metacognition should be mediated by introspection, with a higher-order conscious state allowing the animal to form a judgment of uncertainty in a trial on the basis of its epistemic feelings. These definitional requirements being difficult to operationalize, however, a less contentious distinction has been offered between a primary and a secondary *behavior*, or *goal*. Robert Hampton proposes the following list of 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.

[9] Crystal & Foote, 2009b, p. 54.

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).

On top of these four functional conditions, however, Hampton defines a more restrictive metacognitive capacity, which he calls *"private metacognition"* (in contrast with a "public" form: Hampton, 2009). The functional advantage that private metacognition offers is that it enables animals to respond to uncertainty in a generalized way, through endogenous signals, rather than through separately learnt, task-specific associations available to an external observer. The mechanisms for Private metacognition must fulfill three additional, negative conditions.

i) The metacognitive responses must not be based on response competition (where perceptually presented stimuli are merely selected on the basis of their comparative attraction).

ii) They must not be based on environmental cue association.

iii) They must not be based on behavioral cue associations, i.e. "ancillary responses" such as hesitation, or response latency.

Hampton's three constraints on mechanisms are meant to reveal a capacity for "private" procedural metacognition. We now have, then, three different candidates for a metacognitive function, that might concurrently fulfill the operational definition above: public metacognition (based on publicly available cues), private metacognition (based on internal cues), and mindreading (based on representations of one's mental states). Experimenters aiming to demonstrate private procedural metacognition, Hampton shows, can do so on the basis of a limited number of paradigms. Because it occurs only once a response is given, wagering allows us to disconnect the metacognitive appraisal from the competition of stimuli (condition i). By modifying the stimuli involved in the task, tranfer tests can control for (ii). Finally, checking on latency times should allow (iii) to be controlled for.

Taking all these conditions together, a few paradigms indeed seem to effectively rule out the effect of exogeneous or public influences over metacognitive evaluations. They are the *retrospective gambling paradigm* (also called "wagering"), and some forms of the *prospective opt-out test*, where animals are asked to decide whether or not to perform a task without simultaneously perceiving the test stimuli (Hampton, 2009). Animal research thus seems warranted in claiming that private procedural metacognition is manifested in animals that do not have the ability to read their own minds, or other minds.

2. Developmental evidence favoring a two-function view

Granting that non-humans present procedural metacognition, it would be likely that human children should also do so. Although, as we saw above, developmental evidence has long pointed to late development of epistemic self-monitoring – with a schedule parallel to mindreading – , it is now realized that the evidence for delayed metacognition might be related to the attributive (or "explicit") style of most of the tests that were used (Balcomb & Gerken, 2008). As we have seen above, children of three, when tested verbally about what they know (versus what they guess), normally fail to form correct self-attributions of knowledge. However, dissociations frequently occur, in human cognition, between verbal report and behavioral decision. Given the crucial importance of learning and selective information acquisition in our species, it would be very surprising that infants have no sensitivity to the quality of their informational states.[10]

If metacognition is present in young children, as it presumably is in monkeys, a promising method would consist in studying their epistemic behavior with the paradigms used in comparative psychology. Call & Carpenter (2003), using a set of opaque tubes where food or toys were hidden, showed that 3 yr-old children are able to collect information only when ignorant, with performances similar to those of chimpanzees and orangutans. This study, however, did not allow one to determine whether the secondary behavior was produced by response competition or by access to one's epistemic uncertainty (Hampton, 2009). Another option is to use an opt-out paradigm, which is what Balcomb & Gerken (2008) did: they used Smith et al.'s test of memory-monitoring in rhesus monkeys to test children aged 3 and a half. The children first learn a set of paired pictures, representing an animal (target) and a common object (its match). In the subsequent test, they are shown one item of a pair and two possible associates: the match and a distractor; their task is either to select the match, or decline the trial (the stimuli were arranged so that matches and distractors were equally familiar: familiarity could not be used as a cue). Finally, they are given a forced recognition test where they have to select the match of each animal. This study showed that children were adequately monitoring their memory by opting out on the trials they would have failed. A

[10] It is well-known that babies distinguish novel from familiar stimuli: they seem to prefer looking at a familiar object before becoming habituated (before learning), and at a new object thereafter (Hunter et al., 1983). The function of these preferences is clear: adequately targeted cognitive interest allows infants and adults to optimize learning. Another case in point consists in the capacity of 5-month infants to allocate their attentional resources as a function of the type of information they need to extract (for example: species- or property- level information) (Needham & Baillargeon, 1993, Xu, 1999). These early types of control of attention, however, do not yet qualify as metacognitive to the extent that the secondary behavior (appreciating the degree of familiarity with a stimulus) seems to be directly wired into the infant's learning system; as a result, response competition can explain behavior without invoking a metacognitive decision.

second experiment indicated that they could do so prospectively even when the *only* stimulus presented at the time of decision was the picture of the match (preventing a response competition effect). This experiment thus fulfills the various constraints listed above for metacognition. Furthermore, it also seems to offer evidence for "private metacognition" in children who are not able yet to solve a false belief task.

3. Objection: what if Mindreading is a biological, low-level ability ?

A series of studies, however, suggesting that mindreading is an early biological, rather than cultural ability, surfacing in various implicit forms of social sensitivy to other's intentions and beliefs, has brought a twist in the one/two function debate. Onishi and Baillargeon (2005) reported that 15-month-old infants have insight into whether an agent acts on the basis of a false belief about the world. In addition, Kovacs et al. (2010) present evidence that the mere presence of social agents is sufficient, in 7-month-old infants as well as in adults, to automatically trigger online computations about others' goals.[11] As a consequence, mindreading abilities are seen as an innate "social sense," that is spontaneous, automatic, and effortless. The relevance of this type of evidence is interpreted differently by SAV and by SEV proponents. SAV proponents, when they take these results as reliable evidence for mindreading,[12] may argue that mindreading, with its early influence on behavior, is in a position to drive any form of self-evaluation. They need to assume, however, that additional executive and attentional competences explain the late performance of children on high-level, language-dependent tasks such as completing a false-belief task or offering a verbal epistemic self-evaluation.[13] They need, in addition, to downplay the comparative evidence in favor of private metacognition in monkeys.

SEV proponents may argue, in contrast, that if early forms of mindreading are present in infants, then the first appearance, around 4-5 years of age, of metacognitive competences is no longer correlated with, and explainable by, a newly acquired mindreading ability. Delayed metacognition, and delayed false-belief understanding, might be due to extrinsic competences respectively engaged in each function. One way of adjudicating among these two interpretations would involve exploring the mechanisms that might be respectively engaged

[11] This ability belongs to goal prediction, which has been found to be available to infants in their first year (Gergely et al., 1995). Although this ability is sometimes called "stage-1 mindreading" (Carruthers and Ritchie, this volume), reading a mind is usually defined as a capacity to understand that one's own and others' beliefs can be false.

[12] For an interpretation of Baillargeon's results in terms of behavioral cues, rather than of mindreading, see Perner & Ruffman (2005).

[13] Carruthers, (2009).

in metacognition and in mindreading in the human adult.

II. Do metacognition and mindreading differ in their informational mechanisms?

The most convincing argument in favor of a two-function view would be to show that the informational mechanisms that produce a self-prediction and an other-directed attribution are substantially different, and, to this extent, can produce diverging outcomes. Theorists of noetic judgments have contrasted experience-based and theory-based forms of self-evaluation.[14] Experience consists in feelings, generated by the processes underlying cognitive operations rather than by the agents' attitudes (such as: having a belief) or their outcomes (a belief with a particular content).[15] As we shall see, it can further be hypothesized that the processes that guide self-evaluations in procedural metacognition include a model of the first-order cognitive task; the dynamic properties of the neural vehicle are extracted, and relied upon to model (i.e. monitor and control) the ongoing task. In a nutshell, what makes this model epistemically adequate is that the dynamic properties of the vehicle map the epistemic properties of the computational processes involved.

Mindreading-based metacognition, on the other hand, can develop predictions on the basis of a naive theory of the first-order task, and of the competences it engages. The latter thus requires representing both one's own propositional attitudes (such as beliefs and desires) and their contents (that the chocolate is in the drawer). On a two-function view, theoretical metacognition consists in general of knowledge about cognitive dispositions, whereas procedural metacognition is the ability to conduct cue-based self-evaluations. Although mindreading can redescribe and enrich procedural metacognition, it is, from a SEV viewpoint neither necessary, nor sufficient, to perform contextually flexible metacognitive judgments.

A - A behavioral dissociation between procedural metacognition and theory-based prediction

According to SAV, the same basic informational processes are involved in self- and other-mental attribution. Therefore knowledge made available to oneself through introspection, or self-directed interpretation, should be automatically transferred to others, and reciprocally: knowledge gained about others should be automatically transferred to self.

[14] Koriat & Levy-Sadot, 1999.

[15] See Koriat & Levy-Sadot (1999), Schwarz, (2002). See Dokic's chapter (this volume) for a discussion of the nature and intentional contents of noetic feelings.

Results at variance with this prediction have been obtained by Koriat & Ackerman (2010). Participants are asked to memorize – in a self-paced way – pairs of unrelated words. When they have finished learning a given pair, they are asked to offer a judgment of learning (JOL) about their chances to recall this particular pair. This judgment, however, is elicited in two conditions. In condition A-B, the participants first perform the learning task, with a self-evaluative phase after studying each pair (condition A). They then observe another participant performing the task, and are asked to assess the latter's later ability to recall this particular pair (condition B). In condition B-A, the order is reversed: participants first observe another perform the task and predict her success, then perform it themselves.

A simple SAV prediction is wrong on two accounts. First, *the validity of a judgment of learning for a given pair differs* when participants have performed the task before judging, or merely observed another's performance. When they have performed the task, the participants seem to rely on an implicit Memorizing Effort heuristic, that more study time predicts less recall, which turns out to reliably predict successful performance. In contrast, when predicting another agent's ability *before* having performed the task themselves*, subjects rely on a piece of (wrong) folk-theorizing*, that more study time predicts more recall. This suggests that self-evaluation in A elicits a form of procedural, context-sensitive access to the subjective uncertainty associated with a trial, while other-evaluation in B relies on general background conceptual knowledge about successful learning (disregarding the contextual fact that pairs are of unequal difficulty, and that the time spent on a pair reflects that fact).[16]

Second, *transfer turns out to be different* in the A-B and in the B-A conditions. In the A-B condition, the acquisition and transfer to others in B of the metacognitive knowledge acquired in A, in the experimental settings described above, is found to reliably occur. In the B-A condition, in contrast, participants who, in task B, have merely observed others perform, do not transfer to themselves, in task A, their prediction about others that more time predicts better learning. The reason they do not, clearly, is that engaging in the metacognitive task themselves allows them to extract additional information that they did not have when merely observing others perform the task.

At this point, SAV theorists might object that a subject, when engaged in a metacognitive task, has access to introspective evidence that she fails to have when she is merely observing

[16] There are cases where the dissociation goes the other way round: observers predict more accurately the effects of retention interval for learning in others than in themselves (Koriat et al., 2004). The explanation is the same in both cases, however: procedural metacognition relies on process-based feelings, such as retrieval fluency, which can be a source of illusion, while theory-based control is more prone to involve conceptualizing that time is relevant to prediction of correct retrieval.

another agent. Thus it is expected in SAV terms that i) the validity of the self-evaluations should differ in the two cases, and ii) that the generalization of knowledge should be asymmetric. In response to this objection, however, note that the participants in the Self condition are unaware of using the implicit effort heuristic. None of them reports, after the experiment, having based their own judgment of learning on an inverse relation between study time and learning. In contrast, participants in the Other condition report having used it to predict learning in others. What does this show? The authors observe that a shift has occurred from experience-based to theory-based JOLs, and that this shift is associated with the need to provide an explicit evaluation of learning in others. Indeed this metacognitive task invites subjects to integrate their own experience with someone else's, which might help the participants to make the underlying effort heuristic explicit. The upshot is that participants do not use the same *kind of knowledge* when predicting learning in others in the A-B and the B-A conditions. In the A-B condition, the knowledge collected in A has its source in the experience generated by a metacognitive engagement. The resulting metacognitive decisions, once made, can subsequently be generalized to another performer based on the subject's general inferential abilities. In the B-A condition, however, the prediction of others' learning relies on a tenet of the naïve theory of memory, according to which longer study time predicts better learning.

Thus a more natural explanation for the dissociation discussed above is that procedural metacognition and mental attribution engage two different types of mechanisms. Engaging in a task with metacognitive demands allows the agent to extract "activity-dependent" predictive cues, i.e. associative heuristics that are formed as a result of the active, self-critical engagement in a cognitive task. Predicting success in a disengaged way, in contrast, calls forth theoretical beliefs about success in the task. While activity-dependent cues offer a contextual evaluation, theory-laden cues at work in mindreading rely, rather, on conceptual knowledge, which may fail to be sensitive to causally relevant features of potential success in the task.

Additional evidence in favor of this contrast is offered by a third experiment, where the self-other condition is modified. Now participants learning pairs of words in condition A are *not* invited to form a judgment of learning. Will they still apply the memorizing effort heuristic when subsequently predicting learning in others? Interestingly, they fail to do so, with results closely similar to the Other-first condition. This finding, then, suggests that an implicit heuristics is extracted and used only when the task requires making a judgment of learning for each pair. This makes "activity-dependence" of cue-learning more precise:

Engagement in self-evaluation, rather than mere engagement in a first-order cognitive task, is a precondition to having the relevant experience, and to transferring it to others.

In summary, an experience of active control-and-monitoring of learning - an idiosyncratic interaction between the learner and the items to be learned associated with an evaluative stance – is needed for subjects to form the correct association between study time and successful retrieval. Transfer to others, however, depends on having conceptually represented the regularity- an ability that might not be available to animals with no such conceptual knowledge. Transfer to others of one's metacognitive experience thus requires mindreading – theorizing about mental states as such – as a necessary step.

The next question, then, concerns the mechanisms that might be selectively engaged in procedural metacognition.

B - The double accumulator model: theory and evidence

1. Theory

From classical studies on metacognition and on action, we know that any predictive mechanism needs to involve a *comparator*: without comparing an expected with an observed value, an agent would not be able to monitor and control completion of a cognitive task (Nelson & Narens, 1990). When prediction of ability in a trial needs to be made, the agent needs to compare the cues associated with the present task with their expected values. As we saw above, these cues can, theoretically, be public. For example, the physical behavior that is associated with uncertainty (hesitation, oscillation) might be used as a cue for declining a task (which cue, being of a non-introspective kind, is advanced as a reason to favor SAV: see Carruthers, 2008).

There are more efficient ways of evaluating one's uncertainty, however, which do not depend on actual behavior, but only on the informational characteristics of brain activity. The dynamics of activation in certain neural populations can in fact predict – much earlier and more reliably than overt behavior – how likely it is that a given cognitive decision will be successful. The mechanisms involved in metaperception (i.e., in the control and monitoring of one's perception), described by Vickers and Lee (1998) and (2000), have been called *adaptive accumulator modules* (AAM). An adaptive accumulator is a dynamic comparator, where the values compared are rates of accumulation of evidence relative to a pre-established threshold. The function of this module is to make an evidence-based decision. For example, in a perceptual task where a target might be categorized as an X or as a Y, evidence for the two

alternatives is accumulated in parallel, until their difference exceeds a threshold, which triggers the perceptual decision. The crucial information used here consists in the differential rate of accumulation of evidence for the two (or more) possible responses.

Computing this difference - called the balance of evidence – does not yet, however, offer all the information necessary for cognitive control. Cognitive control depends on a secondary type of accumulator, called "control accumulator". In this second second pair of accumulators, the balance of evidence for a response is assessed against a desired value, itself based on prior levels of confidence associated with that response. Positive and negative discrepancies between the target-level and the actual level of confidence are now accumulated in two independent stores: overconfidence is accumulated in one store, underconfidence in the other. If, for example, a critical amount of overconfidence has been reached, then the threshold of response in the primary accumulator is proportionally reduced. This new differential dynamics provides the system with internal feedback allowing the level of confidence to be assessed and recalibrated over time. [17]

A system equipped to extract this additional type of information can thereby model the first-order task on the basis of the quality of the information obtained for a trial. Genuinely metacognitive control is thus made possible: the control accumulator device allows the system to form, even before a decision is reached, a calibrated judgment of confidence about performance in that trial. Computing the difference between expected and observed confidence helps an agent decide when to stop working on a task (in self-paced conditions), how much to wager on the outcome, once it is reached, and whether to perform the task or not. Granting Vickers & Lee's (2000) assumption that adaptive accumulator modules work in parallel as basic computing elements, or "cognitive tiles", in cognitive decision and self-evaluation, granting them, furthermore, that the information within each module is commensurable throughout the system, a plausible hypothesis is that these accumulators underlie procedural metacognition in non-humans as well as in humans, in perception as well as, mutatis mutandis, in other areas of cognition.

Let us check that our four conditions listed above are fulfilled by a double-accumulator system. There is clearly a *primary behavior*, i.e. a primary perceptual or memory task in which a decision needs to be taken. Second, *variation* in performance, i.e. uncertainty in outcome, is an essential feature of these tasks, generated by endogenous noise and variations in the world. Third, the *secondary behavior*, in control accumulators, consists in monitoring confidence as a function of a level of "caution": a speed-accuracy compromise for a trial

[17] See Vickers and Lee (1998, p.181)

allows the decision threshold to be shifted accordingly. Fourth, secondary behavior obviously *benefits* performance on the primary task, because it guides task selection, optimizes perceptual intake given the task difficulty, the caution needed and the expected reward, and, finally, reliably guides decision on the basis of the dynamic information that it makes available.

2. Evidence for Adaptive Accumulator Modules in procedural metacognition

An empirical prediction of AAM models of cognitive control and monitoring bears on how the temporal constraints applying to a task affect a confidence judgment. When the time for which the stimulus is available in a perceptual task is determined by the experimenter, -- supposing discriminability is constant--, the participant's confidence judgment is a direct function of the time for which the stimulus is available (as the prediction is only based on the difference between rates of accumulation for that duration). If, however, the agents can freely determine how long they want to inspect or memorize the stimulus, other things being equal, the prediction is now based on the comparison of the dynamics of the accumulation of the evidence until the criterion is reached, relative to other episodes. Thus, in a self-paced condition, both probability of correctness and associated confidence are *inversely related to the time needed to complete the task* (Vickers & Lee, 1998, p. 173). These results are coherent with the research conducted on judgments of learning and judgments of confidence for tasks that have either a fixed, or a self-paced, duration (Koriat & al., 2006).

Further experimental evidence in favor of this theoretical construct comes from the neuroscience of decision-making. Here are a few examples. The first concerns the role of accumulators in metacognitive judgments in rodents. Kepecs et al. (2008) trained rats on a two-choice odour categorization task, where stimuli were a mixture of two pure odorants. By varying the distance of the stimulus to the category boundary, the task is made more or less difficult. Rats were allowed to express their certainty in their behavior, by opting out from the discrimination task. Conditions 3 and 4 in Hampton's conditions for procedural metacognition are thus met. The neural activity recorded in the orbitofrontal cortex of rats was found to correlate with anticipated difficulty, i.e. with the predicted success in categorizing a stimulus (with some populations firing for a predicted near-chance performance, and others firing for a high confidence outcome). Furthermore, it was shown that this activity did not depend on recent reinforcement history, and could not be explained by reward expectancy. Vickers' control accumulator model offers an explanation: the distance between decision variables, expressed in the differential evolutions in the firing rates, can provide a reliable estimate of

confidence in the accuracy of the response. No evidence is collected in this study, however, about the control-accumulator described in Vickers & Lee.[18]

Kiani & Shadlen (2009) also use AAMs to account for the capacity of rhesus monkeys to opt out from a perceptual discrimination task, and choose, instead, a "sure target" task, on the basis of the anticipated uncertainty of the task. Interestingly, it is activity of populations of neurons in the monkeys' *lateral intraparietal cortex* that was found to represent both the accumulation of evidence, and the degree of uncertainty associated with the decision. The animals, again, satisfy Conditions 3 and 4 in Hampton's list by opting for the sure target when the stimuli were *either* poorly discriminative *or* briefly presented. Moreover, their accuracy was higher when they waived the option than when the option was not available. Finally, a study by Rolls et al. (2010) explores an alternative model for olfactory decisions in humans, "the integrate-and-fire neuronal attractor network". This model shares with AAMs the notion that decision confidence is encoded in the decision-making process by a comparative, dynamic cue. Here, the information is carried by differences between increments (on correct trials) and decrements (in error trials) as a function of ΔI (relative ease of decision) of the BOLD signal (i.e. the change in blood flow) in the brain regions involved in choice decision-making. These regions involve, inter alia, the medial prefrontal cortex and the cingulate cortex. This model, however, does not clearly raise the question of how confidence is calibrated, and thus fails to explore the structures allowing metacognitive control.

The models presently used for procedural metacognition tend to suggest, then, that it depends on *two* objective properties of the *vehicle* of the decision mechanisms: first the way the balance of evidence is reached carries dynamic information about the validity of the outcomes; second, the history of past errors, i.e. the observed discrepancies beween a target level of confidence and the actual level obtained, carries information about how to adjust the threshold of confidence for a trial, given internal constraints relative to speed and accuracy. Calibration of confidence thus results from a separate dynamic process, storing the variance of the prior positive or negative discrepancies.

In summary, a judgment of confidence is not formed by re-representing the particular content of a decision, or by directly pondering the importance of the outcome. Nor does it require that the particular attitude under scrutiny be conceptually identified (e.g. as a belief). Confidence is directly assessable from the firing properties of the neurons, monitored and stored respectively in the sensory and the control accumulators. A natural suggestion is that

[18] Variance of the decision variables is shown to offer an equivalent basis for confidence judgments, if an appropriate calibration of the criterion value has been made available by prior reinforcement.

metacognitive feelings, such as feelings of perceptual fluency, are associated with ranges of discrepancy in accumulators. [19]

III. Cognition, procedural metacognition and mindreading

Proponents of procedural metacognition as well as supporters of a one-function view might reject the present proposal on various, and indeed incompatible, grounds. Some will find the role of AAMs in procedural metacognition compatible with a no-metacognition view, where secondary behavior is seen as reducible to primary task-monitoring. Others will observe, on the contrary , that adaptive accumulators cannot, as isolated modules, perform all the tasks involved in metacognitive functions. They need to be supplemented by other functional features, such as conscious awareness, attributive and inferential mechanisms, etc., which casts doubt on the claim that procedural metacognition does not need to involve some form of stage-1 mindreading. Finally, it will be observed that the present proposal contrasts two forms of self-knowledge in their respective evolutionary and informational patterns, but does not consider whether, and if so, how, procedural metacognition and mindreading can be integrated into a higher-order form of metacognition.

A. Objection 1: "Procedural metacognition" boils down to primary task-monitoring

The evidence about AAMs summarized above might look too close to usual forms of feedback from action to deserve a qualification as metacognitive. If feelings of uncertainty are emergent on the structural properties of decision processes, are they not, finally, "directed at the world (in particular, at the primary options for action that are open to one), rather than at one's own mental states?", as Carruthers and Ritchie write in this volume?[20] From the viewpoint of the animal, it might be that felt uncertainty, or judgments of confidence, are directed at the problem of *how to act in order to get an optimal reward*. In this case, a motivational explanation should be sufficient to account for the kind of monitoring that is supposed to occur in procedural metacognition. A slightly different interpretation of the evidence would claim that the animal feels a conflict between prior expectation and current belief, as in surprise. The existence of such a feeling of conflict, however, does not yet qualify as *metacognitive*. Any emotion, and even any behavior, will carry information about a

[19] For lack of space, we will not discuss this suggestion in the present chapter. A theory combining some features of Dokic's "Water diviner model" and of the "competence model" (Dokic, this volume) could explain how feelings generated by accumulator discrepancy can predict likely success in a given cognitive performance.

[20] See also Carruthers (2008).

primary task; this does not warrant the conclusion that it is metacognitive.[21]

In order to address these objections, it must first be emphasized that the mechanisms assumed to underly procedural metacognition have an *epistemic* function: this consists in evaluating the validity of a cognitive decision, which contrasts both with a directly *instrumental* function, such as obtaining more reward, and an *executive* function, consisting in allocating more attentional resources to a task. Why might such an epistemic adaptation have evolved? The success of an action – where success is assessed in terms of reward and risk avoidance - presupposes that an organism stores instrumental regularities: in a changing environment, it must be in a position to take advantage of recurring patterns to satisfy its needs. But success of an action also depends on controlling one's cognition, i.e. performing cognitive actions such as directed discriminations or retrievals. This control, however, crucially involves monitoring epistemic deviance with respect to a norm.[22] Just as physical actions are prepared by simulating the act in a context, and need to be evaluated for termination to occur, cognitive actions are prepared by evaluating the probability of the correctness of (and terminated by evaluating the probability of the adequacy of) a given decision. In brief, when predation is high, foraging difficult, or competition high, selective pressure is likely to arise for a capacity to distinguish, on an experiential basis, cases where the world has been changing, or where insufficient information was used to make an epistemic decision. Thus procedural metacognition entails sensitivity to the level of information available; it also entails sensitivity to alternative epistemic norms, such as speed and accuracy, which determine different thresholds of epistemic decision. In contrast with surprise, which is a built-in response meant to increase vigilance, noetic feelings – such as the feeling of confidence – are able to adjust to task and context in a flexible way, as manifested in adequate opting out.

A common mistake in psychophysics consists in failing to distinguish the function of a primary accumulator, which is to make a certainty decision for the current trial, from that of a secondary accumulator, which is to extract the dynamics of error information over successive trials, in order to calibrate the primary accumulator's predictions. The latter function constitutes a different adaptation, as is shown by the fact that, although all animal species have some decision mechanism, few of them monitor the likelihood of error to predictively choose what to do, or to wager about their decision. Indeed the information needed to *make a decision under uncertainty* is not the same as the information used in *assessing one's*

[21] Carruthers (2008).
[22] See Proust (2012).

uncertainty. A decision to do A, rather than B, is made because of A's winning a response competition where the "balance of evidence" is the basis of comparison. Assessing one's uncertainty, in contrast, relies both on the differential dynamics of the response competition throughout the task, and on an additional comparison between the positive and negative discrepancies between the target and the actual levels of confidence across successive trials. From this analysis about function, we can conclude that an accumulator, potentially, can provide epistemic information, rather than merely carrying it, because it carries it as a consequence of having the function of regulating epistemic decisions: thus the information can be put to use, by a more sophisticated mechanism for controlling epistemic decisions. It appears to be the case that some animals do have such a more sophisticated mechanism.

Now an important question is whether the secondary accumulator, or control accumulator, might be interpreted as metarepresenting the cognitive dispositions manifested in the primary accumulator. Metarepresentation, in general terms, applies to propositional contents attributed under an attitude term to an agent or thinker. Here, no such attributional-propositional process is present. There are, however, interesting similarities and differences between a control-accumulator and a metarepresentation. A metarepresentation offers conceptual information about the content of a mental state, e.g. of a belief; it offers a conceptual model for it. A control-accumulator also models thought; it offers, however, non-conceptual, analogic information about the probability of error/accuracy in confidence judgments, which themselves bear on the outcome of a primary cognitive task. In contrast with metarepresentation, no attitude concept is used in a control accumulator. There is, however, a functional coupling between the primary and secondary accumulators, which guarantees that the secondary accumulator predicts confidence based on evidence in the first, and – through its control architecture - that the second is "about" the first. This "aboutness" is reflected in the fact that the noetic feelings are directed at, and concern, the first-order task, i.e. what the animal is trying to do.

Finally, a metarepresentation may allow the organism to predict behavior, but does not have a fixed rational pattern associated with its predictive potential. Here, in contrast, predictions at the control level immediately issue in adapted cognitive behavior: information is process-relative, modular and encapsulated. It only allows an agent to adaptively modify its current cognitive behavior. To explain, and thus remedy persistent discrepancies between expected and observed cognitive success, agents may need to have conceptual knowledge available. Furthermore, various illusions are also created when relying on accumulators to

make confidence predictions for abilities they cannot predict (for example, in judging what one will remember at a retention interval on the basis of felt fluency[23]). This narrow specialization of self-prediction is a signature of procedural, as opposed to analytic metacognition.

B. Objection 2: accumulators are only ingredients in procedural metacognition

A second objection will note, on the contrary, that adaptive accumulators, even if crucial ingredients, are merely ingredients in a larger set of processes involved in metacognition. The indeterminacy of the elements contained in this larger set raises doubts about whether procedural metacognition does not need to involve, for example, stage-1 self-applied mindreading.

It is currently accepted in neuroscience that accumulators are automatic error detection modules, operating in every brain area. Other systems, however, have been proposed to play a role in metacognitive regulation and control. A "conflict monitoring system", located in the anterior cingulate cortex, is known to have the function of anticipating error and correcting it on line. This system is based not on confidence judgments and control accumulators, but on the fact that working memory can activate processing pathways that interfere with each other (by using the same resources or the same structures), a situation which makes processing unreliable.[24] Furthermore, an analytic, conscious, deliberate conceptual system has been found, in humans, to contribute to metacognitive judgment, and sometimes to override confidence judgments resulting from the procedural metacognition.[25] This documented variety of mechanisms, however, does not warrant the one-function view. Rather, it emphasizes the phylogenetic difference between procedural and analytic metacognition. The first type relies on a variety of mechanisms to detect and control error; the second is a distinct adapation, which enables agents to understand error as false belief.

The neurophysiological and experimental evidence discussed above, furthermore, suggests that feelings of confidence are not mediated by a conception of the self nor by higher-order attributional mechanisms. In accord with this evidence, it should be stressed that the brain areas respectively involved in metacognition and in mindreading do not seem to overlap.[26] The first include, in humans, the sensory areas (primary accumulators), the

[23] Cf. Koriat et al. (2004).

[24] Botvinick et al. (2001).

[25] Koriat & Levy-Sadot (1999).

[26] I am deeply indebted on this matter to Stan Dehaene's Lectures on metacognition at the Collège de France, Winter 2011.

dorsolateral prefrontal cortex and the ventro-medial prefontal cortex, in particular area 10 (where control accumulators may be located) and the anterior cingulate cortex. Lesion studies show that the right medial prefrontal cortex plays a role in accurate feeling-of-knowing judgments.[27] Transcranial magnetic stimulation applied to the prefrontal cortex has been further shown to impair metacognitive visual awareness.[28] Mindreading, in contrast, involves the right temporal-parietal junction, the prefrontal antero-medial cortex, and anterior temporal cortex.[29]

Another argument can be drawn from a behavioral phenomenon called " immunity to revision of noetic feelings". In a situation where subjects become aware that a feeling has being produced by a biasing factor, they are in a position to form an intuitive theory that makes subjective experience undiagnostic. In such cases, the biased feelings can be controlled for their effect on decision.[30]. The experience itself, however, survives the correction.[31] Why does experience present this strange property of immunity to correction in the face of evidence?

Nussinson & Koriat (2008) speculate that noetic feelings involve two kinds of "inferences".[32] In a first stage, a "global feeling", such as a feeling of fluency, is generated by "rudimentary cues concerning the target stimulus", which are activity-dependent.[33] In a second stage, a new set of cues are now identified in the light of available knowledge about the stimulus, the context, or the operation of the mind. A new judgment occurs using conscious information to interpret experience. The imperviousness of experience to correction might thus be causally derived from the automatic, unconscious character of the phase 1 processing that generates it. Such a two-stage organization of feelings, and the fact that the experience and associated motivation to act cannot be fully suppressed or controlled, speak in favor of our two-function view.

Conclusion

[27] Schnyer et al. (2004), Del Cul et al. (2009).

[28] Rounis et al. (2010).

[29] Perner & Aichorn (2008).

[30] Unkelbach (2007) shows, for example, that participants can attribute to the same feeling of fluency a different predictive validity in a judgment of truth.

[31] Nussinson & Koriat (2008).

[32] It may be found misleading to use the same term of "inference" for an unconscious predictive process, which seem to rely on the neural dynamics of the activity or, as the authors hypothesize, on implicit heuristics, and for a conscious, conceptual process, which can integrate the subject's knowledge about the world.

[33] In the interpretation offered here, the implicit cues and heuristics ultimately consist in the dynamics of the paired accumulators.

This chapter has defended a two-function view of self-knowledge. One function consists in procedural metacognition, a capacity that has been proposed to depend crucially on the coupling of control and monitoring accumulator mechanisms. Blind to contents, this form of self-evaluation takes as its input dynamic features of the neural vehicle, and yields practical epistemic predictions as output, concerning whether the system can, or cannot, meet a normative standard in a given cognitive task. It is, thus, contextually sensitive to attitudes and to their associated conditions of correction. The other source of self-knowledge is conceptual; mindreading offers human beings a conceptual understanding of their own cognitive dispositions, which in turn allows them to override, when necessary, the decisions of procedural metacognition. These two routes to self-knowledge have a parallel in the so-called "dual-process theory" of reasoning, where "System 1" is constituted by quick, associative, automatic, parallel, effortless, and largely unconscious heuristics (such as the availability heuristics), while "System 2" encompasses slow, analytic, controlled, sequential, effortful, and mainly conscious processes.[34] The present discussion suggests that self-evaluation might similarly depend on two such systems. Noetic feelings seem to be the subjective, emotional correlates of subpersonal accumulator features such as neural latency, intensity and stability; they are also immune to revision: all features associated with System 1. If they deliver consistently inappropriate predictions, i.e. produce metacognitive illusions, controlled processing of System 2 is supposed to step in, as its presumed function is to "decontextualise and depersonalize problems".[35] An open question remains, at this point: is such stepping-in entirely dependent on a mindreading capacity? The present state of the literature suggests a positive answer, but comparative psychology might still surprise us.

Acknowledgments

I am grateful to Richard Carter both for his linguistic help and his comments, and to Michael Beran, David Smith, and Kirk Michaelian, for their critical observations on a prior draft of this chapter.

References

Balcomb, F. K. & Gerken, L. (2008). Three-year old children can access their own memory to guide responses on a visual matching task. Developmental Science, 11: 5, 750-760.

Beran, M. J., Smith, J. D., Coutinho, M. V. C., Couchman, J. J., & Boomer, J. (2009). The

[34] Stanovitch & West, (2000).

[35] Stanovitch & West, (2000), 659.

psychological organization of "uncertainty" responses and "middle" responses: A dissociation in capuchin monkeys (*Cebus apella*). *Journal of Experimental Psychology: Animal Behavior Processes, 35,* 371-381.

Botvinick, M.M., Braver, T.S., Barch, D.M., Carter, C.S., Cohen, J.D. (2001). Conflict Monitoring and Cognitive Control. *Psychological Review,* 108, 3, 624-652.

Call, J., & Carpenter, M. (2001). Do apes and children know what they have seen? *Animal Cognition, 4,* 207–220.

Carruthers, P. (2008). Meta-cognition in animals: A skeptical look. *Mind and Language, 23,* 58–89.

Carruthers, P. (2009). How we know our own minds: The relationship between mindreading and metacognition. *Behavioral and Brain Sciences, 32,* 121–138.

Conant, R. C., and Ashby, W. R. (1970). 'Every good regulator of a system must be a model of that system', *International Journal of Systems Science*, 1: 89-97.

Couchman, J.J., Coutinho, M.V.C., Beran, M.J., Smith, J.D. (2010). Beyond Stimulus Cues and Reinforcement Signals: A New Approach to Animal Metacognition, *Journal of Comparative Psychology*, 124, 4, 356-368.

Crystal, J. D., & Foote, A. L. (2009_a). Metacognition in animals. *Comparative Cognition and Behavior Reviews, 4,* 1-16.

Crystal, J. D., & Foote, A. L. (2009_b). Metacognition in animals: Trends and Challenges. In *Comparative Cognition and Behavior Reviews, 4,* 54-55.

Del Cul, A., Dehaene, S., Reyes, P., Bravo, E. & Slachevsky, A. (2009). Causal role of prefrontal cortex in the threshold for access to consciousness, Brain (2009) 132 (9): 2531-2540.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist* 34 (1979): 906-911.

Flavell, J. H., Green, F. L., & Flavell, E. R. (1995). Young children's knowledge about thinking. *Monographs of the Society for Research in Child Development, 60* (1, Serial No. 243).

Gergely, G., Nadasdy, Z., Csibra, G. & Biro, S. 1995 Taking the intentional stance at 12 months of age. *Cognition* 56, 165–193.

Gopnik, A. 1993. How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences.* 16, 1, 1-14, 29-113.

Gopnik, A. & Astington, J.W. (1988). Children's understanding of representational change and it relation to the understanding of false belief and the appearance-reality distinction. *Child Development,* 59, 1: 26-37.

Hampton, R.R. (2001). Rhesus monkeys know when they remember. *Proceedings of the National Academy of Science USA*, 98, 9, 5359-5362.

Hampton, R.R. (2001). Rhesus monkeys know when they remember. *Proceedings of the National Academy of Sciences U.S.A*., 98, 5359-5362.

Hampton, R.R. (2009). Multiple demonstrations of metacognition in nonhumans: Converging evidence or multiple mechanisms? *Comparative Cognition and Behavior Reviews*, 4, 17-28.

Hampton, R.R.. (2001). Rhesus monkeys know when they remember. *Proceedings of the National Academy of Sciences U.S.A*., 98, 5359-5362.

Hare, B., Call, J., Agnetta, B. & Tomasello, M., (2000), Chimpanzees know what conspecifics do and do not see, *Animal Behaviour* 59, 771-785.

Hunter, M. , Ames, E. &Koopman, R. (1983). Effects of stimulus complexity and familiarization time on infant preferences for novel and familiar stimuli. *Developmental Psychology*, 19, 3, 338-352.

Kepecs, A., Naoshige, U., Zariwata, H. & Mainen, Z.F. (2008). Neural Correlates, computation

and behavioural impact of decision confidence. *Nature*, 455, 227-231.
Kiani, R. & Shadlen, M.N. (2009). Representation of Confidence associated with a decision by neurons in the parietal cortex. *Science:* 324 (5928, 759-764.
Koenig, M. A. & Echols, C.H. (2003). "Infant's understanding of false labeling events: the referential roles of words and the speakers who use them". *Cognition*, 87: 179-208.
Koriat, A. & Ackerman, R. 2010. Metacognition and mindreading: Judgments of learning for Self and Other during self-paced study. *Consciousness and Cognition*, 19, 1, 251-264.
Koriat, A. & Levy-Sadot. 1999. Processes underlying metacognitive judgments: Information-based and experience-based monitoring of one's own knowledge. In S. Chaiken & Y. Trope (Eds.), *Dual Process Theories in Social Psychology*. New-York: Guilford, 483-502.
Koriat, A., Bjork, R.A., Sheffer, L. & Bar, S.K., (2004). Predicting one's forgetting: the role of experience-based and theory-based processes. *Journal of Experimental Psychology: General*, 133, 4, 643-656.
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.
Koriat, A., (2000). The Feeling of Knowing: some metatheoretical Implications for Consciousness and Control. *Consciousness and Cognition*, 9, 149-171.
Kornell, N., Son, L., & Terrace, H. (2007). Transfer of metacognitive skills and hint seeking in monkeys. *Psychological Science, 18,* 64–71.
Kovács, A.M., Téglás, E & Endress, A.D. 2010. The Social Sense: Susceptibility to Others' Beliefs in Human Infants and Adults**,** *Science*, 330, 6012, 1830-1834.
Loussouarn, A., Gabriel, D. & Proust, J. (2011). Exploring the informational sources of metaperception: The case of Change Blindness Blindness. *Consciousness and Cognition*, 20:1489–1501.
Marazita, J.M. & Merriman, W.E. (2004) . Young children's judgment of whether they know names for objects: the metalinguistic ability it reflects and the processes it involves. *Journal of Memory and Language*, 51, 458-472
Needham, A. & Baillargeon, R. (1993). Intuitions about support in 4.5-month-old infants. *Cognition, 47,* 121–148.
Nelson, T. O., and Narens, L. (1990). 'Metamemory: a theoretical framework and new findings', in T. O. Nelson ed., *Metacognition, Core Readings*, (1992) 117-130.
Nussinson, R. & Koriat, A. (2008). Correcting experience-based judgments : the perseverance of subjective experience in the face of the correction of judgment. *Metacognition Learning*, 3: 159-174.
Onishi, K.H. & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, 308, p. 255-258.
Perner & Aichorn (2008). Theory of mind, language and the temporo-parietal junction mystery. *Trends in Cognitive Sciences,* 12, 4,123-126.
Perner J, Ruffman T. (1995). Episodic memory and autonoetic consciousness: Developmental evidence and a theory of childhood amnesia. Special Issue: Early memory. *Journal of Experimental Child Psychology* **59**(3): 516-548.
Perner, J. , Kloo, D., Stöttinger, E. (2007). Introspection & remembering. *Synthese* 159:253–270.
Perner, J. (1991). *Understanding the representational mind,* Cambridge, MIT Press.
Perner, J. & Lang, B. 1999. Development of theory of mind and executive control. *Trends in Cognitive Sciences.* 3, 9: 337-344.
Perner, J. & Ruffman, T. (2005). Infants' insight into the mind: How deep? *Science*, 308, p.

214-216.
Proust, J. (2010) Metacognition. *Philosophy Compass*, 5, 11, 989-998.
Proust, J. (in print). Mental Acts as natural kinds, in: T. Vierkant (ed.), *Decomposing the Will.* Oxford: Oxford University Press.
Proust, J. (in press). Mental Acts as Natural Kinds, in : A. Clark, J. Kiverstein & T. Vierkant (Eds.), *Decomposing the Will*. Oxford : Oxford University Press.
Proust, J. (2007). Metacognition and metarepresentation: Is a self-directed theory of mind a precondition for metacognition?' *Synthese* 2 : 271-295.
Rolls, E.T., Grabenhorst, F & Deco, G. (2010). Decision-making, Errors, and Confidence in the Brain. *Journal of Neurophysiology*, 104: 2359-2374.
Rounis, E., Maniscalco, B., Rothwell, J.C., Passingham, R.E. & Lau, H. (2010). Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cognitive Neuroscience*, 1, 3: 165-175.
Schneider, W & Lockl, K. (2002). The development of metacognitive knowledge in children and adolescents. In T.J. Perfect & B. Schwartz (Eds.), *Applied Metacognition*. Cambridge: Cambridge University Press, 224-257.
Schneider, W. (2008). "The development of metacognitive knowledge in children and adolescents: Major trends and implications for education". *Mind, Brain and Education*, 2: 114-121.
Schnyer, D.M., Verfaellie, M., Alexander, M.P., LaFleche, G., Nicholls, L. & Kaszniak, A.W. (2004). A role for right medial prefrontal cortex in accurate feeling-of-knowing judgments: evidence from patients with lesions to frontal cortex. *Neuropsychologia,* 42: 957-966.
Schwarz, N. (2002). Situated cognition and the wisdom of feelings: Cognitive tuning. In L. F. Barrett & P. Salovey (Eds.), *The wisdom in feelings: Psychological processes in emotional intelligence* (pp. 144–166). New York: Guilford.
Smith, J. D., Beran, M. J., Couchman, J. J., Coutinho, M. V. C., & Boomer, J. B. (2009). Animal metacognition: Problems and prospects. *Comparative Cognition & Behavior Reviews*, 4, 40-53.
Smith, J. D., Beran, M. J., Redford, J. S., & Washburn, D. A. (2006). Dissociating uncertainty states and reinforcement signals in the comparative study of metacognition. *Journal of Experimental Psychology: General, 135,* 282–297.
Smith, J. D., Schull, J., Strote, J., McGee, K., Egnor, R. & Erb, L. (1995), The uncertain response in the bottlenosed dolphin *Tursiops truncates. Journal of Experimental Psychology: General,* 124, 391-408.
Smith, J. D., Shields, W. E., & Washburn, D. A. (2003). The comparative psychology of uncertainty monitoring and metacognition. *Behavioral and Brain Sciences*, 26, 317-373.
Sodian, B, Thoermer, C. & Dietrich, N. (2006). "Two- to four-year old children differentiation of knowing and guessing in a non-verbal task". *European Journal of Developmental Psychology,* 3: 222-237.
Sodian, B., & Wimmer, H. (1987). Children's understanding of inference as a source of knowledge. *Child Development, 58*, 424–433.
Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate. *Behavioral and Brain Sciences, 23:* 645–665.
Surian, L. & Leslie, A.M. 1999. Competence and performance in false belief understanding: Whittlesea, B.W.A. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 19, 6: 1235-1253.
Unkelbach, C. (2007). Reversing the Truth Effect: learning the interpretation of processing fluency in judgments of truth. *Journal of Experimental Psychology*: *Learning, Memory and Cognition.* 33, 1: 219-230.

Vickers, D. & Lee, M.D. (1998). Dynamic Models of Simple Judgments : I. Properties of a Self-Regulating Accumulator Module. *Nonlinear Dynamics, Psychology and Life Sciences*, 2, 3, 169-194.

Vickers, D. & Lee, M.D. (2000). Dynamic Models of Simple Judgments : II. Properties of a Self-Organizing PAGAN Model for multi-choice tasks. *Nonlinear Dynamics, Psychology and Life Sciences*, 4, 1, 1–31.

Washburn, D.A., Smith, J.Δ., & Shields, W.E. (2006). Rhesus monkeys (*Macaca mulatta*) immediately generalize the uncertain response. Journal of Experimental PSychology : Animal Behavior Processes, 32, 85-89.

Xu, F. (1999). Object individuation and object identity in infancy: the role of spatiotemporal information, object property information, and language. *Acta Psychological; Special Issue: Visual object perception*, 102, 2-3, 113-136.