

1 Chapter 5

2 **The emergence of metacognition:**
 3 **affect and uncertainty in animals**

4 Peter Carruthers and J. Brendan Ritchie



5 **Introduction: the meaning of ‘metacognition’**

6 We assume that readers of this volume will by now have some familiarity with the sorts of para-
 7 digms that have been used to provide evidence of metacognition in non-human primates. In a
 8 common type of experiment (e.g. Smith et al. 2008), animals are trained to perform a primary
 9 task such as making a discrimination of some sort between categories (e.g. sparse versus dense) to
 10 achieve a favoured reward (either immediately, or after a delay; Couchman et al. 2010). After
 11 training, the animals are also provided with an ‘opt out’ response of some kind, which they tend
 12 to use in difficult cases where they are more likely to make (or have made) an incorrect judge-
 13 ment. Opting out generally either avoids the penalty that accompanies a mistaken answer (such
 14 as a timeout before there is another opportunity to obtain a reward), or guarantees a less-favoured
 15 reward. Such results are said to show that the animals are aware of their own uncertainty, espe-
 16 cially since similar use of the opt-out response in humans is associated with self-attributions of
 17 uncertainty.

18 We fully accept that this body of work, taken as a whole, cannot be explained in low-level asso-
 19 ciationist terms, as involving mere conditioned responses to stimuli. A great deal of careful
 20 experimentation has been done to demonstrate that this is not the case, and we are happy to
 21 embrace this conclusion (Beran et al. 2009; Couchman et al. 2010; Smith et al. 2010; Washburn
 22 et al. 2010). So it should be agreed that the animals have beliefs about the contingencies of
 23 the experiment and take executively-controlled decisions that depend on those beliefs (as well as
 24 having goals and other states like emotions, which some have been reluctant to attribute to ani-
 25 mals; but see Panksepp 2005).

26 However, to say that the animals’ behaviour is fully cognitive and executively controlled is not
 27 yet to say that it is *metacognitive*, in the sense in which this term is employed throughout cogni-
 28 tive and developmental psychology. For metacognition is generally defined as ‘thinking *about*
 29 thinking’ (Flavell 1979; Dunlosky and Metcalfe 2009), and therefore as involving metarepresenta-
 30 tion. Moreover, metarepresentation in turn is understood to require a representation that repre-
 31 sents another representation, or a mental state whose content represents, and is *about*, another
 32 mental state.

33 This definition of ‘metacognition’ accords with the standard model for classifying and charac-
 34 terizing metacognitive processes in humans (Nelson and Narens 1990; see Fig. 5.1), in which a
 35 metalevel monitors, represents, and controls the processes of object-level cognitive systems. Since
 36 those who study metacognitive processes in animals often cite this model with approval (e.g.
 37 Smith et al. 2003, 2006; Couchman et al. 2010), we assume that it is some version of *this* architec-
 38 ture, or some of its components, that the animals in question are claimed to possess when they are
 39 said to have metacognitive capacities. And it should be noted that an important aspect of the

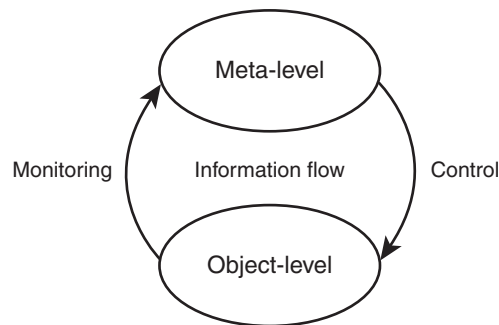


Fig. 5.1 Metacognitive monitoring and control (adapted from Nelson and Narens 1990).

1 Nelson and Narens model has always been that the metalevel contains a *metamodel* of the object-
 2 level, with the metalevel containing metarepresentations of processes and events at the object-level.
 3 (See, for example, Nelson and Narens, 1990, p. 126, Principle 2.)

4 While these definitional issues are important, we should stress that this is only because clarity
 5 is important for the progress of science. What ultimately matters, of course, is knowledge of the
 6 cognitive structures and processes that underlie the animals' behaviour, not the words we use to
 7 express that knowledge. What we will suggest is that some of the behaviour (specifically, so-called
 8 'uncertainty-monitoring' behaviour) that has been claimed to support the presence of a metacog-
 9 nitive architecture can just as well be explained in non-metarepresentational affective terms.

10 Moreover, it should be stressed that even if some of the processes employed by animals in
 11 uncertainty-monitoring experiments might appropriately be described in terms of 'monitoring
 12 and control', it is another matter to claim that the monitoring in question is metarepresenta-
 13 tional, or metacognitive in the standard sense. For there are multiple monitoring and control
 14 processes in human cognition that are not metarepresentational (Metcalfe 2008). Consider, for
 15 example, the use of forward models in the control of action. When motor schemata are activated
 16 and sent to the motor system to initiate an action, an efference copy of those instructions is sent
 17 to an emulator system that constructs a forward model of the expected sensory consequences of
 18 the movement (Wolpert and Kawato 1998; Wolpert and Ghahramani 2000; Jeannerod 2006).¹
 19 This is then received as input by a comparator mechanism that also receives reafferent sensory
 20 feedback, issuing in swift online adjustments in the action when there is a mismatch. Note that
 21 the comparator system is entirely non-metarepresentational in nature: it receives a sensory-coded
 22 representation of the intended outcome and compares this with sensory input from the actual
 23 outcome as it unfolds. When these fail to correspond, it employs an algorithm that adjusts the
 24 motor instructions to bring about a closer match. It doesn't need to represent either the motor
 25 intention or the current experiences resulting from the action as such. No metarepresentations
 26 are needed, and no one in the field of motor processing thinks that they are employed.

¹ This same system is also used offline, when subjects mentally rehearse potential actions for purposes of decision-making. In such cases a motor schema is activated, and although the instructions that would normally be sent to the muscles have been inhibited, the emulator system goes ahead and constructs a representation of the expected sensory consequences. This sensory representation can be 'globally broadcast' (in the sense of Baars, 1988) when attended to, thus being made available to a range of systems to draw inferences and evaluate the action. We return to these points later in the chapter.

1 We will assume, then, that those who propose metacognitive explanations for the behaviour
 2 of animals in uncertainty-monitoring experiments intend this in the standard sense: they are
 3 claiming that the animals metarepresent their own states of uncertainty, and modify their behav-
 4 iour as a result. We will suggest, in contrast, that the data can equally well be explained in non-
 5 metarepresentational terms. First, however, we propose to situate the issue within a wider debate
 6 about the evolutionary emergence of metarepresentational capacities.

7 **The phylogeny of metarepresentation**

8 Metacognition and mindreading (or ‘theory of mind’) are widely believed to overlap (at least) in
 9 their psychological bases and evolutionary histories. This is because both rely, fundamentally, on
 10 metarepresentation: the representation of mental states. In the case of mindreading, this involves
 11 attributing mental states to others, while in metacognition we attribute mental states to ourselves.
 12 When one attempts to explain the adaptive advantage that these capacities supplied to our ances-
 13 tors, a notion of control is invoked in each case. Mindreading allows us to predict the behaviour
 14 of others in order to control our own (social) behaviour. Hence, mindreading is thought to have
 15 evolved to navigate an increasingly complex social world, engaging with multiple conspecifics in
 16 groups with complex social organization. Metacognition, in contrast, allows us to monitor and
 17 control object-level systems in our own mind, enabling us to learn and reason more flexibly.

18 Metarepresentation then features in both mindreading and metacognition, but in the service of
 19 rather different functions (social cognition versus cognitive control). This leads us to ask which
 20 function of metarepresentation is evolutionarily prior (as well as how this bears on the question
 21 of human cognitive architecture). The question of priority naturally suggests two kinds of
 22 account of the evolution of metarepresentational capacities.²

23 According to one approach, the capacity to represent one’s own mental states (or some subset
 24 thereof) evolved first (Couchman et al. 2009), presumably to enable animals to accrue the bene-
 25 fits of metacognitive monitoring and control. Once evolved, the conceptual and inferential
 26 resources involved were later exapted for attributing mental states to other agents. There are two
 27 main ways in which this could have happened, partly motivated by different views of human
 28 mindreading. Either these first-person resources were redeployed to form the basis of a distinct
 29 mindreading faculty of the sort defended by Nichols and Stich (2003), or they were combined
 30 with emerging capacities for imaginative perspective-taking to enable *simulations* of the mental
 31 lives of others (Goldman 2006). We will refer to these as ‘first-person-based’ accounts of the evo-
 32 lution of metarepresentation, while making no attempt to adjudicate between dual-mechanism
 33 and simulationist variants.

34 According to the alternative approach, a capacity to attribute mental states to other agents
 35 evolved first, driven by the exigencies of social living and resulting in an innately channelled min-
 36 dreading faculty of some sort. But this mindreading-based account also admits of two main vari-
 37 ants. According to one, a core *capacity* to make self-attributions would have been present from
 38 the start, since there would have been nothing to prevent subjects from turning their mindreading
 39 abilities on themselves, treating the self as an agent like any other. A disposition to attribute men-
 40 tal states to oneself on a regular basis would only have required the motivation to direct one’s
 41 attention accordingly (Carruthers 2011). According to the other variant of a mindreading-based
 42 account, in contrast, some sort of self-monitoring mechanism was subsequently added to the

.....
 2 We assume that no one should now think that these capacities result from general learning, and that eve-
 ryone should agree that they are innately channelled in development to some significant degree. While
 these assumptions go undefended here, they are in fact supported by large and varied bodies of data. See
 Carruthers (2011) for further discussion.

1 third-person mindreading system, enabling direct (non-sensory) access to one's own mental
 2 states (Frith and Happé (1999) seem to have in mind something like this). In this case we propose
 3 *not* to remain neutral between the two variants, but will work with a self-directed-mindreading
 4 account throughout. This provides the cleanest contrast with first-person-based approaches. And
 5 there is, in fact, a good deal of evidence against the monitoring-mechanism alternative (see
 6 Carruthers 2011).

7 We will shortly compare the first-person-based and mindreading-based accounts of the evolu-
 8 tion of metarepresentation with respect to the predictions that each makes regarding the com-
 9 parative data. But first it is worth noting an apparent anomaly for the former. This is that it is
 10 widely agreed among psychologists that human metacognitive capacities (or at least those of an
 11 uncontroversially metarepresentational sort) are far from impressive. For example, one robust
 12 finding in the literature is that people's metacognitive judgements of learning are only moderately
 13 correlated, at best, with later recall (Leonesio and Nelson 1990; Dunlosky and Metcalfe 2009), and
 14 another is that correlations between metacognitive judgements of text comprehension and tests
 15 of understanding are often close to zero (Lin and Zabrocky 1998; Maki and McGuire 2002).
 16 Moreover, human metacognitive capacities are fragile and cue-based, are late to develop in child-
 17 hood, and are heavily dependent upon individual differences in personality and local cultural
 18 mores for their effectiveness (Stanovich and West 2000; Koriat et al. 2006, 2008; Stanovich
 19 2009).

20 These findings are not what might be expected if metacognitive abilities had a long evolution-
 21 ary history and are innately channelled in development. In contrast, everyone agrees that human
 22 mindreading capacities are remarkably good (although admittedly we lack any shared metric for
 23 comparing mindreading capacities with metacognitive ones). More importantly, we now have
 24 ample evidence of their early emergence in human infancy (Southgate et al. 2007, 2010; Surian
 25 et al. 2007; Song et al. 2008; Buttelmann et al. 2009b; Scott and Baillargeon 2009; Scott et al. 2010).
 26 This is just as might be predicted by a mindreading-based account of the evolution of metarepre-
 27 sentational abilities.

28 It could be replied, of course, that biological structures need only deliver small adaptive advan-
 29 tages in order to be selected for, especially over a long time-frame. And it is possible that metarep-
 30 resentational capacities evolved initially for first-person metacognitive uses, after which the main
 31 adaptive pressure became a social one. This would explain the seemingly poor metacognitive
 32 capacities of humans combined with excellent mindreading. One might expect, however, that if
 33 metacognitive capacities had been selected for among our ancestors, then they would have come
 34 under additional adaptive pressure (leading to further robustness and reliability) when learning
 35 and decision-making become increasingly complex through the evolution of the hominin line. In
 36 any case the contrast between human native capacities for metacognition, on the one hand, and
 37 mindreading, on the other, appears striking, and provides some indirect support for a mindreading-
 38 based account of the evolution of metarepresentation.

39 **Predictions for comparative psychology**

40 If metarepresentational capacities evolved initially for metacognitive monitoring and control,
 41 then one might expect to find creatures capable of metacognition who are *incapable* of mindread-
 42 ing (or at least, who are incapable of mindreading of a sort that requires equivalent metarepresen-
 43 tational resources; see later). At any rate, on this view there must once have been such creatures.
 44 Moreover, if creatures of this sort were now discovered, then it would provide significant support
 45 for a first-person-based account of the emergence of metarepresentation. For the mindreading-
 46 based account predicts, in contrast, that metarepresentational capacities should emerge in parallel
 47 for self and other (while perhaps allowing that other-directed metarepresentation might precede

1 equivalent forms of metacognition, if, for example, the animals aren't initially motivated to
 2 attend to their own mental states). This is because metacognition is held to result from (or at least
 3 to employ the conceptual and computational resources of) self-directed mindreading.

4 The qualification about 'equivalent metarepresentational resources' is important. This is
 5 because it is widely agreed among developmental psychologists that mindreading admits of two
 6 distinct varieties, which emerge at different points in the course of infant development (Wellman
 7 1990; Leslie 1994; Baron-Cohen 1995; Gopnik and Meltzoff 1997; Song and Baillargeon 2008).
 8 One is a form of goal/perception/knowledge-ignorance psychology that appears during the first
 9 year of life. Infants at this stage can represent the goals of other agents, as well as track what
 10 aspects of the world those agents do and do not have perceptual access to. As a result, infants at
 11 this age form appropriate expectations of agents who act in states of knowledge or ignorance
 12 respectively. But at this stage (generally referred to as 'Stage 1'), infants are incapable of represent-
 13 ing the false belief of another agent, or of forming expectations based on how things *appear* to the
 14 other agent. These latter capacities only emerge toward the end of the fourth year of life (in lan-
 15 guage-based tasks), or by the middle of the second year of life (when non-verbal measures of
 16 competence are employed). Moreover, it is widely believed that the difference between Stage 1
 17 and Stage 2 mindreading is one of domain-specific conceptual and/or computational *competence*,
 18 rather than resulting merely from performance factors. For it is thought that the capacity to pass
 19 Stage 2 tasks depends on an appreciation that mental representations can be *incongruent* with
 20 reality (as in a false belief), as opposed to merely *omitting* an aspect of reality (as happens in
 21 ignorance).³

22 There is now evidence of Stage 1 mindreading capacities in non-human animals, not only
 23 among other primates such as chimpanzees and rhesus macaques (Hare et al. 2000, 2001, 2006;
 24 Flombaum and Santos 2005; Melis et al. 2006; Santos et al. 2006; Buttelmann et al. 2007, 2009a),
 25 but also among canids (dogs and wolves; Hare and Tomasello 2005; Hare 2007; Udell et al. 2008),
 26 and corvids (jays, rooks, crows, and the like; Bugnyar and Heinrich 2005, 2006; Dally et al. 2006,
 27 2009; Bugnyar et al. 2007; Stulp et al. 2009). Note that all of these animals live in complex social
 28 groups, suggesting that the pressures of social living might have converged on the evolution of
 29 simple forms of mindreading capacity in widely separated species (Emery and Clayton 2004),
 30 consistent with a version of the 'Machiavellian intelligence' hypothesis (Byrne and Whiten 1988,
 31 1997).

32 Given the presence of Stage 1 mindreading in non-human primates, the finding that they may
 33 be capable of monitoring their own desires (Evans and Beran 2007), their own perceptual access
 34 (Call and Carpenter 2001; Hampton et al. 2004; Krachun and Call 2009), and their own knowl-
 35 edge and ignorance (Hampton 2001, 2005), fails to adjudicate in our dispute. For these findings
 36 are consistent with both self-directed-mindreading and first-person-based accounts.⁴

³ It may yet turn out that this assumption is mistaken. Rather than reflecting differences in mindreading competence, the differences in performance might turn out to result from the differing executive demands of Stage 1 and Stage 2 tasks (Carruthers, forthcoming). If so, then the failures of non-human primates on Stage 2 tasks might likewise result from problems of executive function. This would mean that the meta-cognitive data are incapable of adjudicating in the dispute between first-person-based and mindreading-based accounts of the evolution of metarepresentation. For there would then be no reason to think that non-human primates are capable of forms of metacognition that outstrip their capacities for mindreading, even if they employ Stage 2 metarepresentational capacities in metacognitive tasks.

⁴ In fact we have doubts about the strength of some of this evidence. In particular, success in the memory monitoring experiments conducted by Hampton (2001) does not require attribution of knowledge or ignorance to oneself. It just requires the presence or absence of memory. The animal needs to act in one

1 In contrast, the current consensus among comparative researchers is that no primate species
 2 other than humans is capable of ‘Stage 2’ mindreading, which would include a capacity to attribute
 3 false beliefs to other agents. For all tests of such abilities have proved negative, even when
 4 conducted in competitive situations, and even when paired with knowledge–ignorance tasks that
 5 the animals pass (Hare et al. 2001; O’Connell and Dunbar 2003; Kaminski et al. 2008; Krachun
 6 et al. 2009). So if other primates can attribute such states to themselves, then this would present
 7 an anomaly for a mindreading-based account, while providing corresponding support for a first-
 8 person-based view.

9 While there is no data of quite this kind in the literature, a substantial body of work on uncer-
 10 tainty monitoring aims to show that members of many primate species are capable of monitoring
 11 their own states of certainty and uncertainty, and of choosing adaptively as a result. This might
 12 be taken to demonstrate that these animals are capable of Stage 2 metacognition, suggesting that
 13 they possess the *concept* of false belief, at least, and can apply it in the first person. For one might
 14 think that mastery of the concept of uncertainty requires a capacity to understand that one’s
 15 beliefs are potentially false. Whether or not this is so will be discussed in the next section.

16 **Uncertainty and feelings of uncertainty**

17 Uncertainty, like certainty, is fundamentally a cognitive state, not an emotional one. To be certain
 18 of something is to have a high degree of belief that it is the case. (This might be realized in the
 19 form of an especially strong signal produced by a classifier mechanism, for example, or an espe-
 20 cially strong memory trace.) To be uncertain of something is to have a low degree of belief that it
 21 is so (perhaps realized in a weak signal from a classifier mechanism, or a weak memory trace).
 22 However, each of these states can also give rise to distinctive emotional feelings of confidence or
 23 uncertainty. Moreover, each will have other cognitive and behavioural effects as well, including
 24 *fluent* cognitive processing (in the case of certainty) and *disfluency* (in the case of uncertainty).⁵
 25 These further consequences of uncertainty will be used to undergird our alternative (non-
 26 metarepresentational) explanations of the uncertainty-monitoring data in the next section.

27 If animals self-monitor and metarepresent themselves as uncertain of something, then they
 28 must be representing that they have a low degree of belief in it. This will require that they possess
 29 Stage 2 metarepresentational resources. For it cannot be sufficient to represent that one is certain
 30 of something to represent that one knows it (utilizing one of the concepts from Stage 1), and nor
 31 can it be sufficient to represent uncertainty to think that one is ignorant. This is because neither
 32 knowledge nor ignorance admit of degrees, and nor do they imply some level of incongruency
 33 with the world, as do degrees of belief. (Recall that a capacity to represent that a mental state is
 34 incongruous with the world is thought to be the hallmark of Stage 2 metarepresentation.)
 35 Moreover, in principle the metarepresentational states involved could be based on self-monitoring
 36 that is direct (detecting or introspecting a judgement with a low degree of belief) or indirect
 37 (detecting and classifying sensory or behavioural cues of the underlying state of uncertainty).
 38 Since no one in the human metacognition literature thinks that monitoring is direct, we propose
 39 to dismiss this possibility in respect of animals also (Koriat 2000; Dunlosky and Metcalfe 2009).

40 In fact it should be stressed that there is general agreement among researchers that human
 41 metacognitive judgements are *cue based* (Dunlosky and Metcalfe 2009). Judgements about

way if a memory is present, and to act in another if it is not. But in neither case does it need to entertain a metarepresentation of memory. See Carruthers (2008).

⁵ Cognitive fluency is the ease with which information is processed in the mind, and is signalled by such factors as the speed with which a decision is reached or an item is recognized.

1 whether one has learned something or whether one knows something are grounded in sensorily-
 2 accessible and affective cues, such as the ease with which the item in question is processed or the
 3 feeling of familiarity induced by its presentation. For although Hart (1965) once proposed a sort
 4 of direct-access model in order to explain feelings of knowing, his account has attracted very little
 5 empirical or theoretical support since then (Koriat 2000). We should therefore expect that ani-
 6 mals, too, will need to base their judgements on indirect cues—perhaps their own disfluency, or
 7 perhaps their own feelings of uncertainty.

8 Since humans in uncertainty-monitoring experiments must base their reports of their uncer-
 9 tainty on sensorily-accessible cues of some sort, it is reasonable to assume that the same, or some-
 10 thing similar, is true of non-human primates. So it will be important to know how feelings of
 11 uncertainty should be characterized, as well as what other similar cues might be in the offing.
 12 What we can say with confidence is that often the feelings in question are negatively-valenced
 13 states accompanied by a degree of arousal that is proportional to what is at stake. Feeling uncer-
 14 tain can feel bad (to a greater or lesser degree), and it can also be agitating when concerned with
 15 important matters.⁶

16 Feelings of uncertainty are caused by underlying states of uncertainty (that is, low degrees of
 17 belief). It is a separate question, however, what the negative valence component of the feeling is
 18 directed toward. What is it that one feels bad about, when one feels uncertain? What situation or
 19 state of affairs is it that *seems bad* as a result of negative valence, in the way that fear makes the
 20 threatening object seem bad and anger makes the causes of damage to oneself or to one's own
 21 seem bad? One possibility would implicate metarepresentation in the very feeling of uncertainty
 22 itself, utilizing metarepresentational resources. It may be that what strikes one as bad is that one
 23 has a low degree of belief. On this account, a judgement to the effect that one has a low degree of
 24 belief would be built into (or at least accompany) the feelings in question, providing the intended
 25 object or target of those feelings.

26 What we propose, however, is that feelings of uncertainty (in both humans and animals) are
 27 more plausibly seen as directed at the world (in particular, at the primary options for action that
 28 are open to one), rather than at one's own mental states. Consider what happens when people
 29 engage in the Iowa Gambling Task, for example (Bechara et al. 1994). Subjects are required to
 30 select from four decks of cards with different probabilities of winning or losing. Two of the decks
 31 produce steady gains in the long-run (while sometimes issuing in big losses), while two produce
 32 long-term losses (and yet sometimes issue in big gains). After a while subjects begin to make most
 33 of their selections from the 'good' decks, but before they are capable of explicit recognition that
 34 those decks are better (let alone capable of articulating *why* they are better). Presumably, as a
 35 result of previous learning, the good decks are unconsciously appraised as more likely to issue in
 36 gains. As a result, the thought of selecting from those decks is positively valenced, making those
 37 options seem better. But in addition, some minor degree of arousal is also present, since subjects
 38 display an increased galvanic skin response when reaching toward one of the bad decks.⁷

.....
 6 Note that we are not claiming that there is a unique introspectively-accessible feeling that is distinctive of states of uncertainty. Nor do we think that affective changes are always consciously experienced. All we need to be committed to for present purposes is that there will generally be *some* degree of affective change accompanying states of uncertainty, whether consciously experienced or not, and that these can exert an influence on subsequent behaviour.

7 Amiez et al. (2003) used a decision-making task equivalent to the Iowa Gambling Task with macaques, but found that the galvanic skin response occurred *after* the animals had made their selection, seemingly in anticipation of a reward. Quite how galvanic skin responses in uncertainty tasks like these are supposed to support Bechara and colleagues' own 'somatic marker' account of affective decision-making is a

1 We should stress that in cases of this sort the affective changes can be quite minor, and may pass
 2 unnoticed by the subject. Yet still the good options seem good and the bad options seem bad, with
 3 effects on behaviour that can be quite significant. Certainly in humans, minor forms of affective
 4 priming can have large behavioural consequences. For example, Winkielman et al. (2005) used
 5 briefly presented, backward-masked, happy and angry faces (which were never consciously per-
 6 ceived) before subjects sampled a novel beverage. Thirsty subjects primed with positive affect
 7 drank twice as much of the beverage as those primed with negative affect, and in another condi-
 8 tion, they offered to pay twice as much for a can of the drink having taken just a sip. Yet these
 9 unconscious primes had no discernable effects on the subjects' mood.

10 In fact we think that uncertainty-based decision-making may be best understood as of-a-piece
 11 with affectively-based decision-making generally, of the sort characterized by Damasio (1994),
 12 Gilbert and Wilson (2007), and many others. On this kind of account one runs the instructions
 13 for a motor action offline, using the efference copy to generate a forward model of its outcome (as
 14 described in the first section). When attended to, this is globally broadcast as an imagistic repre-
 15 sentation of the action, which one's evaluative and emotional systems receive and respond to. The
 16 result is some degree of positive or negative affect, which provides the motivation to execute the
 17 action or to seek an alternative means to the goal (or to pursue an alternative goal). On this kind
 18 of account feelings of uncertainty would consist of negatively valenced affect that is caused by the
 19 thought of an otherwise-attractive action, and that is directed toward the situation represented in
 20 the content of that thought. (It is the performance of the action that seems bad as a result, not the
 21 fact that one is thinking about it.)

22 There is some reason to believe that members of other primate species might be capable of such
 23 processes of mental rehearsal and affective evaluation, underlying their limited capacity for
 24 advance planning (Sanz et al. 2004; Mulcahy and Call 2006), and perhaps also explaining instances
 25 of 'insight' behaviour (see Carruthers 2006, for discussion). And indeed, a similar capacity might
 26 be more widespread still. Think of the cat that crouches down as if to leap, *literally* rehearsing (the
 27 first stages of) a difficult leap from a roof to a nearby tree. Presumably the act of representing the
 28 action issues in appraisals of likely success, resulting either in positive affect (felt confidence)
 29 directed at the intended leap, or in negative affect (felt uncertainty), leading the cat to seek other
 30 solutions.

31 In this section we have distinguished uncertainty from the cognitive and affective consequences
 32 of uncertainty, and we have pointed out that animals, like humans, will need to rely on indirect
 33 cues of uncertainty, even if they do metarepresent such states. We have also suggested that the
 34 valence component of feelings of uncertainty is directed at the primary response options, rather
 35 than at one's own mental states. While humans engage in many forms of metacognitive decision-
 36 making, requiring them to metarepresent their own mental states and processes, basic forms of
 37 affectively-based decision-making are *not* metarepresentational in humans. When we represent
 38 and respond affectively to alternative courses of action, no metarepresentations need be involved.
 39 As a result, in the following section we will suggest that the uncertainty-monitoring data may be
 40 explained without ascribing metarepresentational capacities of any sort to the animals involved.⁸

complicated matter, however (Dunn et al. 2006). So it is far from clear that this result undermines their hypothesis. But in any case our view is not committed to the details of this particular theory of the manner in which affective cues influence decision-making. Indeed, our primary focus is on the valence component of affect, rather than on bodily arousal.

⁸ This will mean that even if the evidence suggesting that non-human primates are capable of Stage 1 forms of mindreading proves to be unsound, it will still be the case that the uncertainty-monitoring data fail to

1 **Affective explanations of the evidence**

2 The present section will discuss three distinct non-metarepresentational explanations of the
 3 uncertainty-monitoring data from non-human primates. The first is unsatisfying on its own. But
 4 each of the others provides a viable alternative to a metarepresentational account. We will focus
 5 especially on a valence-based theory that builds on some of the ideas from the previous section.

6 **Degrees of belief**

7 One form of non-metarepresentational explanation is proposed by Carruthers (2008), who
 8 appeals to degrees of belief and desire, together with ordinary practical reasoning, to show how
 9 the uncertainty-monitoring data can be explained. While this account may not be incorrect, it
 10 strikes us as incomplete. This is because it is purely cognitive in nature, and fails to provide for the
 11 emotional character of uncertainty.⁹ Since humans in such experiments report not only that they
 12 *are* uncertain (in the sense of having low degrees of belief) but that they *feel* uncertain (and
 13 indeed, since a judgement that one is uncertain must be grounded in indirect cues such as feelings
 14 of uncertainty), it seems inadvisable to omit an affective component from the explanation. For
 15 the results of uncertainty monitoring experiments with humans can parallel the animal uncer-
 16 tainty-monitoring data quite closely (Smith et al. 2003, 2008; Smith 2005). Accordingly, two
 17 further accounts will be outlined here. Each appeals to the consequences of states of uncertainty,
 18 while differing from one another in the factors that are utilized. We should emphasize, however,
 19 that these accounts are consistent with one another. Each might apply in different kinds of case,
 20 or they might combine together in the same cases.

21 **Affective consequences as cues**

22 One possibility is that the animals in question have learned to use some aspect of their own feel-
 23 ings of uncertainty as a *cue*, but without at any time metarepresenting that they are uncertain (i.e.
 24 without categorizing their affective experience as a feeling *of* uncertainty), or thinking that their
 25 judgements or memories are likely to be false. In effect, they may be following a rule like, ‘When
 26 in a state of *that* sort [uncertainty], opt out and do something different’, which would only
 27 require possession of an indexical, non-mental, concept. This can explain why the animals are
 28 more likely to press the opt-out key in psychophysically difficult cases, and it can also explain how
 29 the animals are able to generalize the use of the opt-out key when presented with it in the context
 30 of a newly learned discrimination task (Son and Kornell 2005; Kornell et al. 2007). But neither the
 31 feeling itself, nor the indexical concept used to identify it, need involve metarepresentation.
 32 Rather, just as humans are apt to do (Koriat 2000), the animals might utilize their own disfluency
 33 or the bodily feelings associated with uncertainty as cues when confronted with difficult cases.

34 We know from the human case that affective states provide experiential cues for metacognitive
 35 deliberation (Koriat 2000; Koriat et al. 2006, 2008). For example, differences in affective states dur-
 36 ing learning or retrieving information are used as cues that reflect the underlying processing
 37 dynamics or processing fluency. In the case of information that is easy to process, greater fluency
 38 results, causing positive affect (Winkielman and Cacioppo 2001). The extent to which people show
 39 confidence in automatic, intuitive, judgements is heavily dependent on processing fluency. This is
 40 true, for example, in low-difficulty recognition tasks, where selection can be based primarily on

support first-person-based views. For uncertainty-monitoring behaviour arguably fails to involve metarep-
 resentations of any sort (whether Stage 1 or Stage 2).

⁹ Emotions might, however, be incorporated into the account as a way of implementing the so-called ‘gating mechanism’ appealed to.

1 the feeling of familiarity without serial recall (e.g. Mandler 1980), and also in implicit learning
2 tasks (Gordon and Holyoak 1983).¹⁰

3 *Disfluent* processing, by contrast, has been suggested to play a role in initiating a transition
4 from more automatic processing to more executively-controlled explicit processing (Alter et al.
5 2007). This role for disfluency has been interpreted as a cue for metacognitive processing, but
6 being sensitive to disfluency need not presuppose any capacity for metarepresentation. For what
7 the experiment by Alter and colleagues shows is that disfluent processing causes changes in *atten-*
8 *tion*, issuing in different forms of cognitive control. And these changes in attention to the task
9 might drive the selection of different reasoning strategies in the absence of metacognitive process-
10 ing. Alternatively, disfluent processing might cause subjects to attend to their own increased
11 arousal, for example, which is taken as a cue to reason differently.¹¹

12 Note that although this sort of account need not involve metarepresentation, it does rely on
13 self-directed forms of attention. For the animals will attend to, and notice, something about
14 themselves (such as their own bodily feelings) in order to learn the cue-based rule in question. So
15 it can appropriately be described as a form of uncertainty *monitoring*, even if the monitoring
16 involved is not metarepresentational. The remaining form of affect-based explanation, in con-
17 trast, is entirely outward-looking or world-directed in character, while likewise finding a basis in
18 what is known about human decision-making.

19 **Directed valence**

20 Suppose that animals, like humans, integrate probabilistic information with intended goal out-
21 comes to issue in appraisals of the likelihood of success of the options available to them in a
22 decision-making context. In that case, when an animal has a low degree of belief in something
23 (that the pattern on the screen is dense rather than sparse, or that it has just touched the longest
24 of the lines on the screen, for example), then actions that depend upon the truth of that belief will
25 be appraised as unlikely to succeed.¹² Consequently the animal will experience some degree of
26 anxiety when it contemplates pressing the ‘dense’ key or the ‘gamble’ option (albeit quite minor,
27 since the stakes are so low). With negative valence directed at the action in question, it will to that
28 extent seem bad or aversive. In such circumstances the primary response options will be seen in a
29 mixed evaluative light. On the one hand they will seem good, since they hold out the possibility
30 of a significant reward; but on the other hand they will seem bad, since they are appraised as
31 unlikely to succeed. The opt-out response, in contrast, will be seen as an unopposed weak posi-
32 tive, since it either advances the animal to a new trial without a time-out or issues in a guaranteed
33 less favoured reward. It is not surprising, then, that the animals should press the opt-out key more
34 often in such circumstances.

35 As we noted earlier, this explanation coheres well with what is known about the decision-
36 making processes employed by humans. When humans are confronted with choices they will
37 generally rehearse the actions involved in implementing those choices. These representations,
38 when taken as input by the individual’s affective mechanisms, will result in some degree of posi-
39 tive or negative affect directed at the option in question. This makes that option seem either good

10 Here we presuppose a processing-fluency view of feelings of familiarity, in the manner of Jacoby (1991).

11 Note that neither interoception nor proprioception, of the sort that might underlie awareness of arousal, are metarepresentational forms of awareness (although in a loose sense they can be described as ‘introspective’). Rather, they issue in awareness of properties of the body.

12 See Balci et al. (2009) for evidence that mice, too, are capable of making swift and accurate assessments of risk. See also Gallistel et al. (2001) for evidence that rats are excellent at tracking random changes in the probability of reward.

1 or bad (attractive or aversive), in many cases issuing in a decision (unless the subject opts to
2 engage in more explicit reflection of some sort).¹³

3 Smith (2005) makes much of the fact that humans in uncertainty-monitoring experiments
4 have response profiles that closely parallel those of the animals (see also Smith et al. 2008). Since
5 the humans report that they opt out in conditions of uncertainty because they are aware of being
6 uncertain, this is said to give us reason to attribute similar awareness to the animals. But it does
7 not. For basic forms of decision-making in humans don't employ metarepresentational aware-
8 ness, as we have seen. So both humans and animals will experience negatively valenced forms of
9 anxiety directed toward the primary response options, resulting from an appraisal of low likeli-
10 hood of success. (The latter in turn is grounded in the low degree of belief that attaches to the
11 categorization or judgement underlying the required discrimination.) This will make those
12 options appear bad or mildly aversive. Such perceptions, when strong enough, will leave the opt-
13 out option as the better-seeming alternative. All of this is entirely non-metarepresentational, as
14 we have noted. But humans, with their highly-developed mindreading capacities, will categorize
15 the state they are in *as* a feeling of uncertainty, either automatically or when asked to explain their
16 choice. This categorization might play no role, however, in their basic decision-making behaviour
17 (unless it is first articulated and treated as a commitment). Indeed, their metacognizing might be
18 largely post hoc.

19 What we suggest, then, is that in humans both uncertainty and its influence on behaviour
20 should be dissociable from metarepresentational awareness of uncertainty. To the best of our
21 knowledge this has not been directly tested. But we predict that subjects who have difficulties with
22 mindreading (including those suffering from autism or schizophrenia) might show capacities to
23 make adaptive use of the opt-out key in uncertainty-responding experiments that are spared in
24 comparison with their capacity to identify themselves *as* uncertain. For example, in one condition
25 subjects might perform the task without making any explicit metacognitive judgements, whereas
26 in another they might be required to make such a judgement before deciding whether or not to
27 opt out. Our prediction is that performance in the former condition should be significantly better
28 than performance in the latter, in these populations.

29 **Further consequences of the accounts**

30 Notice that both of the affect-based explanations mooted here make significant executive demands
31 on the animals in question. In order for feelings of uncertainty to be used as cues to opt out, they
32 have to be attended to. And in order for one to feel anxious at the thought of taking a particular
33 action, that action has to be mentally rehearsed. We should predict, then, that the animals are
34 unlikely to make adaptive use of opt-out behaviour in cases where they are required to execute
35 some concurrent task. Note that this prediction is not made by the degrees-of-belief account
36 alone (independent of any role for epistemic emotions). However, it is also a prediction of the
37 metarepresentational account. So the finding that use of the opt-out response diminishes when
38 animals are required to engage in an ancillary task (Smith 2011), does nothing to support a
39 metarepresentational account of uncertainty monitoring over its affect-based competitors.

40 Moreover, each of the affect-based accounts makes the following empirical prediction. Mood
41 manipulations that are effective in reducing anxiety, or drugs that produce such an outcome,
42 should significantly reduce the extent to which animals opt out in conditions of uncertainty.

¹³ Note that this account deviates slightly from that provided by Damasio (1994), who places more empha-
sis on the arousal and other bodily components of affect, rather than on the valence component as we do
here. For discussion and defence, see Carruthers (2011). And note, too, that even if arousal *is* involved it
can be quite subtle, perhaps depending on what Damasio calls '*as if* affect.'

1 In contrast, the degrees-of-belief account fails to make any such prediction. For it is purely cog-
 2 nitive in nature. Moreover, any metacognitive account that is cast in purely cognitive terms (merely
 3 maintaining that the animals are aware of their uncertainty, for example) will likewise fail to make
 4 such a prediction. However, mood manipulations, even if successful, would not necessarily sup-
 5 port an affect-based account of uncertainty-monitoring behaviour over a metarepresentational
 6 one. For metacognitive theorists can presumably claim that what is represented is an *emotional*
 7 state of uncertainty, and in that case manipulations that reduce anxiety will have the effect of
 8 making it harder to monitor and metarepresent the relevant state.

9 None of the tests that have been employed to date are capable of discriminating between
 10 metarepresentational and non-metarepresentational explanations of uncertainty-monitoring
 11 behaviour. So we are forced to fall back on indirect reasons that might favour one or other kind
 12 of explanation, of the sort that have been in play up to now. Some further considerations of this
 13 kind will form the topic of the next section.

14 **Species differences and individual variation**

15 Smith (2005) and others have argued that differences in uncertainty-monitoring behaviour
 16 between species support a metarepresentational account. In this section we challenge this inter-
 17 pretation, while also arguing that individual differences in such behaviour among humans may be
 18 problematic for first-person-based accounts to accommodate.

19 **Species differences**

20 We agree that differences in uncertainty monitoring behaviour across species favour a metarep-
 21 resentational account over an associative learning competitor, since the species that fail in these
 22 tasks (rats and pigeons) excel at such learning (Smith 2005; Smith et al. 2009). But they don't
 23 support a metarepresentational account over either of the affectively-based proposals discussed
 24 earlier in the 'Affective explanations of the evidence' section. This is because there may be species
 25 differences in the extent to which anxiety is created in foraging situations, or differences in the
 26 extent to which members of a given species pay attention to or notice their own bodily feelings,
 27 or differences in capacities to engage in mental rehearsal of action. None of these differences is yet
 28 confirmed. But until they are ruled out, we have no positive reason to believe that the difference
 29 between the species is a metarepresentational one.

30 It might be claimed that differences in uncertainty-monitoring behaviour among distinct spe-
 31 cies of monkey provide a greater challenge for non-metarepresentational accounts (Beran et al.
 32 2009; cf. Basile et al. 2009). Capuchin monkeys, in particular, rarely if ever make use of the opt-
 33 out response, even after numerous trials, and even under conditions designed to bias the mon-
 34 keys toward using the opt-out response. Macaque monkeys, in contrast, show response profiles
 35 that closely parallel those of humans. It should be obvious from the previous discussion, however,
 36 that there are multiple types of resource that could potentially be used to explain these differences
 37 without appealing to metarepresentational capacities, and some of these explanations are
 38 independently plausible.¹⁴

.....
 14 We note that Beran et al. (2009) themselves offer accounts of the failure of capuchins in these tasks that
 don't seem to depend on an absence of metarepresentational capacities. They suggest, for example, that
 capuchins may lack the ability to appreciate the abstract and indirect benefit of selecting the uncertainty
 response to maximize reward, leading them to focus on the primary, directly rewarding, options. If this is
 transposed into a positive account of macaques' success in these tasks, then the account is no longer a
 metacognitive one. For appreciating an indirect benefit need not require metarepresentation.

1 It is possible that the two species differ in the extent to which they are apt to experience anxiety
 2 in foraging situations. In particular, if capuchins feel little or no anxiety when confronted with a
 3 difficult discrimination task to gain a food reward, then they will not be motivated to use the opt-
 4 out key. If macaques are more like humans in this respect, however, then the primary response
 5 options will be experienced as aversive in cases of difficult discrimination, making it more likely
 6 that the animals will use the opt-out response. Alternatively, capuchins might experience anxiety,
 7 but not know what to do with it (i.e. what control operation to adopt). (Compare people who are
 8 used to dealing with high degrees of anxiety and those who are not.) These suggestions could be
 9 motivated by the following ecological facts.

10 Capuchins are arboreal, living locally in forest environments that provide ample sources of
 11 fruits, nuts, leaves, and insects that constitute their primary diet. Although they experience food
 12 competition within groups, adults are known to share food with unrelated infants, and adults will
 13 often share food with one another (De Waal 2000). Macaques, in contrast, are often semi-nomadic
 14 with broad ranges, and have colonized a wide set of ecologies, with the largest distribution of any
 15 non-human primate genus (Fleagle 1998). Illustrating their flexibility in adapting to new envi-
 16 ronments, ‘weed’ macaques (such as the rhesus macaque) have been able to thrive in human
 17 environments (Richard et al. 1989). Although they, too, are omnivorous, they are subject to
 18 intense food competition within groups (Sterck and Steenbeek 1997). It would not be surprising,
 19 then, that they might have become adapted to experience and deal with anxiety in difficult forag-
 20 ing situations, since they face far more uncertainties when foraging than do capuchins.

21 Individual differences

22 Smith (2005) also notes that both humans and the other primates in these experiments display
 23 similar ranges of individual difference. Some people, and some animals, never make use of
 24 the opt-out key, and confine themselves to the primary response options, whereas others opt out
 25 adaptively in circumstances where they are likely to make (or to have made) a mistake. It is
 26 unclear why this should be thought to support a metarepresentational account, however. (Indeed,
 27 we will suggest in a moment that it may cause problems for that account.) In any case, each of the
 28 two affect-based theories is capable of explaining this fact.

29 In the first place, it is well known that there are chronic differences among people in the extent
 30 to which they pay attention to the bodily—arousal—component of their emotional states (Barrett
 31 1998; Gasper and Clore 2000; Barrett et al. 2004), and one might expect the same to be true of
 32 other primates. Such individuals are unlikely to notice the manifestations of their own state of
 33 uncertainty, and so will be less likely to learn to use them as cues to opt out. It is also well known
 34 that there are chronic differences between people (and presumably other primates) in the extent
 35 to which they become anxious in everyday situations. Those who aren’t easily made anxious will
 36 fail to see the primary response options as bad or aversive, and so will lack any motivation to use
 37 the opt-out response, whereas those who are more easily made anxious will opt out more often.

38 In contrast, while all views might predict that there will be individual differences in the *extent*
 39 to which people make use of the opt-out option, the fact that some people (and animals) almost
 40 *never* employ it is harder for a metacognitive theorist to accommodate. *How* hard it is will be a
 41 function of the proportion of subjects who never opt out. If such people are rare, then they might
 42 be considered the tail-ends of a normal distribution. But if they are numerous, and the normal
 43 distribution curve is not very steep, then this will be more problematic. For recall that the concep-
 44 tual and inferential resources necessary to monitor one’s own mental states are claimed to have
 45 been selected for precisely because of the adaptive advantages that they yield in situations like this.
 46 It would therefore be puzzling if there should turn out to be many individuals who nevertheless

1 fail to make use of those resources. (It would much as if we found a significant proportion of
 2 people who never make use of episodic memory.) For there is surely just as much need for people
 3 to monitor uncertainty as there ever was in our evolutionary past. If metarepresentational resources
 4 evolved, in part, to enable animals to monitor their own uncertainty and respond adaptively, then
 5 one would expect that those resources would be regularly and reliably employed by the vast
 6 majority of normal individuals. The affect-based accounts, in contrast, can appeal to widespread
 7 individual differences that are already known to exist.

8 It seems, then, that in the absence of a direct experimental test, there are no indirect reasons to
 9 favour a metarepresentational account of the uncertainty-monitoring data over its affect-based
 10 competitors; indeed, there are some reasons to prefer the latter.¹⁵

11 Conclusion

12 We conclude that existing uncertainty-monitoring experiments with non-human primates fail to
 13 discriminate between a metacognitive (metarepresentational) account and those that rely on
 14 non-metarepresentational uses of feelings of uncertainty. Until experiments that might tease
 15 apart these differing explanations have been done, a metarepresentational account of the uncer-
 16 tainty-monitoring data is unsupported. As a result, while we have good reason to think that these
 17 animals are capable of taking executively controlled decisions in many ways like our own, we
 18 presently have no reason to prefer a first-person-based account of the evolutionary emergence of
 19 metacognition over its mindreading-based competitor.

20 Acknowledgements

21 We are grateful to Michael Beran and an anonymous reviewer for their insightful comments on
 22 an earlier draft of this chapter.

23 References

- 24 Alter, A., Oppenheimer, D., Epley, N., and Eyre, R. (2007). Overcoming intuition: Metacognitive difficulty
 25 activates analytic reasoning. *Journal of Experimental Psychology: General*, 136, 569–76.
- 26 Amiez, C., Procyk, E., Honoré, J., Sequeira, H., and Joseph, J.-P. (2003). Reward anticipation, cognition,
 27 and electrodermal activity in the conditioned monkey. *Experimental Brain Research*, 149, 267–75.
- 28 Baars, B. (1988). *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.
- 29 Balci, F., Freestone, D., and Gallistel, C. R. (2009). Risk assessment in man and mouse. *Proceedings of the*
 30 *National Academy of Sciences*, 106, 2459–63.
- 31 Baron-Cohen, S. (1995). *Mindblindness*. Cambridge, MA: MIT Press.
- 32 Barrett, L. (1998). Discrete emotions or dimensions? The role of valence focus and arousal focus. *Cognition*
 33 *and Emotion*, 12, 579–99.
- 34 Barrett, L., Quigley, K., Bliss-Moreau, E., and Aronson, K. (2004). Interoceptive sensitivity and self-reports
 35 of emotional experience. *Journal of Personality and Social Psychology*, 87, 684–97.
- 36 Basile, B., Hampton, R., Suomi, S., and Murray, E. (2009). An assessment of memory awareness in tufted
 37 capuchin monkeys (*Cebus apella*). *Animal Cognition*, 12, 169–80.

¹⁵ There is one final possibility that has not been discussed here. This is that epistemic emotions like uncer-
 tainty are *non-conceptual* forms of first-person metarepresentation in virtue of their function of inform-
 ing us of the underlying risk of epistemic failure. See Proust (2009a, 2009b) for defence of a view of this
 sort. (Note, however, that Proust herself declines to use the language of ‘metarepresentation’ in this con-
 nection.) For a critique of this idea see Carruthers (2011).

- 1 Bechara, A., Damasio, A., Damasio, H., and Anderson, S. (1994). Insensitivity to future consequences
2 following damage to human prefrontal cortex. *Cognition*, 50, 7–15.
- 3 Beran, M., Smith, J., Coutinho, M., Couchman, J., and Boomer, J. (2009). The psychological organization
4 of ‘uncertainty’ responses and ‘middle’ responses: A dissociation in capuchin monkeys (*Cebus apella*).
5 *Journal of Experimental Psychology: Animal Behavior Processes*, 35, 371–81.
- 6 Bugnyar, T. and Heinrich, B. (2005). Food-storing ravens differentiate between knowledgeable and
7 ignorant competitors. *Proceedings of the Royal Society of London B*, 272, 1641–6.
- 8 Bugnyar, T. and Heinrich, B. (2006). Pilfering ravens, *Corvus corax*, adjust their behavior to social context
9 and identity of competitors. *Animal Cognition*, 9, 369–76.
- 10 Bugnyar, T., Stöwe, M., and Heinrich, B. (2007). The ontogeny of caching in ravens. *Animal Behavior*, 74,
11 757–67.
- 12 Buttelmann, D., Carpenter, M., Call, J., and Tomasello, M. (2007). Enculturated chimpanzees imitate
13 rationally. *Developmental Science*, 10, F31–38.
- 14 Buttelmann, D., Call, J., and Tomasello, M. (2009a). Do great apes use emotional expressions to infer
15 desires? *Developmental Science*, 12, 688–98.
- 16 Buttelmann, D., Carpenter, M., and Tomasello, M. (2009b). Eighteen-month-old infants show false belief
17 understanding in an active helping paradigm. *Cognition*, 112, 337–42.
- 18 Byrne, R. and Whiten, A. (Eds.) (1988). *Machiavellian Intelligence*. Oxford: Oxford University Press.
- 19 Byrne, R. and Whiten, A. (Eds.) (1997). *Machiavellian Intelligence II*. Cambridge: Cambridge
20 University Press.
- 21 Call, J. and Carpenter, M. (2001). Do apes and children know what they have seen? *Animal Cognition*, 4,
22 207–20.
- 23 Carruthers, P. (2006). *The Architecture of the Mind*. Oxford: Oxford University Press.
- 24 Carruthers, P. (2008). Meta-cognition in animals: A skeptical look. *Mind and Language*, 23, 58–89.
- 25 Carruthers, P. (2011). *The Opacity of Mind*. Oxford: Oxford University Press.
- 26 Carruthers, P. (forthcoming). Mindreading in infancy.
- 27 Couchman, J., Coutinho, M., Beran, M., and Smith, D. (2009). Metacognition is prior. *Behavioral and*
28 *Brain Sciences*, 32, 142.
- 29 Couchman, J., Coutinho, M., Beran, M., and Smith, J. D. (2010). Beyond stimulus cues and reinforcement
30 signals: A new approach to animal metacognition. *Journal of Comparative Psychology*, 124, 356–68.
- 31 Dally, J., Emery, N., and Clayton, N. (2006). Food-caching western scrub-jays keep track of who was
32 watching when. *Science*, 312, 1662–5.
- 33 Dally, J., Emery, N., and Clayton, N. (2009). Avian theory of mind and counter espionage by food-caching
34 western scrub-jays. *European Journal of Developmental Psychology*, 7, 17–37.
- 35 Damasio, A. (1994). *Descartes’ Error*. London: Papermac.
- 36 De Waal, F. (2000). Attitudinal reciprocity in food sharing among brown capuchin monkeys. *Animal*
37 *Behaviour*, 60, 253–61.
- 38 Dunlosky, J. and Metcalfe, J. (2009). *Metacognition*. Thousand Oaks, CA: Sage Publications.
- 39 Dunn, B., Dalgleish, T., and Lawrence, A. (2006). The somatic marker hypothesis: A critical evaluation.
40 *Neuroscience and Biobehavioral Reviews*, 30, 239–71.
- 41 Emery, N. and Clayton, N. (2004). The mentality of crows: Convergent evolution of intelligence in corvids
42 and apes. *Science*, 306, 1903–7.
- 43 Evans, T. and Beran, M. (2007). Chimpanzees use self-distraction to cope with impulsivity. *Biology Letters*,
44 3, 599–602.
- 45 Flavell, J. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry.
46 *American Psychologist*, 34, 906–11.
- 47 Fleagle, J. (1998). *Primate Adaptation and Evolution*, 2nd Edn. New York: Academic Press.
- 48 Flombaum, J. and Santos, L. (2005). Rhesus monkeys attribute perceptions to others. *Current Biology*, 15,
49 447–52.

- 1 Frith, U. and Happé, F. (1999). Theory of mind and self-consciousness: what is it like to be autistic? *Mind*
2 *and Language*, 14, 1–22.
- 3 Gallistel, R., Mark, T., King, A., and Lantham, P. (2001). The rat approximates an ideal detector of rates of
4 reward. *Journal of Experimental Psychology: Animal Behavior Processes*, 27, 354–72.
- 5 Gasper, K. and Clore, G. (2000). Do you have to pay attention to your feelings to be influenced by them?
6 *Personality and Social Psychology Bulletin*, 26, 698–711.
- 7 Gilbert, D. and Wilson, T. (2007). Propection: Experiencing the future. *Science*, 317, 1351–4.
- 8 Goldman, A. (2006). *Simulating Minds*. New York: Oxford University Press.
- 9 Gopnik, A. and Meltzoff, A. (1997). *Words, Thoughts, and Theories*. Cambridge, MA: MIT Press.
- 10 Gordon, P. and Holyoak, K. (1983). Implicit learning and generalization of the ‘mere exposure’ effect.
11 *Journal of Personality and Social Psychology*, 45, 492–500.
- 12 Hampton, R. (2001). Rhesus monkeys know when they remember. *Proceedings of the National Academy of*
13 *Sciences of The United States of America*, 98, 5359–62.
- 14 Hampton, R. (2005). Can Rhesus monkeys discriminate between remembering and forgetting? In H.
15 Terrace and J. Metcalfe (Eds.) *The Missing Link in Cognition*, pp. 272–95. Oxford: Oxford University
16 Press.
- 17 Hampton, R., Zivin, A., and Murray, E. (2004). Rhesus monkeys (*Macaca mulatta*) discriminate between
18 knowing and not knowing and collect information as needed before acting. *Animal Cognition*,
19 7, 239–46.
- 20 Hare, B. (2007). From nonhuman to human mind: What changed and why? *Current Directions in*
21 *Psychological Science*, 16, 60–4.
- 22 Hare, B. and Tomasello, M. (2005). Human-like social skills in dogs? *Trends in Cognitive Sciences*,
23 9, 439–44.
- 24 Hare, B., Call, J., Agnetta, B., and Tomasello, M. (2000). Chimpanzees know what conspecifics do and do
25 not see. *Animal Behavior*, 59, 771–85.
- 26 Hare, B., Call, J., and Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal*
27 *Behavior*, 61, 139–51.
- 28 Hare, B., Call, J., and Tomasello, M. (2006). Chimpanzees deceive a human competitor by hiding.
29 *Cognition*, 101, 495–514.
- 30 Hart, J. (1965). Memory and the feeling of knowing experience. *Journal of Educational Psychology*,
31 56, 208–16.
- 32 Jacoby, L. (1991). A process dissociation framework: Separating automatic from intentional uses of
33 memory. *Journal of Memory and Language*, 30, 513–41.
- 34 Jeannerod, M. (2006). *Motor Cognition*. Oxford: Oxford University Press.
- 35 Kaminski, J., Call, J., and Tomasello, M. (2008). Chimpanzees know what others know, but not what they
36 believe. *Cognition*, 109, 224–34.
- 37 Koriat, A. (2000). The feeling of knowing: Some metatheoretical implications for consciousness and
38 control. *Consciousness and Cognition*, 9, 149–71.
- 39 Koriat, A., Ma’ayan, H., and Nussinson, R. (2006). The intricate relationships between monitoring and
40 control in metacognition: Lessons for the cause-and-effect relation between subjective experience and
41 behavior. *Journal of Experimental Psychology: General*, 135, 36–69.
- 42 Koriat, A., Nussinson, R., Bless, H., and Shaked, N. (2008). Information-based and experience-based
43 metacognitive judgments: Evidence from subjective confidence. In J. Dunlosky and J. Bjork (Eds.)
44 *Handbook of Metamemory and Memory*, pp. 117–34. Mahwah, NJ: Erlbaum.
- 45 Kornell, N., Son, L., and Terrace, H. (2007). Transfer of metacognitive skills and hint seeking in monkeys.
46 *Psychological Science*, 18, 64–71.
- 47 Krachun, C. and Call, J. (2009). Chimpanzees (*Pan troglodytes*) know what can be seen from where.
48 *Animal Cognition*, 12, 317–31.
- 49 Krachun, C., Carpenter, M., Call, J., and Tomasello, M. (2009). A competitive nonverbal false belief task for
50 children and apes. *Developmental Science*, 12, 521–35.

- 1 Leonesio, R. and Nelson, T. (1990). Do different metamemory judgments tap the same underlying aspects
2 of memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 464–70.
- 3 Leslie, A. (1994). ToMM, ToBy, and Agency: Core architecture and domain specificity. In L. Hirshfeld and
4 S. Gelman (Eds.) *Mapping the Mind*, pp. 39–67. Cambridge: Cambridge University Press.
- 5 Lin, L. and Zabrocky, K. (1998). Calibration of comprehension: Research and implications for education
6 and instruction. *Contemporary Educational Psychology*, 23, 345–91.
- 7 Maki, R. and McGuire, M. (2002). Metacognition for text: Findings and implications for education.
8 In T. Perfect and B. Schwartz (Eds.) *Applied Metacognition*, pp. 39–67. Cambridge: Cambridge
9 University Press.
- 10 Mandler, G. (1980). Recognizing: The judgment of previous occurrences. *Psychological Review*, 87, 252–71.
- 11 Melis, A., Call, J., and Tomasello, M. (2006). Chimpanzees (*Pan troglodytes*) conceal visual and auditory
12 information from others. *Journal of Comparative Psychology*, 120, 154–62.
- 13 Metcalfe, J. (2008). Evolution of metacognition. In Dunlosky, J. and Bjork, J. (eds.) *Handbook of*
14 *Metamemory and Memory*, pp. 29–46. New York: Psychology Press.
- 15 Mulcahy, N. and Call, J. (2006). Apes save tools for future use. *Science*, 312, 1038–40.
- 16 Nelson, T. and Narens, L. (1990). Metamemory: a theoretical framework and new findings. In G. Bower
17 (Ed.) *The Psychology of Learning and Information* (Vol. 26), pp. 125–73. London: Academic Press.
- 18 Nichols, S. and Stich, S. (2003). *Mindreading*. New York: Oxford University Press.
- 19 O’Connell, S. and Dunbar, R. (2003). A test for comprehension of false belief in chimpanzees.
20 *Evolution and Cognition*, 9, 131–40.
- 21 Panksepp, J. (2005). Affective consciousness: Core emotional feelings in animals and humans.
22 *Consciousness and Cognition*, 14, 30–80.
- 23 Proust, J. (2009a). The representational basis of brute metacognition: A proposal. In R. Lurz (Ed.)
24 *The Philosophy of Animal Minds*, pp. 165–83. Cambridge: Cambridge University Press.
- 25 Proust, J. (2009b). Overlooking metacognitive experience. *Behavioral and Brain Sciences*, 32, 158–9.
- 26 Richard, A., Goldstein, S., and Dewar, R. (1989). Weed macaques: The evolutionary implications of
27 macaque feeding ecology. *International Journal of Primatology*, 10, 569–94.
- 28 Santos, L., Nissen, A., and Ferrugia, J. (2006). Rhesus monkeys (*Macaca mulatta*) know what others can
29 and cannot hear. *Animal Behavior*, 71, 1175–81.
- 30 Sanz, C., Morgan, D., and Gulick, S. (2004). New insights into chimpanzees, tools, and termites from the
31 Congo basin. *American Naturalist*, 164, 567–81.
- 32 Scott, R. and Baillargeon, R. (2009). Which penguin is this? Attributing false beliefs about object identity at
33 18 months. *Child Development*, 80, 1172–96.
- 34 Scott, R., Baillargeon, R., Song, H., and Leslie, A. (2010). Attributing false beliefs about non-obvious
35 properties at 18 months. *Cognitive Psychology*, 61, 366–95.
- 36 Smith, J. D. (2005). Studies of uncertainty monitoring and metacognition in animals and humans.
37 In H. Terrace and J. Metcalfe (Eds.) *The Missing Link in Cognition*. Oxford: Oxford University Press.
- 38 Smith, J. D. (2011). Presentation at the American Association for the Advancement of Science,
39 Washington DC, 20 February 2011.
- 40 Smith, J. D., Beran, M., Redford, J., and Washburn, D. (2006). Dissociating uncertainty responses and
41 reinforcement signals in the comparative study of uncertainty monitoring. *Journal of Experimental*
42 *Psychology: General*, 135, 282–97.
- 43 Smith, J. D., Beran, M., Couchman, J., and Coutinho, M. (2008). The comparative study of metacognition:
44 Sharper paradigms, safer inferences. *Psychonomic Bulletin & Review*, 15, 679–91.
- 45 Smith, J. D., Beran, M., Couchman, J., Coutinho, M., and Boomer, J. (2009). The curious incident of the
46 capuchins. *Comparative Cognition & Behavior Reviews*, 4, 61–4.
- 47 Smith, J. D., Redford, J., Beran, M., and Washburn, D. (2010). Rhesus monkeys (*Macaca mulatta*)
48 adaptively monitor uncertainty while multi-tasking. *Animal Cognition*, 13, 93–101.

- 1 Smith, J. D., Shields, W., and Washburn, D. (2003). The comparative psychology of uncertainty monitoring
2 and meta-cognition. *Behavioral and Brain Sciences*, 26, 317–73.
- 3 Son, L. and Kornell, N. (2005). Meta-confidence judgments in rhesus macaques: Explicit versus implicit
4 mechanisms. In H. Terrace and J. Metcalfe (Eds.) *The Missing Link in Cognition*, pp. 296–320. Oxford:
5 Oxford University Press.
- 6 Song, H. and Baillargeon, R. (2008). Infants' reasoning about others' false perceptions. *Developmental*
7 *Psychology*, 44, 1789–95.
- 8 Song, H., Onishi, K., Baillargeon, R., and Fisher, C. (2008). Can an actor's false belief be corrected by an
9 appropriate communication? Psychological reasoning in 18.5-month-old infants. *Cognition*,
10 109, 295–315.
- 11 Southgate, V., Senju, A., and Csibra, G. (2007). Action anticipation through attribution of false belief by
12 2-year-olds. *Psychological Science*, 18, 587–92.
- 13 Southgate, V., Chevallier, C., and Csibra, G. (2010). Seventeen-month-olds appeal to false beliefs to
14 interpret others' referential communication. *Developmental Science*, 13, 907–12.
- 15 Stanovich, K. (2009). *What Intelligence Tests Miss: The Psychology of Rational Thought*. New Haven, CT:
16 Yale University Press.
- 17 Stanovich, K. and West, R. (2000). Individual differences in reasoning: Implications for the rationality
18 debate. *Behavioral and Brain Sciences*, 23, 645–726.
- 19 Sterck, E. and Steenbeek, R. (1997). Female dominance relationships and food competition in the
20 Sympatric Thomas Langur and long-tailed Macaque. *Behavior*, 134, 749–74.
- 21 Stulp, G., Emery, N., Verhulst, S., and Clayton, N. (2009). Western scrub-jays conceal auditory information
22 when competitors can hear but cannot see. *Biology Letters*, 5, 583–5.
- 23 Surian, L., Caldi, S., and Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological*
24 *Science*, 18, 580–6.
- 25 Udell, M., Dorey, N., and Wynne, C. (2008). Wolves outperform dogs in following human social cues.
26 *Animal Behavior*, 76, 1767–73.
- 27 Washburn, D., Gulledge, J., Beran, M., and Smith, J. D. (2010). With his memory magnetically erased, a
28 monkey knows he is uncertain. *Biology Letters*, 6, 160–2.
- 29 Wellman, H. (1990). *The Child's Theory of Mind*. Cambridge, MA: MIT Press.
- 30 Winkielman, P. and Cacioppo, J. (2001). Mind at ease puts a smile on the face: Psychophysiological
31 evidence that processing facilitation leads to positive affect. *Journal of Personality and Social Psychology*,
32 81, 989–1000.
- 33 Winkielman, P., Berridge, K., and Wilbarger, J. (2005). Unconscious affective reactions to masked happy
34 versus angry faces influence consumption behavior and judgments of value. *Personality and Social*
35 *Psychology Bulletin*, 31, 121–35.
- 36 Wolpert, D. and Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nature*
37 *Neuroscience*, 3, 1212–17.
- 38 Wolpert, D. and Kawato, M. (1998). Multiple paired forward and inverse models for motor control.
39 *Neural Networks*, 11, 1317–29.