The Philosophy of Metacognition: Mental Agency and Self-Awareness

By Joëlle Proust

Oxford University Press, 2014. xii + 366 pp. £40.00.

In this ambitious book (which includes both previously published articles and new material), Joëlle Proust sets out to defend her 'evaluativist' conception of metacognition against the rival 'attributivist' conception and to draw out the implications of the former for a range of issues in philosophy of mind and epistemology. Metacognition is often defined as 'thinking about thinking'. That definition, however, risks begging the question against evaluativism, and, less memorably but more neutrally, Proust defines metacognition as 'the set of capacities through which an operating cognitive subsystem is evaluated or represented by another subsystem in a context-sensitive way' (14). These capacities centrally include the capacity for metamemory, which manifests itself, for example, in the familiar 'tip of the tongue' state, a state which turns out to be a reliable indicator of one's ability to retrieve the information for which one is searching.

Chapter 1 provides a brief historical sketch of research on analytic and procedural forms of metacognition – though psychologists initially concentrated their attention on the former, the latter has become increasingly prominent – and distinguishes between attributivism (defended, e.g. by Carruthers) and evaluativism. These two conceptions of metacognition are described in detail in Chapters 2 and 3, which focus on four key points of disagreement. First, whereas evaluativism claims that metacognition pertains essentially to one's own cognition, attributivism views it as a special case of a more general mindreading capacity, a capacity that can be directed either at one's own mind or at the mind of another agent. Second, evaluativism views metacognition as a matter of the dynamic, adaptive monitoring and control of mental processes; attributivism sees it in terms of static propositional representations of first-order mental states. Third, the evaluativist grants that metacognition may sometimes – when it has an analytic form – involve metarepresentation but maintains that it does not always involve metarepresentation; the attributivist, on the other hand, maintains that all metacognition is metarepresentational. Finally, while evaluativism sees metacognition as linked to mental agency, attributivism largely denies this link.

Having set out the rival conceptions, Proust launches her attack on attributivism in Chapter 4, making a case for viewing metacognition as being activity-dependent, in the sense that it is concerned with monitoring properties of cognitive processes, as opposed to the contents they produce, in sharp distinction to mindreading. She argues further that metacognition involves implicit selection of the norms against which processes are evaluated and that it is directly (automatically) action-guiding, additional features that would be difficult to reconcile with an attributivist conception. The attack continues as Chapter 5 reviews evidence for metacognition in non-human primates, arguing against several hypotheses meant to explain away this capacity. For example, on the belief competition hypothesis, apparent cases of animal metacognition do not in fact involve anything higher-order – anything properly 'meta' – but only competing first-order beliefs (e.g. about whether a given stimulus is present), together with rules for resolving conflicts among such beliefs. While acknowledging

Analysis Reviews Vol 75 | Number 2 | April 2015 | pp. 349–351 doi:10.1093/analys/anu159

that early results reported in the animal metacognition literature are compatible with this hypothesis, Proust argues that subsequent findings obtained using more sophisticated experimental paradigms rule it out. Her preferred alternative is a double accumulator model, according to which metacognitive monitoring and control depend on 'adaptive accumulators' which enable an agent to compute the difference between expected and observed confidence in his performance on a given task. Breaking the link between animal metacognition and mindreading further undermines the attributivist view. Chapter 6 then takes up the question of the representational format of procedural metacognition, arguing for an account in terms of 'feature-based thoughts' which builds on Strawson's notion of feature-placing thoughts: feature-placing thoughts are nonconceptual, nonpropositional means of picking out affordances available at certain locations and times in the subject's environment; analogously, feature-based thoughts provide a non-conceptual, non-propositional means of picking out *cognitive* affordances (e.g. the ability to retrieve the answer to a question from memory) available to the agent. Feature-based thoughts are well-suited to play a role in procedural metacognition, including in animals incapable of conceptual, propositional thought.

Turning to mental action, Chapter 7 argues that mental action constitutes a natural kind distinct from bodily action, since mental actions are performed to satisfy basic informational needs, and since they are subject to specific epistemic norms. Chapter 8 takes up the norms involved in mental action, focusing on the particular case of acceptance. Proust argues that there are multiple forms of acceptance, associated with multiple epistemic norms (accuracy, exhaustivity, coherence, consensus and so on) and that acceptance requires a 'two-tiered' account that distinguishes between 'epistemic' acceptance (treating a proposition as if it were true) and 'strategic' acceptance (choosing to act on the proposition). Epistemology comes more clearly into focus in Chapter 9, which argues that, while it is natural to suppose that the cognitive role of metacognition supports an internalist perspective in epistemology, in fact the apparent kinship between metacognition and internalism is largely illusory, since a subject may be aware of his metacognitive feelings without being aware of the objective basis of their reliability (e.g. fluency); externalism is thus better placed to provide an account of the role of metacogition.

In Chapter 10, Proust explores the role of metacognition in mental agency, arguing that knowledge of one's own mental actions depends in part on the sensitivity to the adequacy of those actions that is provided by metacognitive self-probing and post-evaluation. Chapters 11 and 12 continue to investigate the connections among metacognition, the self, and agency, looking at how metacognition might underwrite the capacity for thinking of oneself as the same individual over time and at how the sense of agency may be disturbed in schizophrenia. Changing pace, Chapter 13 relates metacognition research to research on embodied communication, defending the view that there is a class of metacognitive conversational gestures – bodily movements that function to communicate metacognitive states, such as scratching one's head to indicate effort – that have so far not been acknowledged. This is a novel proposal, and the chapter provides a nice illustration of the way conceptual philosophical work might feed back into empirical metacognition research. Chapter 14 sums up the argument and discusses the implications of procedural metacognition for the personal/subpersonal and system 1/system 2 distinctions, using metacognition as a test case to argue that the personal/subpersonal distinction is orthogonal to the system 1/system 2

distinction, in the sense that, if system 1 is viewed as a nonconceptual, featural system, its output may be available at the personal level; what is distinctive of system 1 is primarily its inflexibility, which derives from its nonconceptual character.

This highly original book is the fruit of many years of thought about metacognition and will no doubt stand as the major philosophical contribution to the area for some time to come. Readers with a psychology background will benefit from the conceptual clarity with which the book views experimental results, and philosophers will find in it an accessible introduction to a field of psychological research with important philosophical dimensions and implications. There are, naturally, many points at which one might disagree with details of Proust's argument – for example, given the deep differences between analytic and procedural metacognition, one might suspect that the two capacities should not be grouped together under a common heading – but the book's major limitation has less to do with its content than with its form. At 14 densely written chapters, the book is an exceptionally demanding read, going into great detail about the relevant theoretical debates and making heavy use of novel technical terminology. Indeed, given the sheer level of detail included, the reader at times (e.g. in the chapter on conversational metacognition, which is only loosely related to the remainder of the book) risks losing track of the thread of the argument. Lest the reader be overwhelmed by unfamiliar jargon, Proust has included a glossary, but it would have been preferable to simplify the language and exposition to the extent possible. Given that there are few philosophical works devoted primarily to metacognition, the value of a book such as this lies as much in its demonstration of the philosophical relevance of empirical metacognition research as it does in the particular arguments and positions it defends, and there is a risk that this relevance will be obscured by the wealth of detail included in the book.

KOURKEN MICHAELIAN
Department of Philosophy
University of Otago
Dunedin 9054 New Zealand
michaelian.kourken@gmail.com

The Limits of Kindness

By CASPAR HARE

Oxford University Press, 2013. xii + 230 pp. £25.00.

In his *The Limits of Kindness*, Caspar Hare attempts to make progress in normative ethics by taking what he calls a foundational approach. This approach involves starting with the normative claims that are regarded as 'prima facie obvious' as our starting points. These claims will then be combined with minimal assumptions about rationality in order to reach conclusions about '*the thing to do*' (222). Hare then proceeds in the three sections of his book to address issues about (i) saving others from harm when we know the identity of the persons in peril, (ii) saving others from

Analysis Reviews Vol 75 | Number 2 | April 2015 | pp. 351–353 doi:10.1093/analys/anv008