

Two Levels of Metacognition

Santiago Arango-Muñoz

Received: 10 August 2010 / Accepted: 15 August 2010 /
Published online: 23 September 2010
© Springer Science+Business Media B.V. 2010

Abstract Two main theories about metacognition are reviewed, each of which claims to provide a better explanation of this phenomenon, while discrediting the other theory as inappropriate. The paper claims that in order to do justice to the complex phenomenon of metacognition, we must distinguish two levels of this capacity—each having a different structure, a different content and a different function within the cognitive architecture. It will be shown that each of the reviewed theories has been trying to explain only one of the two levels and that, consequently, the conflict between them can be dissolved. The paper characterizes the high-level as a rationalizing level where the subject uses concepts and theories to interpret her own behavior and the low-level as a controlling level where the subject exploits epistemic feelings to adjust her cognitive activities. Finally, the paper explores three kinds of interaction between the levels.

Keywords Metacognition · Metarepresentation · Control · Epistemic feelings · Dual-process theory

Introduction

Current discussions of metacognition have focused on questions like the following: What is the nature of metacognition? What is the function of this mental capacity? What is the content and epistemic status of metacognitive assessments? Which living beings are endowed with it? At the present stage of the discussion, two main theories have been proposed concerning the proper set of answers to these questions. One claims that metacognition is a metarepresentational capacity to *self-ascribe* mental states, whereas the other claims that it is mainly a capacity to *evaluate and control* our cognitive processes via a mental simulation of them.

S. Arango-Muñoz (✉)
Werner Reichardt Center for Integrative Neuroscience (CIN), Eberhard Karls Universität Tübingen,
Paul-Ehrlich-Str. 17, 72076 Tübingen, Germany
e-mail: santiagoarangom@gmail.com

My suggestion is that our metacognitive capacity can be understood as involving two different levels of complexity (as has been originally suggested by Koriat (2000)), each having a different structure, a different content and a different function within the cognitive architecture, and that each of the competing theories has been addressing a different level. Thus, in the end, the two theories can be shown to be compatible since they provide explanations of different levels of metacognition. Moreover, my argument also shows that any theory that aims to explain human metacognition should be able to account for both levels and their interactions.

The paper will proceed as follows: “Two Theories of Metacognition” introduces the two main theories of metacognition together with the empirical data that they cite in their favor. “Two Levels of Metacognition” develops the idea of two levels of metacognition and characterizes each level according to the *dual-process* theories of cognition. Finally, “Interactions, Mechanisms and Advantages of the Two-Level Account” considers some interactions between the two levels and the question of how to individuate the levels.

Two Theories of Metacognition

Metarepresentational Theory of Metacognition

From the point of view of the metarepresentational theory, metacognition refers to “thinking about thinking”, i.e., to the self-ascription of mental states carried out by forming a second order thought about a first order one, and more generally forming an $n+1$ -order thought about an n -order thought. Such self-ascription depends on a more general mindreading capacity consisting in an inferential capacity to attribute mental states in order to interpret and rationalize other people’s behavior. Thus metacognition in this sense is no more than “turning our mindreading capacities upon ourselves” (Carruthers 2009a, p.8; see Larkin 2010; Gopnik 1993; Wegner 2002; Bogdan 2001, 2005; Flavell 2004). In this respect, this view runs together with the Theory-theory of mindreading.

Mindreading requires the possession of mental concepts by the subject in order to apply them to other people and interpret their behavior. Psychological concepts are concepts referring to propositional attitudes¹ such as perceptions, feelings, intentions, knowledge, beliefs and expectations, among others. Therefore, the necessary structure of metacognitive judgments is composed by: 1) a proposition (e.g. “it rains”), 2) a first-order attitude directed to that representation, such as believing or intending, denoted by a *psychological concept*, and 3) a second-order attitude, namely a metacognitive judgment, directed to the first order attitude (2) and its proposition (1) (Proust 2007). In other words, the content of a second-order

¹ It is still a matter of hot debate whether all these mental attitudes are propositional; especially in the cases of perception, emotion and feeling. And it is also unclear if all the metarepresentational theorists hold a propositional view of attitudes. To my knowledge, within the group of metarepresentational theorists, at least Carruthers (2009b) is committed to this view. My own view is that the content of perception, emotion and feelings is nonconceptual and non-propositional, though I am not going to claim so here.

representation is necessarily constituted by the self-attribution of a psychological concept together with a first-order representation:

[3] I believe that [2] I *KNOW* (or *PERCEIVE*, *BELIEVE*, *FEEL*, *ETC*) that [1] it rains.

It should be highlighted then that a necessary condition to form thoughts with this structure is to possess and be able to apply *psychological concepts*. Thus, from an evolutionary point of view, only beings endowed with the capacity for mindreading would be able to form metacognitive judgments. However, empirical studies using the different paradigms have suggested that *stricto sensu* non-human animals lack the mindreading capacity that would allow them to form metarepresentations (Bermúdez 2009; Call and Tomasello 2008; Carruthers 2008; Povinelli and Vonk 2003, 2006) and thus are also incapable of metacognition in this sense. From a developmental point of view, infants are only capable of self-ascriptions of mental attitudes after they have acquired a theory of mind that provides them with mastery of the relevant psychological concepts (Wellman 1990; Baron-Cohen 1995; Gopnik and Melzoff 1997). Most of these theorists acknowledge that metacognition in this sense starts between the ages of 3 and 4, but others hold that it does not occur until the age of 5 or even later (see Bogdan 2001, 2005).²

From an epistemological point of view, there should be *almost* no difference between the knowledge that a subject has about herself and her knowledge about others because both are based on the similar behavioral cues,³ use the same conceptual resources to make inferences and are produced by the same cognitive mechanism. Both kinds of knowledge derive from mindreading and thus have the same epistemic status. For example, the following judgments: “He intends to listen to the teacher” and “I intend to listen to the teacher” are based on the observation of some bodily behavior (as turning the head in some direction) together with contextual factors and other cues, and are produced by the mindreading capacity using the same concepts and therefore have the same epistemic status.

Control Theory of Metacognition

The control view on metacognition claims that it is mainly a capacity to evaluate and control our cognitive processes and mental dispositions by means of mental simulation. In Joëlle Proust’s words: “The aim is, rather, to evaluate one’s present mental dispositions, endorse them, and form epistemic and conative commitments” (Proust 2010). The main point is that this evaluative capacity does not follow from a theoretical capacity to meta-represent attitudes (mindreading), i.e., subjects do not need to form a second-order representation about their first order attitudes in order to evaluate and control them. As Pamela Hieronymi puts it: “The forming and revising

² However, there is a lively debate about the onset of mindreading abilities, with some researchers claiming that infants of 15 months already understand false beliefs (see e.g. Onishi & Baillargeon 2005). I thank an anonymous peer-reviewer for this remark.

³ I say “almost” because Carruthers’ later account of mindreading acknowledges that first person mindreading has access to some cues (such as visual, auditory and motor imagery and inner speech) that the third person mindreading does not have access to. This creates a quantitative difference but not a qualitative one: access in both cases is interpretative and thus its epistemic status remains the same.

of beliefs and intentions is not voluntary nor does it require the same kind of reflective distance or awareness” (Hieronymi 2009).⁴ Thus, it does not require the possession of psychological concepts, a theory of mind or mindreading. Control theorists speculate that such evaluative capacity derives from an off-line simulation of the cognitive process in question which permits predicting and adjusting future cognitive performance on a given task (Proust 2007, 2008, 2009a; Peacocke 2007, 2008, 2009), as happens in the case of bodily actions where the subject runs an off-line motor program to predict her future performance (Grush 2004). In this respect, the view runs together with the Simulation theory of mindreading.

Inspired by the psychological literature on human and non-human metacognition, the control theory holds that the postulated evaluative capacity can be understood as a capacity to monitor and control cognitive activities, such as remembering or perceiving, which is more primitive than mindreading. Empirical data supporting control theories come from the domains of experimental and animal psychology: 1) Different behavioral paradigms support the claim that *very often* humans *do not rely* on metarepresentations to control their cognitive activities (Koriat 2000, Reder 1996; Paynter et al. 2009; Walsh and Anderson 2009). For example, Reder and Schunn (1996) have shown that subjects are able to assess and decide among possible cognitive strategies, e.g., whether they will be able either to remember or calculate a math problem, based on a feeling of knowing produced by the properties of the math problem and the retrieval process but not by the content of the solution. And 2) some non-human animals that lack psychological concepts and mindreading seem nonetheless to be endowed with the capacity to monitor and control their cognitive performance in memory and perception tasks, allowing them to have an accurate performance similar to human behavior in perception and memory (see Smith 2009 for a review). These findings seem to support a distinction between mindreading and metacognition, and also suggest the idea that metacognition is a function exerted by means of a different representational basis in both cases: it does not rely on conceptual representations in order to monitor and control cognitive processes (Proust 2009b).

Two Levels of Metacognition

Theorists of both sides have been accusing each other of misinterpreting the phenomenon and giving the wrong account of metacognition. While metarepresentation theorists accuse control theorists of putting too much weight on a sub-personal mechanism, a “gate-keeping mechanism” (Carruthers 2008, 2009a), control theorists accuse the former of over-intellectualizing a more basic phenomenon that does not need mindreading capacities (Proust 2007, 2009b, c). My suggestion is that we can dissolve this conflict if we analyze our metacognitive capacity as comprising two different levels of complexity—each having a different structure, a different content and a different function within the cognitive architecture (as has been originally suggested by Koriat (2000)). Each of the competing theories has been addressing a

⁴ Even if Hieronymi never speaks of metacognition as such, it seems to me that her theory of mental control can be understood as a control theory of metacognition, regardless of whether or not she uses the term and of whether or not we accept her account.

different level. So, in the end there is no real disagreement between both theories because they are trying to explain different phenomena. It is also worth highlighting that once the two levels are functioning, they interact and influence each other dynamically, although neither of the two theories has taken these interactions into account. Each of these two levels may be associated with one of the two cognitive systems proposed by *dual-process* theories of mind (Evans 2008, 2009; Thompson 2009). The high-level has a metarepresentational structure and it is associated with an analytic theory-based level belonging to system 2, whereas the low-level is a control structure, an experience-based level belonging to system 1.

High-Level: Theory-Based Metacognition

I largely agree with Carruthers' (2009a) characterization of the high-level. From his point of view, the subject self-ascribes some mental states, properties or capacities based on her beliefs about her cognitive capacities and some perceptual or quasi-perceptual cues (such as visual, auditory and motor imagery and inner speech). Thus, this is a self-interpretative level where the subject tries to make sense of the observation of her behavior, in other words to rationalize and justify what she does (Evans and Over 1996; Thompson 2009). In order to do so, the subject needs a) some psychological concepts that permit her to self-attribute psychological states, b) a language in which to formulate her judgments (Bermúdez 2003, 2009; Evans 2009), and c) a theory of mind, understood as a set of beliefs concerning the functioning of the mind and allowing her to make inferences. In short, the subject needs a mindreading capacity to self-ascribe psychological states. These characteristics seem to point to the kind of processing associated with what has been called *system 2* by cognitive psychologists and characterized as being slow, analytic, controlled and conscious (Evans 2008; Thompson 2009). Metarepresentational theories have been trying to explain this level of metacognition.

High-level metacognition thus deals with conceptual content and its main structure is inferential. Subjects interpret their behavior and make inferences thanks to a folk theory of mind they possess (Flavell et al. 1998). Metarepresentational judgments are thus “drawn upon the *content* of domain-specific beliefs and knowledge that are retrieved from memory” (Koriat 2007, p.298). A striking example of the interpretative nature of this level is revealed when subjects in an experiment are given an alternative explanation of the origin of their feelings of familiarity according to which they are unreliable; consequently, this *new theory* prevents them from relying on those feelings. Based on this new theory, subjects will be less likely to self-attribute knowledge or memories and to rely on the fluency of retrieval as a cue, as they did before they were given the *new theory* (Schwarz and Vaughn 2002; Sanna and Schwarz 2003). In the same way, Whittlesea et al. (1990) found that when subjects were told that fluency of perception was actually being manipulated, perceptual fluency no longer felt like familiarity, but instead felt like current perceptual clarity and thus it no longer lead to illusions of familiarity, memory or knowledge.

Given that the main cognitive function of high-level metacognition is interpretive, it follows that high-level metacognition should have co-evolved with mindreading (Carruthers 2009a; Bogdan 2001, 2005; Jacob 2005) and subjects may often be

wrong in their self-interpretative judgments about their own propositional attitudes and cognitive capacities: “People will (falsely) confabulate attributions of judgments and decisions to themselves in a wide range of circumstances, while being under the impression that they are introspecting” (Carruthers 2009a, p.10). On the one hand, people can be easily deluded concerning the content of their memory since “recognition or direct questioning can have ‘contaminating’ effects on memory” (Loftus et al. 1996). On the other hand, people seem to hold false theories about their memory or their perception. For example, they often think that their visual field is like a TV screen or that their memory is like a hard disk.

What is striking, however, is that subjects often do not rely on such theories or confabulations to control their cognitive behaviour. In other words, what they believe they do and what they actually do are not consistent. They do not behave as if their memory was perfect; they often make little mnemonic notes and consult their notebooks when they feel uncertain. They do not behave as if their perceptual field was a TV screen; they constantly scan the visual scenes in order to grasp all the relevant details. Koriat and Ackerman’s (2010) recent study provides an interesting example of this inconsistency: subjects were presented with a learning task and they had to make judgements of learning (JOL) concerning how well they had learned the items. Their overall behaviour was based on the *implicit* heuristic that the more study time they invest in an item, the less likely they judge that they are to recall it later. However, when they were *explicitly* asked about the basis of their judgment, they reported different, inconsistent stories. These behavioural facts give rise to the idea that normal behaviour is not driven by high-level metacognition and metarepresentational beliefs (though it can sometimes be driven by it, as I show below in “Interactions Between the Two Levels”) but by low-level metacognition and epistemic feelings.⁵ Along the same lines, many experiments on reasoning have pointed out that many of the reasoning biases, such as the belief bias or the myside bias, are actually caused by the subjects’ propensity to accept uncritically (i.e., without an analysis or revision by S2) a heuristic response (Thompson 2009).

Low-Level: Experience-Based Metacognition

Control theories focus on actual behavior rather than the judgments that the subject forms about her mental events, dispositions and capacities. Proponents of this view have remarked that our cognitive behavior is often caused not by reflexive thinking and second-order thoughts but by emotional states (de Sousa 2008; Berkowitz 2000; Carver and Scheier 1998). For example, a subject confronted with a cognitive problem such as a multiplication task has to select a cognitive strategy to solve it. She has to choose either to remember the answer (in the case of a familiar problem) or to calculate. This decision does not seem to be based on a reflexive process considering all the possible alternatives and the pros and cons of each one (a maximizing and metarepresentational process), but on a feeling that affords or makes salient one of the possible strategies (Walsh and Anderson 2009; Paynter et al. 2009;

⁵ This might seem to commit me to Evans’ (2008) and Carruthers’ (2009b) thesis that behaviour is *always* driven by system 1. However, my claim is that it is so driven when system 1 can cope with the situation or problem; otherwise system 2 is activated, as Thompson claims (2009).

Kahneman 2003). Feelings are one kind of output of what cognitive psychologists have called *System 1* and which has been characterized as being fast, based on heuristics, mostly automatic and unconscious (Evans 2008).

But how is *system 1* supposed to generate such feelings? A possible solution is to postulate that metacognition is a “criticism system” or set of criticism systems provided with a list of heuristics in the form of conditionals prescribing the production of a given epistemic feeling for a given situation: if P-event then Q-feeling, if R-event then S-feeling, and so on. For example, in the case of memory, the criticism system would diagnose a good performance by a feeling of familiarity if the stimulus is perceptually fluent⁶ (Whittlesea 1993; Whittlesea and Williams 2001). The feeling itself is metacognitive in the sense of being directed towards a mental disposition (knowledge, uncertainty, ignorance, etc.), but the content of the epistemic feeling that determines decision-making is non-conceptual and thus not metarepresentational. The feeling points or is directed to a mental property which is not necessarily within the gaze of the subject. For instance, a feeling of uncertainty points to a lack of knowledge or indicates that something is wrong with our perceptual or mnemonic activity, allowing a subsequent correction or improvement, without the need for an introspective effort by the subject. In a nutshell: low-level metacognition is the capacity of a being a) to entertain epistemic feelings that *nonconceptually* point to mental dispositions and b) to be able to exploit such feelings in order to control its cognitive activities (Proust 2009a).

Interactions, Mechanisms and Advantages of the Two-Level Account

Interactions Between the Two Levels

If my analysis is right, then the two competing theories of metacognition are indeed trying to explain two different levels of this metacognitive capacity instead of a single phenomenon. The high-level addressed by the metarepresentational theory is a rationalizing level where the subject uses concepts and theories to interpret her own behavior. The low-level is by default a controlling level where feelings induce the subject to adjust her cognitive activities in different ways without the need to engaging in second order thought. It obviously does not follow that they are totally independent mechanisms and that they cannot affect each other. Let us consider three possible interactions:

- 1) On the one hand, one may have something like bottom-up causation: a feeling of error or uncertainty, for instance, elicited by the low-level may trigger an inferential process of verification carried out by the high-level (Thompson 2009). What is interesting in these cases is that feelings seem to emerge without the need for any higher-order belief concerning the cognitive processes and

⁶ Fluency is a property of the way a cognitive system treats a piece of information. A piece of information is more or less fluent according to the easiness or difficulty of the processing of the signal. For example, adding some noise to a perceptual signal makes perceptual processing less fluent. (Whittlesea 1993; Whittlesea et al. 1990; Whittlesea and Williams 2001)

come to indicate a way of acting. The tip-of-the-tongue phenomenon is a good example. You unsuccessfully try to recall some piece of information together with the unpleasant feeling that you are in possession of such information. Notice that the higher-order belief that the unsuccessful retrieval should elicit is a negative one (“I don’t know the information”), while the feeling points in the opposite direction. Then you might persist trying to recall it using your best known mnemonic strategies, or give up your attempt based on the metacognitive belief that when this happens it is better to distract yourself while waiting for the information to come to mind spontaneously.

- 2) On the other hand, one may have something like top-down causation in the following cases: making some concept salient, such as the concept of forgetting, may facilitate the production of some experiences in the subject, such as the feeling of forgetting something that would not be elicited otherwise (Koriat et al. 2004). This phenomenon suggests that epistemic feelings are not cognitively impenetrable, as perceptual experience seems to be (e.g., in the classic case of the Müller-Lyer illusion), that having some beliefs may trigger some particular experience concerning your cognitive activity. For example, thinking about the fallibility and unreliability of your memory may trigger a strong feeling of uncertainty that might even interfere with your normal cognitive performance (Pieschl 2010).
- 3) Another possibility is a top-down effect causing an inhibition of the low-level: when, e.g., the second-order belief that you are going to make a mistake activates the high-level as the controlling level and then inhibits low-level responses (Thompson 2009). The same case occurs in the example cited above, when subjects inhibit their propensity to rely on the feeling of familiarity after they are given a *theory* concerning the unreliability of the feeling (Schwarz and Vaughn 2002; Sanna and Schwarz 2003).

One or Two Mechanisms?

These considerations suggest a dichotomy: Either (A) these two levels are parts of the same mechanism or (B) they constitute two (or several) different mechanisms? To answer this question, we should establish an individuation criterion for cognitive mechanisms. Normally, the criterion used in cognitive science is to individuate a mental mechanism according to the mental operations that it carries out and its cognitive function (Bechtel 2008). The first possibility (A) would imply that high-level metacognition (mindreading) is grounded on low-level metacognition, as some of the simulation theorists hold (Peacocke 2007, 2009; Goldman 1993, 2006). And then we should expect a parallelism between judgments concerning the self and the others.⁷ This seems unlikely given the differences in function, content and structure that have been analyzed so far. Moreover, recent research in cognitive psychology has provided evidence of differences in the bases of metacognitive judgments about self and others (Koriat and Ackerman 2010).

⁷ Carruthers’ metarepresentational view of metacognition in terms of mindreading would belong to option (A) since it presuppose the existence of only one mechanism, but it cannot be considered to be in competition with the other accounts since it begins by rejecting the existence of low-level metacognition.

The second possibility is that both levels are distinct mechanisms that have evolved in virtue of different evolutionary pressures to carry out different cognitive functions (Nichols and Stich 2003; see also Carver 2003; Evans 2008; Thompson 2009). This seems a more plausible option given the deep differences in structure, content and cognitive function that have just been described. However, even if I agree with Nichols and Stich (2003) in their claim that there are at least two different mechanisms, I deeply disagree with them in the main function they attribute to low-level metacognition. Whereas for them its main function is to provide introspective self-knowledge of our propositional attitudes, for me its main function is rather to control cognitive activities without the need for metarepresentation or introspection. In evolutionary terms, we can presume that the low-level mechanisms evolved in order to monitor and control our own cognitive activities and thus improve their functioning, whereas the high-level mechanism evolved mainly to interpret others' behavior, only afterwards being turned towards the self.

Some Advantages of the Two-Level Account

Some of the advantages of conceiving metacognition as involving two levels are the following:

- a) By accepting and explaining low-level metacognition in terms of experiences and heuristics, we do justice to the empirical findings on animal and infant metacognition that demonstrate the presence of this capacity in non-linguistic or mindreading beings. Thus, we can avoid the counterintuitive consequence of the metarepresentational view that we have to endow animals with a psychological concepts and theory of mind (see Gennaro 2009; DeGrazia 2009). This also allows us to understand that there is a low level of mental self-control, a *primitive mental agency*, that is non-conceptual, non-reflexive in terms of high-order thought and that does not require an inferential capacity. This nicely draws a line of continuity between human and non-human cognition: it is not the case that we humans have a *sui generis* form of cognition, but some of the capacities already present in animals became much more complex in the human mind.
- b) By accepting and explaining high-level metacognition in conceptual and metarepresentational terms, we do justice to the empirical findings that show that only beings that possess and are able to apply psychological concepts are able to self-attribute mental states. In other words, high-level metacognition is consistent with the developmental symmetry of self/other psychological ascriptions. It also allows us to explain the confabulation data as the result of self-interpretation based on a folk theory of mind.
- c) A key question for my account is why we should consider these as two levels of the same capacity and not just two different and independent phenomena. My answer to this question is that actually they have evolved as different and independent mechanisms (or sets of mechanisms) to carry out different cognitive functions, but from the moment in which they start to interact and influence each other, as I have sketched in "[Interactions Between the Two Levels](#)", we have to consider both as forming one complex capacity to evaluate and self-ascribe psychological properties (from the low-to-high level) and self-ascribe and

control (from the high-to-low level), that would not have emerged in the absence of either of the two levels.

Conclusion

If my argument is right, the two competing theories are trying to explain two different levels of metacognition instead of a single phenomenon, and therefore the conflict is resolved. The high-level is a rationalizing level where the subject uses concepts and theories to interpret her own behavior. The low-level is a controlling level where the subject exploits epistemic feelings to adjust her cognitive activities. As I have tried to show, each has a different structure, a different content and a different function in the cognitive architecture. Moreover, my argument also shows that any theory that aims to explain human metacognition should be able to account for both levels and their interactions in order to produce a satisfactory account of this capacity.

Acknowledgments I am particularly indebted to Fabián Bernache, Jerome Dokic, Joëlle Proust, Kirk Michaelian, Tobias Schlicht and three anonymous peer-reviewers for their thorough comments and corrections to this paper. I also received valuable comments from David Papineau, Albert Newen, Reinaldo Bernal and David Fajardo. This work was supported by the Werner Reichardt Center for Integrative Neuroscience (CIN), Tübingen, Germany.

References

- Baron-Cohen, S. (1995). *Mindblindness. An essay on autism and theory of mind*. Cambridge: MIT.
- Bechtel, W. (2008). *Mental mechanisms, philosophical perspectives in cognitive science*. New York: Taylor & Francis Group.
- Berkowitz, L. (2000). *Causes and consequences of feelings*. Cambridge: Cambridge University Press.
- Bermúdez, J. L. (2003). *Thinking without words*. Oxford: Oxford University Press.
- Bermúdez, J. L. (2009). Mindreading in the animal kingdom. In R. Lurz (Ed.), *The philosophy of animal minds* (pp. 145–164). Cambridge: Cambridge University Press.
- Bogdan, R. J. (2001). Developing mental abilities by representing intentionality. *Synthese*, 129(2), 233–258.
- Bogdan, R. J. (2005). Why self-ascriptions are difficult and develop late? In B. F. Malle & S. D. Hodges (Eds.), *Other minds. How humans bridge the divide between self and the others* (pp. 190–206). New York: Guilford.
- Call, J., & Tomasello, M. (2008). Does the Chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Science*, 12, 187–231.
- Carruthers, P. (2008). Meta-cognition in animals: a sceptical look. *Mind and language*, 23, 58–89.
- Carruthers, P. (2009a). How we know our own minds: the relationship between mindreading and metacognition. *Behavioral and Brain Sciences*, 32, 1–18.
- Carruthers, P. (2009b). Invertebrate concepts confront the generality constraint (and win). In R. Lurz (Ed.), *The philosophy of animal minds* (pp. 89–107). Cambridge: Cambridge University Press.
- Carver, C. S. (2003). Pleasure as a sign you can attend to something else: Placing positive feelings within a general model of affect. *Cognition and Emotion*, 17(2), 241–261.
- Carver, C. S., & Scheier, M. F. (1998). *On the self-regulation of behaviour*. New York: Cambridge University Press.
- DeGrazia, D. (2009). Self-awareness in animals. In R. Lurz (Ed.), *The philosophy of animal minds* (pp. 201–217). Cambridge: Cambridge University Press.
- de Sousa, R. (2008). Epistemic feelings. In G. Brun, U. Doğuoğlu, & D. Kuenzle (Eds.), *Epistemology and emotions* (pp. 185–204). Hampshire: Ashgate Publishing Limited.

- Evans, J St B T. (2008). Dual-processing accounts of reasoning, judgment and social cognition. *Annual Review of Psychology*, 59, 255–278.
- Evans, J St B T. (2009). How many dual-process theories do we need: One, two or many? In J St B T Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 33–54). Oxford: Oxford University Press.
- Evans, J St B T, & Over, D. E. (1996). *Rationality and reasoning*. Hove: Psychology.
- Flavell, J. H. (2004). Theory-of-mind development: Retrospect and prospect. *Merrill-Palmer Quarterly*, 50, 274–290.
- Flavell, J. H., Green, F. L., & Flavell, E. R. (1998). The mind has a mind of its own: developing knowledge about mental uncontrollability. *Cognitive Development*, 13, 127–138.
- Gennaro, R. (2009). Animals, consciousness, and I-thoughts. In R. Lurz (Ed.), *The philosophy of animal minds* (pp. 184–200). Cambridge: Cambridge University Press.
- Goldman, A. (1993). The psychology of folk psychology. *Behavioral and Brain Sciences*, 16, 15–28.
- Goldman, A. (2006). *Simulating minds: the philosophy, psychology and neuroscience of mind-reading*. Oxford: Oxford University Press.
- Gopnik, A. (1993). How we know our minds: the illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences*, 16(1–15), 90–101.
- Gopnik, A., & Melzoff, A. N. (1997). *Words, thoughts, and theories*. Cambridge: MIT.
- Grush, R. (2004). The emulation theory of representation: motor control, imagery, and perception. *Behavioral and Brain Sciences*, 27, 377–442.
- Hieronymi, P. (2009). Two kinds of agency. In L. O'Brien & M. Soteriou (Eds.), *Mental actions and agency* (pp. 138–162). Oxford: Oxford University Press.
- Jacob, P. (2005). First-person and third-person mindreading. In P. Giampieri-Deutsch (Ed.), *Psychoanalysis as an empirical, interdisciplinary science*. Vienna: Austrian Academy of Sciences Press.
- Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American Psychologist*, 58, 697–720.
- Koriat, A. (2000). The feeling of knowing: some metatheoretical implications for consciousness and control. *Consciousness and Cognition*, 9, 149–171.
- Koriat, A. (2007). Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch & E. Thompson (Eds.), *Cambridge handbook of consciousness*. New York, USA: Cambridge University Press.
- Koriat, A., & Ackerman, R. (2010). Metacognition and mindreading: judgments of learning for self and other during self-paced study. *Consciousness and Cognition*. doi:10.1016/j.concog.2009.12.010.
- Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: the role of experience-based and theory-based processes. *Journal of Experimental Psychology: General*, 133, 643–656.
- Larkin, S. (2010). *Metacognition in young children*. New York: Routledge.
- Loftus, E. F., Coan, J. A., & Pickrell, J. E. (1996). Manufacturing false memories using bits of reality. In L. Reder (Ed.), *Implicit memory and metacognition*. New Jersey: LEA.
- Nichols, S., & Stich, S. P. (2003). *Mindreading: An integrated account of pretence, self-awareness, and understanding other minds*. Oxford: Oxford University Press.
- Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, 308 (5719), 214. doi:10.1126/science.1111656.
- Paynter, C. A., Reder, L., & Kieffaber, P. D. (2009). Knowing we know before we know: ERP correlates of initial feeling-of-knowing. *Neuropsychologia*. doi:10.1016/j.neuropsychologia.2008.12.009.
- Peacocke, C. (2007). Mental action and self-awareness (I). In B. McLaughlin & J. Cohen (Eds.), *Contemporary debates in the philosophy of mind* (pp. 358–376). Oxford: Blackwell.
- Peacocke, C. (2008). *Truly understood*. Oxford: Oxford University Press.
- Peacocke, C. (2009). Mental action and self-awareness (II): Epistemology. In L. O'Brien & M. Soteriou (Eds.), *Mental actions and agency* (pp. 192–214). Oxford: Oxford University Press.
- Pieschl, S. (2010). Epistemological beliefs and learning. Talk at the APIC Seminar, Institut Jean-Nicod, ENS, EHESS, Paris 22-01-2010.
- Povinelli, D. J., & Vonk, J. (2003). Chimpanzee minds: suspiciously human? *Trends in Cognitive Science*, 7, 157–160.
- Povinelli, D. J., & Vonk, J. (2006). We don't need a microscope to explore the chimpanzee's mind. In S. Hurley & M. Nudds (Eds.), *Rational Animals?* Oxford: Oxford University Press.
- Proust, J. (2007). Metacognition and metarepresentation: is a self-directed theory of mind a precondition for metacognition? *Synthese*, 159, 271–295.
- Proust, J. (2008). Epistemic agency and metacognition: a externalistic view. *Proceedings of the Aristotelian Society*, 108(3), 241–268.

- Proust, J. (2009a). It there a sense of agency of thought? In L. O'Brien & M. Soteriou (Eds.), *Mental actions and agency* (pp. 253–275). Oxford: Oxford University Press.
- Proust, J. (2009b). The representational basis of brute metacognition: a proposal. In R. Lurz (Ed.), *The philosophy of animal minds* (pp. 165–183). Cambridge: Cambridge University Press.
- Proust, J. (2009c). Overlooking the metacognitive experience. *Behavioral and Brain Sciences*, 32, 38–39.
- Proust, J. (2010). What is Metacognition? *Philosophical compass*.
- Reder, L. (1996). *Implicit memory and metacognition*. New Jersey: LEA.
- Reder, L., & Schunn, C. (1996). Metacognition does not imply awareness: strategy choice is governed by implicit learning and memory. In L. Reder (Ed.), *Implicit memory and metacognition* (pp. 45–77). New Jersey: LEA.
- Sanna, L. J., & Schwarz, N. (2003). Debiasing the hindsight bias: the role of accessibility experiences and (mis)attributions. *Journal of experimental social psychology*, 39, 287–295.
- Schwarz, N., & Vaughn, L. A. (2002). The availability heuristic revisited: ease to recall as distinct sources of information. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: the psychology of intuitive judgment*. NY: Cambridge University Press.
- Smith, J. D. (2009). The study of animal metacognition. *Trends in Cognitive Sciences*, 13(9), 389–396.
- Thompson, V. A. (2009). Dual-process theories: a metacognitive perspective. In J. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 171–195). Oxford: Oxford University Press.
- Walsh, M. M., & Anderson, J. R. (2009). The strategic nature of changing your mind. *Cognitive psychology*, 58, 416–440.
- Wegner, D. M. (2002). *The illusion of the conscious will*. Cambridge, MA: MIT Press.
- Wellman, H. (1990). *The child's theory of mind*. Cambridge: MIT.
- Whittlesea, B. W. A., Jacoby, L. L., & Girard, K. (1990). Illusions of immediate memory: evidence of an attributional basis for feelings of familiarity and perceptual quality. *Journal of Memory and Language*, 29, 716–732.
- Whittlesea, B. (1993). Illusion of familiarity. *Journal of Experimental Psychology*, 19(6), 1235–1253.
- Whittlesea, B., & Williams, L. (2001). Source of the feeling of familiarity: the discrepancy-attribution hypothesis. *Journal of Experimental Psychology*, 26(3), 547–565.