

Non-human Metacognition

Joëlle Proust

Abstract

There is as yet no consensus about what "metacognition" requires, and in particular, on how essential it is for genuine metacognition to include the declarative form of awareness that humans seem to enjoy when they "think that they think". Some primates, rodents, and pigeons seem to be able to reliably predict or evaluate their own performance in perceptual or memorial tasks. Are these abilities merely based on predicted reward, or do they rely on metarepresentations of first-order states? Psychological data and neuroscientific evidence, however, suggest that neither of these views is correct. Nonhuman metacognition seems to be rather based on activity-dependent information, i.e., on the dynamic aspects of the mind-brain activation that a given task triggers. On the view defended below, non-humans contextually experience their perceptual or memorial dispositions, as humans do, in feeling-based evaluative attitudes called "affordance sensings".

5991 words

Introduction

There is evidence that non-human animals that have not evolved a mindreading capacity, such as macaques and rodents, are nevertheless able to appropriately evaluate their self-confidence level in perceptual and memory tasks. This creates a puzzle because self-knowledge seems to require embedding a representation into another, i.e. *metarepresenting* one's own states, as exemplified in mindreading. Part of the puzzle has to do with disunified terminology. "Meta", it is often claimed, means "being about". "Metacognition", then, is taken to refer to cognition *about* one's own cognition, "thinking about one's own thinking", or, in short, to meta-knowledge. This terminology owes its influence to the early models of the relations between the control and monitoring aspects of metamemory. To philosophers, this acceptance of metacognition is consonant with deeply entrenched views about the exclusively human character of rationality. Caution, however, is needed to disentangle the terminological from the empirical issues: terminology should follow, rather than preempt research. Section 1 will present the evidence. Section 2 will discuss the view that metacognition is self-directed metarepresentation. Section 3 will discuss the "no-metacognition" view, which claims that

animals merely rely on observable stimuli – such as an oscillating behavior – or on anticipated reward to decide what to do. Section 4 will present the "experience-based" accounts, in which metacognition is neither a mere matter of first-order cognition, nor of metarepresentation. It will be proposed that animal metacognition depends on a non-propositional evaluative attitude called affordance-sensing, which is common to human and some nonhuman cognitive systems.

1 - Experimental evidence for nonhuman metacognition

There is now ample comparative evidence that many nonhumans, although unable to read minds, manifest the ability to evaluate what they can perceive or remember as humans do. The experimental paradigms used to elicit metacognitive evaluations include three main types:

1) Tasks requiring the animals to seek information before acting (Call 2010), or obtain it from a helper, at a cost.

2) Tasks allowing the animals to choose to perform or reject cognitive tasks as a function of their difficulty: for example categorizing perceived densities as low or sparse (Smith et al. 2008), or retrieving paired items presented earlier (Hampton 2001).

3) Tasks requiring the animals to wager on their own cognitive decision right after having made it.

Rhesus monkeys (Smith et al. 2010) and rats (Foote and Crystal 2007) were found to be able to evaluate their own perceptual access to stimuli. Rhesus monkeys (Hampton 2001), pigeons (Adams and Santi (2011), chimpanzees and orangutans Suda-King (2008) have been shown to flexibly search for needed information and to reliably monitor their memory; rhesus monkeys have shown an ability to express their retrospective confidence in a response by wagering (Kornell et al. 2007). Crucially, the response patterns in each case are similar to those of humans in the same task (see discussion in Couchman et al. 2012; Beran 2014). Other animal

species, however, have been found to fail on the tasks above, for example capuchin monkeys. (Basile et al. 2015, Beran et al. 2011). The tasks 2 and 3 listed above have been used in behavioral and in neuroscientific studies, where the activation of neural assemblies involved in metacognitive decisions by rhesus monkeys or rodents was tracked over time (Fleming, 2012, Kepecs et al. 2008). Computational modelling has been made possible by precise quantitative data being collected with either behavioral (Smith 2008) or neuroscientific methods (Kepecs and Meinen, 2012). How can the evidence of animal success in the tasks above be explained?

2. Metacognition as meta-knowledge (MK)

2.1 Stronger version

This is a class of views that take metacognition to involve some form or other of metarepresentation. On the richer MK view, metacognitive animals have a disposition to *know that* they themselves are in a given mental state, for example that they are trying to remember whether they have perceived a given stimulus in a prior occasion. "Knowing that", by definition, has propositional content. As a consequence, embedded contents should include concepts of mental states and of oneself as the target knower. Evaluation, in other words, cannot occur without metarepresentation of one's own mental states. Asking oneself whether one remembers something presupposes that one possesses the concept of memory, of remembering (and cognate notions), and, furthermore, the capacity to represent one's own self as endowed with various states and attitudes, each having a specific cognitive content.

Three main arguments have been provided in favour of metacognitive judgments being metarepresentational. First, flexible prospective assessments of uncertainty, in the absence of the primary test stimuli, express a declarative, i.e. an explicit representation of knowledge in a non-human (Hampton, 2003, Smith et al. 2003). Second, the evidence for easy transfer of

retrospective assessments of confidence across tasks suggests that monkeys express their declarative knowledge about their first-order epistemic states (Son et al. 2003). Finally, the similarity of pattern in uncertainty responses in humans and in rhesus monkeys indicates that a metarepresentational account is justified in both cases (Son et al. 2003).

Discussion

With the benefit of hindsight, the arguments offered by MK-strong fail to be convincing. Flexibility in self-evaluation clearly shows that animals rely on a context to decide how to act – independently of the stimulus presented in the first-order task; they are not merely conditioned to act. But it is premature to conclude that metarepresentation explains flexibility. Second, the "reportlikeness" of the response is merely a paradigm effect; animals do not report or declare what they know, they decide how to act. Finally the similarity of performances between nonhuman and human primates does not entail that metarepresentations underly these performances, but merely that a similar kind of information is available in both cases.

A major weakness of the strong MK-view is that, granting that rodents and pigeons, cannot represent second-order attitudes through concepts such as "believing", "knowing", "being uncertain", etc., they should not be able to evaluate *to what extent* they themselves can remember or perceive, which they were found able to do. they cannot attribute beliefs to others or to themselves, and hence, the MK-view predicts that they cannot evaluate their own perception or their own memory. Against this objection, it has been claimed that metacognitive ability is a matter of degree (Smith, 2009). It has also been proposed that success in metacognitive tasks constitute evidence for an animal's ability to form mental concepts *about its own states*, although not about mental states in general. Such a defense, however, is unparsimonious and ad hoc. The proposal that nonhumans might represent "only in their own case" that they believe, perceive, and so on, infringes the Generality Constraint as applied to what it is to possess the concept of a mental state (Evans, 1982). The MK definition of

metacognition, in its strong form, is therefore difficult to reconcile with current evidence. If non-mindreaders such as rhesus monkeys can evaluate their memory, then their metamemory can be exercised with no concept of memory. Unless, obviously, the studies reported above have conflated metacognition and mere cognition (see section 3)

2.2. MK as Introspection

An alternative MK view has been proposed: even though the animals successfully tested for metacognition cannot read their own minds the way humans do, they might *still have conscious self-reflective access to their uncertainty, and hence qualify as genuinely metacognitive* (Son and Kornell 2005; Metcalfe and Kober 2005). Self-reflectiveness is a particular way of being conscious: not merely having conscious experiences of the world (such as seeming to perceive red objects), but having conscious experiences of oneself as having these experiences (such as feeling uncertain of being able to discriminate red objects).

This view, however, still retains the gist of MK: a metacognitive ability is constituted by a meta-level reflecting on the representational content of an object-level. Metacognition is a form of introspection – a mechanism akin to sensory perception, but directed inward. For example, animals might know what they know (or don't know) by glancing at the content of their memory, as hypothesized in the human case by Josef T. Hart (Hart 1965). In contrast with the preceding view, however, awareness of a *nonconceptual thought content* (an animal, for example, being aware of its "memory strength") is seen as sufficient to adaptively guide behavior. You no more need to know what memory is to gain access to its content than you need to know what hunger is to feel hungry. Metacognition could, then, qualify as a primary form of self-awareness (Metcalfe and Kober 2005).

Discussion

First, note the conceptual difficulty of this view. Reflective self-consciousness is supposed to

embed a first-order thought (for example: a memory state "glanced at") in a second order thought ("I have this memory state"). The view implies that the animals, lacking the appropriate concepts, can still non-conceptually metarepresent their own internal impressions and feelings. But what can it mean to metarepresent a thought, when the state so metarepresented is non-propositional and when the metarepresentation includes no concepts for categorizing its relation with the embedded content? (For an extensive discussion, see Proust 2013 : 142-144).

A second problem for the view is that, in humans, introspection can, at best, be claimed to access one's own sensory states and emotions. The propositional contents of one's own attitudes, in contrast, are not directly perceivable by an inner eye (Carruthers 2011). On this view, inferring the contents of one's attitudes requires conceptual interpretations of what one believes, desires or intends to do, as well as the directing of mindreading towards oneself, in order to know what type of attitude is activated. Hence, nonhumans cannot infer what they believe or are uncertain about.

Another objection to the intermediate MK model, however, points to an alternative account of the informational source of metacognition, which will be discussed in section 4. Human agents cannot know what they remember (or don't remember) by introspecting the content of their memory, as was first hypothesized by Hart (Koriat and Goldsmith 1996). They rather use heuristic strategies based on predictive, activity-dependent cues, in order to predict their ability to perceive or remember. These heuristics elicit noetic feelings that guide cognitive decisions (Koriat 1993). On this view, introspection to inner thought contents is no longer required, while the notion that nonconceptual cues may be used to monitor epistemic success is fully endorsed.

Some comparative psychologists, however, have attempted to exercise parsimony by reducing metacognition to cognition. They claimed that the information that animals use when performing tasks qualified as "metacognitive" are of a behavioral nature, and that the

corresponding abilities actually reflected reinforcement and sensitivity to payoff rather than a sensitivity to subjective uncertainty. We now turn to this family of accounts, the "no-metacognition" views.

3. The no-metacognition views

3.1 Associative accounts (AA)

In stark contrast with the MK views, some theorists have claimed that the relevant information, in tasks (1)-(3), is exclusively behavioral: opting out, and the other tasks reviewed can be solved on the basis of operant conditioning.¹⁶ (Crystal and Foote, 2009; Le Pelley, 2014). An animal's willingness to opt out from a cognitive task, from this viewpoint, *depends on a state of the world represented as worth producing*, rather than on an internal evaluation of the agent's own uncertainty or on an emotional appraisal. This interpretation has motivated computer simulations able to model relevant correlations between animals' decisions and low-level cues, where the latter should only consist in observable properties, such as time elapsed, oscillatory behavior and associated reward. A major objection that AA has raised to non-parsimonious accounts is that opt-out experiments do not allow a clear distinction to be made between the animal's decision - performing the task or opting out - and the confidence report. It has also been noted that confidence is indistinguishable from variables such as attention and vigilance, which can be conditioned and influence performance. One may summarize AA, in terms of the following four claims:

- 1) The information involved in cognitive monitoring by animals is behavioral.
- 2) Opting out is a cognitive (rather than metacognitive) decision based on stimulus features and payoff.

¹⁶ In operant conditioning, an action that turns out to have positive consequences for the agent tends to be reinforced, i.e. reproduced.

- 3) Pay-off is what motivates decisions.
- 4) The form of learning involved is operant conditioning through past trials.

In response to AA claim 3, experimenters suppressed direct reinforcement from the opt-out task paradigm; in spite of lacking information about reward, animals maintained their ability to opt out reliably (Smith et al. 2006). Another way of circumventing an associative account was to compare free-choice with forced-choice decisions for the same stimuli: a systematic increase in successful freely chosen trials versus forced ones suggests that endogenous cues were involved in guiding decisions (Hampton 2001). A third tack consisted in showing that computer simulations based on AA claim 2 were actually unable to track animals' monitoring-based response patterns (Smith et al. 2008 ; 2014).

Another way of addressing AA, however, is to directly question its first clause by combining a computational method with an analysis of the brain correlates of perceptual decisions under uncertainty in animals. The most compelling study uses a post-decision report of confidence, which allows the collecting, in contrast with an opt-out paradigm, of both an answer to the first-order cognitive task *and* a confidence evaluation in each trial (Kepecs and Mainen, 2012). The authors are able to demonstrate that confidence level plays a role in decision-making that is independent of reward. Correct and erroneous choices for the same stimulus are found to systematically vary in their confidence levels. Single cell recordings in other studies make it clear that the kind of cues that rats are using to evaluate their own uncertainty include late onset, intensity (firing rates are higher at chance performance) and the coherence of the neural activity currently triggered by a first-order cognitive task (Kiani and Shadlen, 2009). These cues, however, *are not related to stimulus information, nor to recent reward history*: they are generated by the current first-order neurocognitive activity. Disclaiming 1 and 3 in turn disposes of AA as articulated above.

3.2. The executive account (EA)

Another version of the no-metacognition view relies on a contrast between action guidance and self-attribution. The tasks (1) - (3) above do not qualify as "*metacognitive*", because they are not about one's own cognition at all. A key issue is the status of noetic feelings. "What is it that one feels bad about, when one feels uncertain?", Carruthers and Ritchie (2012) ask. Is it about the likelihood of a cognitive perceptual or memory task being correct/incorrect? Is it not, rather, about the likelihood of a reward being missed/obtained? Carruthers and Ritchie find the second hypothesis both more plausible and more parsimonious. Monkeys' epistemic decisions are embedded in a rewarded task, where the monkeys rely on their emotions to select a reward-conducive decision. There are in the same situation as human subjects completing the Iowa Gambling Task, where subjects have to select cards from four decks with different probabilities for winning or losing various amounts of money (Bechara et al. 1994). The point is that selecting the right deck requires extracting a subpersonal heuristic for selecting the "good" decks of cards, that produces steady gains in the long run. Similarly, the authors claim, feelings of uncertainty are directed at the primary actions open to them, rather than at their own mental states. "It is the performance of the action that seems bad, not the fact that one is thinking about it" (Carruthers and Ritchie, 2012: p. 82). There is nothing specifically metacognitive about the feelings that guide animals' decisions in the tasks described above.

Discussion

EA contains a grain of truth: given that metacognition includes a control dimension, it partly depends on executive capacities – i.e. the abilities involved in selecting a behavior as a function of one's goal, in inhibiting it, shifting it, and updating it. But it also includes a monitoring dimension, whose function is not to select, but to predict epistemic outcome. Concentrating on the monitoring part reveals that there is more to animal and human metacognition than appetitive control. In metacognition, incentive gradient (expected reward or penalty) affects the

amount of effort expended, hence the success of the outcome. Agents, accordingly, tend to scale up their confidence level as a function of their own invested effort (Koriat et al. 2014). This "goal-driven" dimension of metacognitive evaluation, however, is complemented by a "data-driven" dimension, namely the feedback from the task being performed. This dimension is reflected in the remarkable *comparative* stability of evaluations across incentives: Agents assess similarly their *relative* confidence across stakes. In other words, even though confidence is enhanced by added time and effort, relative uncertainty across trials remains similarly influenced by subjectively felt difficulty at each incentive level.

This behavioral finding in humans is compatible with the view that animals use two separate subsystems for evaluating respectively subjective uncertainty and reward (see also Kornell et al., 2007). In an experiment by Zakrzewski et al. (2014), monkeys have proved able to avoid risky trials (where, if there is failure, all the accumulated food tokens can be lost) *only when* they feel unsure that they will correctly carry them out. Higher accumulated numbers of tokens make the monkeys more conservative about performing the primary task. Hence, conservativeness of the response criterion (knowing when to perform) is modulated *both* by the payoff schedule and by the confidence level. Behavioral evidence from comparative studies, then, confirm the neural-computational argument offered at the end of the last section. EA, as a result, is not the most parsimonious hypothesis for explaining animal evidence: it does not account for the combination of epistemic sensitivity and reward appetite in animals.

4. Evaluative-associative accounts of animal metacognition

An alternative theory hypothesizes that the information that is crucial for metacognition lies in the vehicle of thought rather than its content, and that this information is meant to guide cognitive actions, rather than world-directed action. This insight is at the core of a set of views

that we will call "Evaluative-associative accounts" (EAA). Four claims are common to these accounts:

- 1) The information involved in cognitive monitoring by animals is endogenous rather than behavioral: it consists in activity-dependent information generated in a current cognitive task.
- 2) The information that predicts success/failure in a cognitive task is carried by the vehicle of mental activity (e.g. the dynamic features of the firing rates), not by its content.
- 3) A noetic feeling associated with the subpersonal heuristics motivates an epistemic decision to act.
- 4) Learning consists in a) extracting implicit decisional cues and b) calibrating decision thresholds over time through reinforcement.

4.1 Experience-based (EB) metacognition

An analysis involving these four claims can be extracted from Asher Koriat's characterization of human metacognition and applied to animal performances. In "experience-based" metacognition, agents form metacognitive predictions on the basis of a variety of associative cues elicited by the current cognitive activity and assembled in subpersonal heuristics [claim 1]. These cues are "structural" in the sense that they belong to the activity elicited by a task, not to its content [Claim 2]. Recognizing the role of nonconceptual information in epistemic decision-making is a remarkable step, first established through behavioral experiments, and now supported by neuroscientific evidence. A conscious noetic feeling is generated by the discrepancy between stored and observed values of the heuristic standard (Koriat, 2000). This conscious feeling in turn motivates and flexibly controls the decision to act [Claim 3]. The expression of 'epistemic decision' is justified, to the extent that the cues used in prediction are

those that, in the past, have had "diagnostic" value (namely those that carried accurate information about future success or failure) (Koriat, 2012: 227) [claims 2 and 4]. The valence and intensity of noetic feelings is the most parsimonious way of explaining why agents so readily and flexibly rely on subpersonal heuristics to guide their decisions. Feelings, however, are "noetic" in the sense that their valence is not reducible to their reward value.

In Koriat's model, humans additionally enjoy "concept-based" metacognition, through which, for example, they apply to their performances what they believe to be the case concerning their own cognitive abilities. Experience-based metacognition, however, does not depend on such beliefs, and can thus be operational in agents that do not have the conceptual equipment for reading minds, such as animals and young children.

In summary: A prominent feature of EB models is that they are *evaluative* rather than declarative. An evaluative attitude represents a gradient, which adjusts epistemic decision to the evidence accumulated in its favor. This feature suggests a continuity between nonhumans and humans through a non-linguistic, nonconceptual informational route based on the heuristics cum feelings elicited by a cognitive task. An additional argument in favor of this continuity is that the cues assembled in the evaluative heuristics cannot be reported by human agents,²⁸ which suggests that they have been acquired through *implicit* learning, rather than by self-description (by "know-how" rather than by "know-that"). Animal and human implicit learning systems, then, use Bayesian predictive processing to extract heuristics from the associations between prior cues and decision feedback.²⁹

4.2 Animal metacognition relies on cognitive affordance-sensings (AS)

²⁸ See Koriat and Ackerman, (2010).

²⁹ Such learning ability does not build up metacognitive heuristics in every animal species. While some species, such as rhesus monkeys, have high predictive needs associated with non-cooperative foraging, others, such as capuchin monkeys do not, because for them food is plentiful and foraging is cooperative.

This section presents a philosophical conception of the representational format of metacognition compatible with the evidence reviewed above. It is based on an architecture of the mind in which nonconceptual, emotional information plays a crucial evaluative role. The type of evaluative attitude – called "affordance-sensing" (AS) – that will now be defended fulfils the four functional characterizations listed above which, taken together, constitute experience-based metacognition. Its semantics, its scope, its structure, its functions will be briefly described.

1) The semantics of Affordance Sensings

The predictive function of interest for animal metacognition is exercised in a non-propositional attitude, associated with cognitive actions, – actions whose goal is to acquire or retrieve information–, called "affordance-sensing" (AS). AS applies in other domains beyond metacognition whenever an opportunity (either positive or negative) is detected, estimated, and used or rejected in taking action. Under a variety of names,³¹ affordance-sensings have been claimed by philosophers to form a specific set of mental states with an associative structure, available to humans as well as nonhumans. Affordance-sensings can be described as "pushmi-pullyu" attitudes, as defined by Ruth Millikan (Millikan (1995)). Granting, however, that the conditions of correctness for belief and desire are incompatible, the semantics of AS is not constituted by subject-predicate propositions, but by properties assessed from a subjective viewpoint, and endowed with a gradient of valence and intensity. As the following examples illustrate, world-directed and cognitive affordance-sensings have a similar relational structure; they engage a subjective experience and a motivation to act on it; they depend both on

³¹ Bermudez's frames (2009), Cussins' NASAS, (2012), Dreyfus and Kelly's affordance sensings, (2007), Gendler's aliefs, (2008), Griffiths and Scarantino's emotional representations (2009), and Nanay's pragmatic representations, (2013).

subjective needs and perceived opportunities, rather than on objective (subject-independent) world properties:

1. When hungry, an animal perceives *this* food (or this food smell) as an affordance of a given valence and intensity.
2. When trying to jump between two somewhat distant rocks, an animal must evaluate the valence and intensity of this jumping affordance (whether it feels safe to jump that far), given its present motivation and effort readiness.
3. When trying to categorize this display (as dense or sparse), or to remember whether this icon was perceived earlier, an animal must evaluate its present perceptual or memorial affordance (does this task look achievable to me?).

Feelings, i.e. AS, optimally represent these pairings because they make a property salient from a specific subjective viewpoint (e.g., [food for me to eat]). In addition, they provide relevant directional information about how to act.

2) Structure

Having an essential relation to an occurrent event or affordance, the semantic structure of AS is indexical: [Here is this substance for me!] The content so indexed is an *occurrent* (relational) affordance, rather than an individual event or object. Here is how this indexical meaning decomposes:

{Affordance_a [Place=here], [Time= Now/soon], [Valence_{a=+}], [Intensity_{a=.8} (comparatively specified on a scale 0 to 1)], [motivation to act of degree_d according to action program_a] }

The subscript "a" is meant to indicate that all the elements having this subscript characterize the type of affordance as what it contextually appears to be.

3) The scope of Affordance Sensings

Similarity of structure in AS does not entail that affordances are equally detected and relied upon across domains of interest. Sensing affordances is a matter of the opportunities that a lifestyle makes salient to agents. Based on the dual system discussed in section 3, it is justified to distinguish *Appetitive* AS, which are elicited by bodily needs and opportunities, from *Metacognitive* AS, which are elicited by cognitive needs and opportunities. In contrast to appetitive AS, Metacognitive AS do not involve a specification of place. They are elicited by the activity, but they neither coincide with the appetitive experience driving the cognitive motivation itself (such as a food affordance when trying to remember where the food is), nor metarepresent the mental states associated with it.

4) Function: the role of Gradiency

As a graded detector of affordance intensity, affordance sensings depend on metacognitive emotions for providing an *immediate* evaluation of the *degree* of subjective likelihood of a cognitive opportunity, and for immediately activating the selection and execution of an appropriate and timely action (Barrett and Bar 2009, Griffiths and Scarantino, 2009, Proust 2015b). Reactivity in detection and guidance explains why emotions have been recruited in forming evaluative representations. This also explains why cognitive affordance sensings are involved in monitoring fluency, informativeness and relevance of signals and messages in nonhuman and human communication (Proust 2016).

5) Function: Prediction vs Explanation

An association between cues is learned through reinforcement learning; it allows a form of restricted Bayesian reasoning, i.e., prediction of task success. When monitoring their own cognitive actions, animals can transfer their ability to form evaluations to new tasks of the same type, but cannot generalize their findings beyond present opportunities (inference, in contrast, through concept use, enables theory building and hypothetical reasoning). Contrary to

propositional thought, AS has no combinatorial semantics and no truth conditions. It has conditions of appropriateness, however, related to the actual predictive power that it makes available to an organism. Appropriateness requires integration of affordances, which AS is tailored to do. In particular, a cognitive affordance can be combined with a reward affordance of the distal goal on the basis of their respective weighted gradients of valence and intensity (De Martino et al., 2013). Affordance sensings are a common currency for decision making in a world of conflicting opportunities.

Conclusion

The evaluation-based view explains how metacognitive awareness develops in rhesus monkeys, in rodents, and in human infants in the absence of a mindreading ability. Evaluating a cognitive task as feasible, or a cognitive outcome as satisfactory, depends on an associative, comparative process of affordance-sensing. Affordance sensings allow agents to assess and exploit what appears to them as opportunities when deciding how to act.

Acknowledgements

I am grateful to Richard Carter and to Nicole Hall, for their linguistic help, to Mike Beran, for personal communication about his recent experiments, to Kristin Andrews and Reinaldo Bernal for their critical remarks and useful suggestions.

This article was prepared thanks to an advanced grant of the European Research Council # 269616.

REFERENCES

- Adams, A., and Santi, A. (2011). Pigeons exhibit higher accuracy for chosen memory tests than for forced memory tests in duration matching-to-sample. *Learning and behavior*, 39(1), 1-11. Doi: 10.1007/s13420-010-0001-7
- Barrett, L. F., and Bar, M. (2009). See it with feeling: affective predictions during object perception. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364(1521), 1325-1334. Doi: 10.1098/rstb.2008.0312.
- Basile, B. M., Schroeder, G. R., Brown, E. K., Templer, V. L., & Hampton, R. R. (2015). Evaluation of seven hypotheses for metamemory performance in rhesus monkeys. *Journal of Experimental Psychology: General*, 144(1), 85-102. doi: 10.1037/xge0000031.

- Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, 50(1), 7-15. [http://dx.doi.org/10.1016/0010-0277\(94\)90018-3](http://dx.doi.org/10.1016/0010-0277(94)90018-3)
- Beran, M. J. and Smith, J.D. (2014). The uncertainty response in animal-metacognition researchers, *Journal of Comparative Psychology*, 128(2), 155-159. <http://dx.doi.org/10.1037/a0036564>.
- Bermúdez, J. L. (2009). Mindreading in the animal kingdom. In R.W.Lurz (ed.), *The Philosophy of Animal Minds*, Cambridge: Cambridge University Press. 145-164.
- Call, J. (2010). Do apes know that they could be wrong?. *Animal cognition*, 13(5), 689-700. doi:10.1007/s10071-010-0317-x
- Carruthers, P. (2011). *The Opacity of Mind: an integrative theory of self-knowledge*. OUP Oxford.
- Carruthers, P., and Ritchie, J. B. (2012). The emergence of metacognition: affect and uncertainty in animals. In Beran, M.J., Brandl, J., Perner, J. and Proust, J. (Eds.), *Foundations of Metacognition*, Oxford: Oxford University Press.76-93.
- Couchman, J.J., Beran, M.J., Coutinho, M.V.C., Boomer, J. and Smith, J.D. (2012). Evidence for animal metaminds. In Beran, M.J., Brandl, J., Perner, J. and Proust, J. (eds.), *Foundations Metacognition*, Oxford: Oxford University Press, 21-35.
- Crystal, J. D., and Foote, A. L. (2009). Metacognition in animals: Trends and challenges. *Comparative Cognition and Behavior Reviews*, 4, 54-55. doi: 10.3819/ccbr.2009.40001
- Cussins, A. (2012). Environmental Representation of the Body. *Review of Philosophy and Psychology*, 3(1), 15-32. Doi: 10.1007/s13164-012-0086-3.
- De Martino, B., Fleming, S. M., Garrett, N., and Dolan, R. J. (2013). Confidence in value-based choice. *Nature neuroscience*, 16(1), 105-110. doi:10.1038/nn.3279.
- Dreyfus, H., and Kelly, S. D. (2007). Heterophenomenology: heavy-handed sleight-of-hand. *Phenomenology and the Cognitive Sciences*, 6(1-2), 45-55. Doi:10.1007/s11097-006-9042-y.
- Evans, G. (1982). *The Varieties of Reference*. Oxford: Oxford University Press.
- Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. *Phil. Trans. R. Soc. B*, 367(1594), 1338-1349. Doi: 10.1098/rstb.2011.0417
- Foote, A. L., and Crystal, J. D. (2007). Metacognition in the rat. *Current Biology*, 17(6), 551-555. doi:10.1016/j.cub.2007.01.061
- Gendler, T. S. (2008). Alief and belief. *The Journal of Philosophy*, 105(10), 634-663.
- Griffiths, P. E., and Scarantino, A. (2009). Emotions in the wild: The situated perspective on emotion. In Aydede, M., and Robbins, P. (Eds.). *The Cambridge Handbook of Situated Cognition*. New York, NY: Cambridge University Press, pp. 437-453.
- Hampton, R. R. (2001). Rhesus monkeys know when they remember. *Proceedings of the National Academy of Sciences*, 98(9), 5359-5362. doi: 10.1073/pnas.071600998.
- Hampton, R. R. (2003). Metacognition as evidence for explicit representation in nonhumans. *Behavioral and Brain Sciences*, 26(03), 346-347. Doi: <http://dx.doi.org/10.1017/S0140525X03300081>
- Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of educational psychology*, 56(4), 208-16. <http://dx.doi.org/10.1037/h0022263>
- Kepecs, A., and Mainen, Z. F. (2012). A computational framework for the study of confidence in humans and animals. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 367(1594), 1322-1337. doi: 10.1098/rstb.2012.0037.
- Kepecs, A., Uchida, N., Zariwala, H. A., and Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature*, 455(7210), 227-231. doi:10.1038/nature07200.
- Kiani, R., and Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science*, 324(5928), 759-764. doi: 10.1126/science.1169405

- Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological review*, 100(4), 609-639. doi:10.1037/0033-295X.100.4.609
- Koriat, A. (2000). The feeling of knowing: Some metatheoretical implications for consciousness and control. *Consciousness and cognition*, 9(2), 149-171. <http://dx.doi.org/10.1006/ccog.2000.0433>
- Koriat, A. (2012). The subjective confidence in one's knowledge and judgments: Some metatheoretical considerations. In Beran, M.J., Brandl, J., Perner, J. and Proust, J. (Eds.), *Foundations of Metacognition*, Oxford: 215-233.
- Koriat, A. & Ackermann, R. (2010). Metacognition and Mindreading: Judgments of Learning for Self and Other During Self-Paced Study. *Consciousness and Cognition*, 19, 1: 251-264. <http://dx.doi.org/10.1016/j.concog.2009.12.010>
- Koriat, A., and Goldsmith, M. (1996). Memory metaphors and the real-life/laboratory controversy: Correspondence versus storehouse conceptions of memory. *Behavioral and Brain Sciences*, 19(02), 167-188. Doi:10.1017/S0140525X00042114.
- Koriat, A., and Levy-Sadot, R. (2000). Conscious and unconscious metacognition: A rejoinder. *Consciousness and Cognition*, 9(2), 193-202. doi:10.1006/ccog.2000.0436.
- Koriat, A., Nussinson, R., and Ackerman, R. (2014). Judgments of learning depend on how learners interpret study effort. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(6), 1624-1637. doi:10.1037/xlm0000009
- Kornell, N., Son, L. K., and Terrace, H. S. (2007). Transfer of metacognitive skills and hint seeking in monkeys. *Psychological Science*, 18(1), 64-71. doi: 10.1111/j.1467-9280.2007.01850.x
- Le Pelley, M. E. (2014). Primate polemic: commentary on Smith, Couchman, and Beran. *Journal of Comparative Psychology*, Vol 128(2), 132-134. <http://dx.doi.org/10.1037/a0034227>.
- Metcalf, J., and Kober, H. (2005). Self-reflective consciousness and the projectable self. In H.S.Terrace and J. Metcalfe (Eds.), *The Missing Link in Cognition: Origins of self-reflective Consciousness*, Oxford: Oxford University Press. 57-83.
- Millikan, R. (1995) Pushmi-Pullyu Representations, *Philosophical Perspectives*, 9:185-200.
- Nanay, B. (2013). *Between Perception and Action*. Oxford: Oxford University Press.
- Proust, J. (2013). *The Philosophy of Metacognition. Mental agency and self-awareness*. Oxford: Oxford University Press.
- Proust, J. (2015a). The Representational Structure of Feelings, in T. Metzinger and J.M. Windt, (eds.) *Open Mind*. www.open-mind.net. doi: 10.15502/9783958571044.
- Proust, J. (2015b) Time and action: Impulsivity, habit, strategy? In *The Review of Philosophy and Psychology* 6(4), 717-743. doi10.1007/s13164-014-0224-1.
- Proust, J. (2016). The Evolution of Primate Communication and Metacommunication. *Mind & language*, 31(2), 177-203. doi: 10.1111/mila.12100
- Smith, J. D., Beran, M. J., Couchman, J. J., & Coutinho, M. V. (2008). The comparative study of metacognition: sharper paradigms, safer inferences. *Psychonomic bulletin & review*, 15(4), 679-691. doi:10.3758/PBR.15.4.679
- Smith, J. D., Beran, M. J., Redford, J. S., & Washburn, D. A. (2006). Dissociating uncertainty responses and reinforcement signals in the comparative study of uncertainty monitoring. *Journal of Experimental Psychology: General*, 135(2), 282-97. <http://dx.doi.org/10.1037/0096-3445.135.2.282>.
- Smith, J. D., Redford, J. S., Beran, M. J., and Washburn, D. A. (2010). Rhesus monkeys (*Macaca mulatta*) adaptively monitor uncertainty while multi-tasking. *Animal cognition*, 13(1), 93-101. doi: 10.1007/s10071-009-0249-5.

- Smith, J. D., Couchman, J. J., and Beran, M. J. (2014). Animal metacognition: A tale of two comparative psychologies. *Journal of Comparative Psychology*, 128(2), 115-131. <http://dx.doi.org/10.1037/a0033105>.
- Son, L. K., and Kornell, N. (2005). Meta-confidence judgments in rhesus macaques: Explicit versus implicit mechanisms. In H. S. Terrace and J. Metcalfe (eds.), *The missing link in cognition: Origins of self-reflective consciousness*, 296-320.
- Suda-King, C. (2008). Do orangutans (*Pongo pygmaeus*) know when they do not remember? *Animal cognition*, 11(1), 21-42. doi:10.1007/s10071-007-0082-7.
- Zakrzewski, A. C., Perdue, B. M., Beran, M. J., Church, B. A., and Smith, J. D. (2014). Cashing out: The decisional flexibility of uncertainty responses in rhesus macaques (*Macaca mulatta*) and humans (*Homo sapiens*). *Journal of Experimental Psychology: Animal Learning and Cognition*, 40(4), 490-501. [doi.org/10.1037/xan0000041](http://dx.doi.org/10.1037/xan0000041).

Further readings

Articles in a (2014) issue of the *Journal of Comparative Psychology* by Nate Kornell offer interesting critical views about animal metacognition (128(2), 143-149 and 160-162, <http://dx.doi.org/10.1037/a0033444-a0036194>). For a presentation of the neuroscientific correlates of metacognition in nonhumans and humans, see S.M. Fleming and C.D. Frith (eds.), *The Cognitive neuroscience of Metacognition* (Berlin and Heidelberg: Springer-Verlag, 2016). *Foundations of Metacognition* offers an interdisciplinary discussion of the emergence of metacognition, of its criteria and of its functions (M.J. Beran, J.L. Brandl, J. Perner and J. Proust eds.)

Notes on contributors

Joëlle Proust works at Institut Jean-Nicod, Ecole Normale Supérieure (Paris, France), as a Director of Research for Pierre-Gilles de Gennes Foundation. She has headed several interdisciplinary collaborative projects on animal metacognition and on the development and cross-cultural diversity of human metacognition.