

1 Chapter 18

2 **Anoetic, noetic, and**
 3 **autonoetic metacognition**

4 Janet Metcalfe and Lisa K. Son

5 **Introduction**

6 Metacognition can take many guises. Consider, first, one contestant of several, W., playing a tel-
 7 evision game show that tests general knowledge by presenting whimsically phrased cues. As a
 8 question ‘What is Will’s Quill?’ is displayed on the screen, W. very quickly retrieves bits of infor-
 9 mation regarding what may possibly be or be related to the answer, based on the question and his
 10 understanding of the natural language associates to it. He accumulates the fragmentary informa-
 11 tion resulting from his memory search quickly as the clock ticks. If the information count reaches
 12 a criterion, but one far less than is necessary for complete access to the answer, W. buzzes in to
 13 beat out his opponents and to indicate that he thinks that he *will know* the answer given an addi-
 14 tional 5 seconds, even though he does not know it yet. If the accumulation of partial information
 15 does not reach criterion, W. declines to respond, letting the opposition buzz in, instead. Using
 16 this ‘game-show’ strategy (Reder 1987) based on the metacognitive feeling that he will know, W.
 17 is nearly always—roughly 85% of the time—able to come up with the answer later when he thinks
 18 he will be able to do so. And, by combining his encyclopaedic knowledge, his lightening speed,
 19 and his sophisticated metacognitive strategy, W. becomes the new world Jeopardy champion.

20 Now imagine L., who is playing a memory gambling game. He is presented with the target—a
 21 complex picture—in a flash on the screen. The picture disappears, and nine alternative pictures
 22 appear on the screen simultaneously. L. looks through them considering each in turn and upon
 23 seeing what he thinks is the target picture in the array, he touches it, and they all disappear. Then,
 24 though, he has to give his confidence in his answer. He can either ‘pass’—choose not to wager—
 25 or he can ‘double down’—wager big. Two betting icons appear on the screen. Nothing further
 26 will happen until he makes this retrospective decision about whether he thinks he was right or
 27 wrong. In this case, L. chooses the ‘double down’ icon, and he wins three tokens, which fall into
 28 his hopper, to be redeemed later when he has accumulated enough tokens for a prize. Had he
 29 pushed the ‘pass’ icon, he would have gotten only one token. But had he touched the wrong pic-
 30 ture in the 9-option task, and then ‘doubled down’, he would have seen three tokens fly out of his
 31 hopper and disappear. L. is known, by other gamblers, as having a *serious* emotional reaction
 32 when this happens. But, fortunately, it doesn’t happen often. And he does get paid off with prizes
 33 from time to time. Like other gamblers, L. is happy to play this game of making metacognitive
 34 bets on his own memory hour after hour, day after day.

35 Finally, imagine S. trying to retrieve the name of the famous Canadian author who wrote *The*
 36 *Last Spike*. A nagging feeling of having the answer right on the tip of his tongue plagues S. But S.
 37 cannot retrieve the answer no matter how hard he tries, and he is trying hard. His friends tell him
 38 to give up. None of them are Canadians, and they neither know the answer, nor care, to be sure. But
 39 S. refuses to listen. His mind is screaming with this impossible-to-resist emotional premonition

1 that the answer is imminent (see Schwartz and Metcalfe, 2011). And he is right, statistically, at
 2 least. When people have this feeling, they nearly always get the answer eventually. But it is hours,
 3 not moments away. Having been driven almost to distraction by this tantalizing gap in his knowl-
 4 edge, and knowing that the answer, oddly, is ‘almost’ a French name, and that the first letter of
 5 that first name is P, finally, in a flash of insight the answer—Pierre Berton—appears, seemingly
 6 unbidden out of the blue (previous intense efforts to find it notwithstanding).

7 Which one of the above individuals is metacognitive? Which was making an assessment
 8 about an internal representation? Which, by virtue of this metacognitive reflection, has a self?
 9 Insofar as all three of these cases represent what many researchers in the field affirm as true met-
 10 acognition—knowing about what one knows—then, it would seem that the case could be made
 11 that all three of them involve these characteristics and each of W., L., and S. exhibit self-awareness.
 12 Indeed, a number of distinguished thinkers have forwarded the idea that a central reason for
 13 interest in metacognition, above and beyond its functional usefulness in allowing people better
 14 control of their thinking and their action, is that metacognition is the key to a special kind of
 15 human self-reflective consciousness that is the very essence of our humanness.

16 Metacognition, by this view, is thought to be what we might call self-perspectival (see Descartes
 17 1637; Husserl 1929; Searle 1992). The emphasis on the relation of metacognition to the self
 18 undoubtedly stars the work of Descartes, who reflected about his reflections and perceptions, and
 19 in so doing made the claim—that he certainly believed was self-evident and irrefutable—that the
 20 fact that he was able to do this reflection provided incontrovertible evidence for the self. ‘I think
 21 therefore I am’ with the ‘I’ highlighted. The reflection gave the proof of his self. While some mod-
 22 erns, notably Bertrand Russell (1997),¹ are not so sure, it is a fascination with the *self* in self-
 23 reflection—that this kind of recursive cognition gives rise to consciousness and self-awareness
 24 and proof that an internal person exists—that provides the intellectual glitter giving studies in
 25 metacognition their panache.

26 The modern theorist most associated with this view is Rosenthal (2000). In advancing his
 27 ‘higher order thought’ (HOT) hypothesis, he argues that consciousness is essentially metacogni-
 28 tion, which, classically (see Nelson and Narens 1990) entails the reflection at the metalevel upon
 29 a lower, basic, level. Rosenthal notes: ‘The leading idea behind the HOT hypothesis is that a men-
 30 tal state is conscious only if one is, in some suitable way, conscious *of* that state ... A conscious
 31 state is a state one is conscious of oneself as being in’ (pp. 231–2). Rosenthal’s HOTs involve
 32 something more than just a metalevel reflection on a basic level representation: *self*-consciousness
 33 is implied. He does not necessarily endorse an elaborate folk-theoretic notion of what self con-
 34 sciousness entails including being explicitly conscious of oneself as the subject, or of having all of
 35 one’s conscious thoughts and experiences come together mentally. Self-consciousness could be
 36 much more pared down: ‘HOTs can, instead, represent the self in some minimal way, for exam-
 37 ple, in terms simply of a distinction between oneself and everything else’. But, even though mini-
 38 mal, some form of self-consciousness is implied. Furthermore, Rosenthal says that such
 39 consciousness can only be found in creatures; presumably, computers need not apply. But, per-
 40 haps nonhuman animals could.

¹ Russell notes (p. 17): “‘I think, therefore I am’ says rather more than is strictly certain. It might seem as though we were quite sure of being the same person to-day as we were yesterday, and this is no doubt true in some sense. But the real Self is as hard to arrive at as the real table and does not seem to have that absolute, convincing certainty that belongs to particular experiences. When I look at my table and see a certain brown colour, what is quite certain at once is not “I am seeing a brown colour”, but rather, “a brown colour is being seen”. This of course involves something (or somebody) which (or who) sees the brown colour; but it does not of itself involve that more or less permanent person whom we call “I”.’

1 Animal metacognition researchers almost invariably allude to the self-awareness aspect of
 2 metacognition in motivating their investigations of whether animals might be able to do meta-
 3 cognitive tasks. For example, Smith et al. (2009) justify their research on animals by saying:
 4 ‘Metacognition is linked to self-awareness ... because doubt is so personal and self-oriented.
 5 Metacognition is linked to declarative consciousness, because we can introspect and declare states
 6 of knowing. Thus, metacognition is a sophisticated capacity in humans that might be uniquely
 7 human’ (p. 40). Smith (2009) says ‘one of comparative psychology’s current goals is to establish
 8 whether nonhuman animals (hereafter, animals) share humans’ metacognitive capacity. If they
 9 do, it could bear on their consciousness and self-awareness too’ (p. 389). Foote and Crystal
 10 (2007), who investigated metacognition in rats, say ‘People are sometimes aware of their own
 11 cognitive processes. Therefore, studies in metacognition test the hypothesis that animals behave
 12 functionally the same as an organism that is aware of its own cognitive state’ (p. 1).

13 And, while, if W., L., and S. were all people, we would have no qualms about admitting that the
 14 stream and quality of the metacognitive thought processes would allow us to attribute selfhood to
 15 each—they ‘feel’ like people—when we realize that two of these three were not even humans, we
 16 might balk at this conclusion. And, indeed, W. in our earlier example, is Watson, the IBM com-
 17 puter who recently made front page news by beating out previous Jeopardy champions to become
 18 the new world champion. The feat is impressive, but does it imply that W. is conscious and has a
 19 self? And L. is Lashley, a rhesus monkey. S. is human, with the initial chosen for ‘Self.’ In that light,
 20 S.’s musings about his tip-of-the-tongue state leave little doubt, in most people’s minds, that he
 21 has mind, consciousness, and self-awareness. But while, intuitively, we reject the idea that Watson
 22 might have a self, and remain agnostic about Lashley (while perhaps swayed toward the possibil-
 23 ity by the metacognitive data), the question remains: if the evidence for self awareness is metacog-
 24 nition, why do we accept that evidence for Self but not for Watson? Perhaps we are merely
 25 exhibiting an anthropocentric bias, and the impressive performance on the metacognitive tasks,
 26 by all three actors, should mean that we should, rationally, be compelled to abandon our preju-
 27 dices against machine or monkey and attribute consciousness and a self to all three. One possibil-
 28 ity, though, which we explore in this essay, is that perhaps it is only *certain* metacognitive tasks,
 29 with *particular* characteristics that imply high-level consciousness and selfhood. We will here
 30 endeavour to analyse tasks that have been labelled as ‘metacognitive’ into three different levels,
 31 borrowed from Tulving’s (1985) analysis of different levels of consciousness: *anoetic*, *noetic*, and
 32 *autonoetic*.

33 **Three levels of consciousness and metacognition**

34 Before analysing various metacognitive tasks we will first review Tulving’s (1984; Wheeler et al.
 35 1997; Rosenbaum et al. 2005) distinction between three different levels of consciousness.

36 **Anoetic consciousness**

37 At the lowest level, Tulving defines anoetic consciousness as a state that is temporally and spatially
 38 bound to the current time. Although it is a kind of consciousness, it is not one that allows escape
 39 in any way from the here and now, and so an animal functioning at this level of consciousness is
 40 stimulus bound. A judgement that refers to something in the world even though that something
 41 is interpreted through the viewer’s perceptual biases and learning would, then, be anoetic. Thus,
 42 if a person were learning to discriminate between Pinot Gris and Pinot Grigio, for example, and
 43 made judgements, based on tastes of various wine samples, these judgements—being about
 44 something in the world, even though the internal percept experienced is, undoubtedly, biased by
 45 the learning mind—would be anoetic. Note that while mental processes and past discrimination

1 learning may interact with just what the subject perceives (we make no claim that perception is
 2 naive) the percept, itself, is bound to the moment. It is not a representation or a memory of Pinot
 3 Grigio, but rather the percept of the wine itself that is being judged (and so is neither a judgement
 4 about an internal representation nor, indeed, is it a judgement about the judgement). By some
 5 definitions (see Metcalfe and Kober 2005; Carruthers 2011) a judgement at this level would not
 6 be considered metacognitive at all. It would simply be a judgement about the world as perceived.
 7 But other researchers (e.g. Reder and Schunn 1996; Smith 2009) have labelled such judgements
 8 metacognitive. The framework specified by Nelson and Narens (1990), proposed that there are at
 9 least two levels of cognition interacting to form a metacognitive system, a basic level and a meta-
 10 level. The basic level, in this anoetic case, would not be a representation at all, however, but rather
 11 a percept, and so it is not clear that the word meta-‘cognition’, should be applied to judgements,
 12 such as these, concerned with percepts. They might better be called metaperceptual. But perhaps
 13 to overcome the definitional disputes about whether judgements about objects or events in the
 14 world as perceived by the subject are metacognitive, and, hopefully to forward our understanding
 15 of whether or not self-awareness is involved, we could agree to call such judgements anoetic
 16 metacognition. Anoetic consciousness, of course, makes no reference to the self. Similarly, anoetic
 17 metacognition could not be considered to involve self-awareness.

18 **Noetic consciousness**

19 This kind of consciousness involves internal representations, and is associated with semantic
 20 memory. It allows an organism to be aware of, and to cognitively operate on, objects and events,
 21 as well as relations among objects and events, in the absence of the physical presence of those
 22 objects and events. Noetic metacognition would be a judgement that is made about a representa-
 23 tion. The object on which the judgement is made has to be mental and internal rather than
 24 physically present, to qualify as being noetic rather than anoetic. To our knowledge, all research-
 25 ers agree to call such judgements about mental representations metacognition. However, noetic
 26 consciousness, while a form of consciousness as the name implies, does not necessarily involve
 27 the self or anything self-referential.

28 **Autonoetic consciousness**

29 This is the highest form of consciousness and is self-reflective or self-knowing. For the first time,
 30 the self, then, is intimately involved. This level of consciousness is often, in Tulving’s framework,
 31 related to human adult episodic memory, which may involve mental time travel of the self.
 32 Autonoetic consciousness is thought to be necessary for the remembering of personally experi-
 33 enced events, as long as the memory of those events is self-referential. An individual could not
 34 remember something that they experienced in a noetic manner, if they did not know that they
 35 had explicitly experienced it, as has been shown to be the case with certain amnesic patients, such
 36 as K.C., who are thought to lack autonoetic memory (Rosenbaum et al. 2005). But when a normal
 37 person remembers an event in which they participated, he or she is normally thought to be aware
 38 of the event as a veridical (or sometimes non-veridical) part of his own past existence, and the
 39 involvement of the self is a necessary component in this kind of consciousness. Autonoetic con-
 40 sciousness is not mere depersonalized knowledge. Rather, as James (1890) says: ‘this central part
 41 of the Self is *felt* ... and no *mere* summation of memories or *mere* sound of a word in our ears. It
 42 is something with which we also have direct sensible acquaintance, and which is as fully present
 43 at any moment of consciousness in which it *is* present, as in a whole lifetime of such moments’
 44 (p. 299). A normal healthy person who possesses autonoetic consciousness is capable of becom-
 45 ing aware of her own projected future as well as her own past; she is capable of mental time travel,

1 roaming at will over what has happened as readily as over what might happen, independently of
 2 physical laws that govern the universe. According to Tulving (2005) only humans past infancy
 3 possess auto-noetic consciousness.

4 Do any kind of metacognitive judgements necessarily involve auto-noetic consciousness? It
 5 would seem that if the judgement makes specific reference to the self it would qualify. A metacog-
 6 nition at the auto-noetic level might also be a judgement about one's own personal memories of
 7 one's own personal past. From the standpoint of relating metacognition to self-awareness, then,
 8 these particular kinds of metacognitions, if there are any such, are of particular importance, since
 9 it is only these that involve self-consciousness.

10 In the sections that follow we will sort metacognitive tasks that have been conducted, both in
 11 humans and in animals, into anoetic, noetic, and auto-noetic metacognition, with the view to
 12 clarifying the use of this reflective (but perhaps not *self*-reflective) processing as a litmus test for
 13 ascertaining whether or not particular creatures and, indeed, sophisticated machines, might have
 14 self-awareness.

15 **Anoetic metacognition: stimulus-driven judgements**

16 The lowest level of metacognition is anoetic. Any judgement where the individual is evaluating an
 17 external stimulus is here categorized as anoetic. Consider the simple example when judging the
 18 value of an item, say, a mug. One could say that a mug is worth £10. One's judgement of the mug
 19 changes, though, depending on who owns the mug (Kahneman et al. 1990). While the object is
 20 'endowed' with higher value when possessed by the individual (Thaler 1980), as given by his or
 21 her subjective judgement, the judgement is, nevertheless, of an external stimulus rather than a
 22 representation; it is anoetic and no self-awareness is involved. The judgement of the Pinot Grigio
 23 mentioned earlier, whether by a trained or untrained palate, also falls into this category, as do all
 24 such perceptual/categorical judgements.

25 While Foote and Crystal (2007) have argued that rats are able to reflect on their own mental
 26 processes, their task was anoetic. The experimenters had their rats learn by reinforcement to dis-
 27 criminate between the duration of two-tone classes. Then they combined this task with one in
 28 which the animals, before making the discrimination choice, could pick one response if they
 29 wanted their upcoming discrimination choice to let the response count and another (a 'pass'
 30 response) if they did not. When the stimulus duration was in the middle of the two learned classes
 31 some, but not all, of the rats chose the 'pass' response. Although arguments have been made that
 32 the entire sequence was simply a complex chain of conditioned responses (Staddon et al. 2007),
 33 even if we allowed that the rats really made a choice to take the test or not, the task is nevertheless
 34 anoetic. It was about a categorization of a stimulus in the world not a representation and was, in
 35 no way, self-relevant.

36 Similarly, the classic 'escape' studies in dolphins are anoetic. In one such study (Smith et al.
 37 1995), dolphins were required to discriminate the auditory frequencies of two tones by respond-
 38 ing with one of two responses. If a 2100-Hz tone was sounded, the dolphin was rewarded when it
 39 responded to a '2100-Hz' icon; for all lower frequencies, the dolphin was rewarded when it
 40 responded to a '<2100-Hz' icon. An error terminated the trial without reinforcement and resulted
 41 in a punishment in the form of a time out. A response to a third 'escape' icon also terminated the
 42 trial, but with neither reward nor punishment. It simply acted as an expression of 'I'd rather opt
 43 out of this question' and moved onto a new trial. Dolphins could do this task, and sometimes
 44 chose to escape rather than take the test. Even allowing that their doing so was a judgement, it was
 45 an anoetic judgement, and hence does not imply self-awareness. Other 'escape' type studies (e.g.
 46 Smith et al. 1998; Shields et al. 2005; Washburn et al. 2010), where the probe or percept, and not
 47 a internal representation, gives rise to the judgement, would also be included as examples of

1 anoetic metacognition (see Terrace and Son (2009), for a review of yet other cases of anoetic
2 metacognition using the escape paradigm).

3 It is possible, of course, that monkeys, dolphins, and even rats, have self awareness. But none of
4 the tasks outlined in this section require it. Even those tasks that require a human to simply make
5 a judgement about the world is not evidence that people are self aware (indeed, it can be argued
6 that such a confidence judgement is not metacognitive, but simply, a memory judgement). In
7 the case of humans, however, any judgement that is categorized as ‘anoetic’ might include self
8 awareness—and thus, be truly metacognitive—given that we can make further judgements about
9 our judgements, verbally. Non-verbal animals are not as fortunate. Even if we agree that anoetic
10 metacognition *is* metacognition—a proposition that we might consider to be stretching the defi-
11 nition of metacognition to the breaking point—it is still anoetic, and does not imply anything
12 about whether or not the organism showing such a capability has a self, or can reflect upon that
13 self in any way.

14 **Noetic metacognition: judgement about an internal representation**

15 Noetic consciousness allows an organism to be aware of, and to cognitively operate on, objects
16 and events, and relations among objects and events, in the absence of those objects and events.
17 The main difference between noetic metacognition, and anoetic metacognition is that with the
18 former the judgement is made about an internal representation that is no longer present in space
19 and time, rather than about a stimulus that is present in the world.

20 Classic cue-only delayed judgements of learning are a typical case of noetic metacognitive
21 judgements. A learning event, consisting of a cue and a to-be-learned target, is presented, and
22 then at some later time, the person is given the cue and asked to make a judgement about whether
23 he or she will later be able to give evidence that they know the target. If they think they will know
24 it, they give it high judgement; if not then they give a low judgement of learning. Note, if people
25 mentally projected their selves into the future to see whether they would get the answer this
26 judgement would be considered autoanoetic. However, the data on what people actually do to
27 make this assessment suggest that they do not so mentally time travel. The most compelling evi-
28 dence for a lack of mental future projection is that people’s judgements of learning do not distin-
29 guish between whether the test will be 5 minutes or 1 year hence (Koriat et al. 2004)—a distinction
30 that would be large were people really mentally projecting into the future. What they appear to do
31 instead (Son and Metcalfe 2005; Metcalfe and Finn 2010) is first try to recognize the cue. If they
32 cannot do so, they say that they don’t know and give a fast low rating. If they do recognize it, they
33 then attempt to retrieve the target, with judgements of learning getting lower and lower the longer
34 it takes them to do so. Thus, the judgement is about the current retrievability of the cue and
35 target, and hence noetic in nature.

36 Another case of what is probably a noetic metacognitive judgement occurs in the hindsight bias
37 paradigm. After a person has made an assessment about some event and is then given feedback
38 concerning the correct answer, they are asked to remember what their earlier judgement was.
39 They tend to think that their earlier judgement was much closer to the correct answer, which they
40 now know, than it really was (Hoffrage and Pohl 2003). This reflects a hindsight bias or a ‘knew it
41 all along’ effect. Hawkins and Hastie (1990) defined hindsight as ‘a projection of new knowledge
42 into the past accompanied by a *denial that the outcome information has influenced judgement*’
43 (p. 311). In contrast to this idea, though, it seems plausible that the hindsight bias results from a
44 *lack of projection of the self back into its past state of knowing*. The failure to do the past projec-
45 tion, itself, results in the bias. If so, then the judgement is noetic: based, not on mental time travel
46 but rather on current knowledge.

1 While many experiments indicate that animals have anoetic metacognition, examples of noetic
 2 metacognition in animals are much rarer. There are two cases, however, that qualify. In a sequence
 3 of trials, Hampton's (2001) monkeys were shown a target picture to study. Then, after a short
 4 delay (which was important because it meant that the monkey had to rely on a representation
 5 rather than a stimulus currently present in the world), they saw the target picture again, along
 6 with three distractor pictures. The monkeys' task was to select the target. However, after seeing
 7 the sample and prior to receiving the test, Hampton gave the monkeys the choice of either taking
 8 the test, or opting out. On some mandatory trials, though, they had to take the test. The finding
 9 of most interest was that the monkeys were more accurate on self-selected test trials than on man-
 10 datory trials, suggesting that the monkeys opted out when they knew they did not know the answer.
 11 Crucially, they did so when no external stimuli were available as cues at the time of their decision,
 12 which means that the judgements were based on internal representation and hence were noetic.
 13 However, insofar as no self-reference was necessary, these judgements were not auto-noetic.

14 Finally, Kornell et al. (2007), asked monkeys to make retrospective judgements after they took
 15 a memory test. In one such task, monkeys performed a memory task and were then asked to
 16 'wager' on the accuracy of their memories. They first studied six images that were presented
 17 sequentially on a touch-sensitive computer screen. Then, one of the six images was presented
 18 along with eight distractors and the task was to touch the picture that was already seen in the
 19 initial exposure sequence. Once a monkey had touched his choice, he made a wager. Making a
 20 'high' wager meant that he would earn three tokens if his memory response had been correct, and
 21 lose three tokens if it had been wrong. Making a 'low' wager meant that he would earn one token,
 22 regardless of the accuracy of the memory. Tokens were accumulated at the bottom of the screen
 23 and could be exchanged for food pellets when a criterion was reached. The monkeys in this task
 24 tended to choose the 'high' icon after correct responses and the 'low' icon after incorrect responses.
 25 Moreover, they did so within the first few trials of transferring to this task (the monkeys had pre-
 26 viously been trained to respond metacognitively in other, perceptual, tasks; see Son and Kornell
 27 2005). It seems, then, that they had learned a broad metacognitive skill that could generalize to
 28 new circumstances. Crucially, the monkeys appear to have represented two internal responses: a
 29 recognition memory response and a confidence judgement, as measured by their wagers. These
 30 data do not imply that the monkeys, one of whom was Lashley, by the way, had self-awareness. They
 31 do, however, imply that the animals could monitor their confidence in their own memories—a true
 32 metacognitive judgement (for recent reviews of animal metacognition research, see Kornell
 33 (2009), Smith (2009), and Terrace and Son (2009)).

34 **The ambiguous case of Panzee the chimp: noetic or** 35 **auto-noetic metacognition?**

36 Panzee, a female chimpanzee, had been taught to use over 100 lexigrams, at the time of the
 37 'experiment' in which one keeper hid 26 food objects and seven non-food objects in a large forest
 38 field, an area that Panzee knew from her past, but had not visited in 6 years (Menzel 2005). Panzee
 39 was able to recruit the assistance of other caretakers (who knew nothing about the objects being
 40 hidden) and 'tell' them where the objects were hidden. Because these new caretakers were not
 41 aware of the 'experiment' at all, let alone where the objects were hidden, when objects were found,
 42 it was thought to be the result of Panzee's 'own initiative' (Menzel 2005, p. 199). The uninformed
 43 caretaker found all 34 objects as a result of Panzee's behaviour! And, furthermore, Panzee had
 44 indicated on her lexigram board 84% of the time, which particular item had been hidden in each
 45 location, and correctly identified these items at delays, for some items, of over 90 hours from
 46 the original hiding event. Evidence in support of metacognition was seen in Panzee's behaviour:

1 The caretaker noted and responded to Panzee's relative degree of excitement—a seemingly spon-
 2 taneous metacognition, since it directly reflected the distance to the target. Panzee kept pointing,
 3 showed intensified vocalization, shook her arm, and bobbed her head or body as the caretaker got
 4 closer to the site (see Menzel 2005, p. 202). In addition, Menzel reported that Panzee seemed to
 5 do whatever it took to catch the caretaker's attention and, only once joint attention was estab-
 6 lished, touched the lexigram corresponding to the type of object hidden, pointed outdoors, some-
 7 times went outdoors (if the caretaker followed), and continued to point manually toward the
 8 object and vocalize until the caretaker found the object. As noted by Kohler (1925), the 'time in
 9 which chimpanzees live' and whether they are able to freely mentally time travel, as auto-noetic
 10 consciousness requires, remains an open question, but it seems, from these data that Panzee
 11 could, at the very least, freely recall which one of at least 20 types of objects she had been shown
 12 at a distance and at a long delay, and that she was highly certain, and highly keyed up, of her own
 13 knowledge—a feat that begins to look a lot like human auto-noesis.

14 **Auto-noetic metacognition: self-referential judgements** 15 **about internal representations**

16 There are several kinds of metacognitive judgements that seem auto-noetic. The criterion is that
 17 the judgement be specifically self-referential. The three main categories of research that conform
 18 to this definition of auto-noetic metacognition are source judgements, remember/know judge-
 19 ments, and agency judgements.

20 **Source judgements**

21 While there is a large literature on source judgements (see Johnson et al. 1993; Mitchell and
 22 Johnson 2009), most of that literature is not specifically self-referential. For example, much effort
 23 has been invested in determining when and under what circumstances people are able to distin-
 24 guish one person from another as the source of an utterance, but neither person is the self, or
 25 whether the original input was auditory or visual, say, or whether the background colour was red
 26 or blue. Young children and older adults (Craik et al. 1990; Henkel et al. 1998) have especially
 27 difficulties with source judgements. But none of them qualify as necessarily being auto-noetic.

28 However, certain source judgement are necessarily auto-noetic, if the distinction the individual
 29 must make involves the self as compared to another, or the self in one form (imagining speaking,
 30 say) as compared to in another form (actually speaking). People with schizophrenia have particu-
 31 lar difficulty with this kind of judgements (Wang et al. 2010). Furthermore, deficits in self-other
 32 source (but note, these are often not distinguished from non-self-referential source judgements
 33 in the literature) appear to be related to positive symptoms of schizophrenia such as hallucina-
 34 tions and delusions.

35 Many of the results in the source monitoring literature focus on the details of memories of past
 36 events, and some of these studies—those that are particularly relevant for self-consciousness—
 37 investigate the extent and manner of self-involvement in those memories. However, it could be
 38 argued that a simpler kind of metacognition—that involving adjectival checklists, or self referen-
 39 tial statements—is also a kind of metacognitive judgement that is also auto-noetic. When a person
 40 is asked to decide whether they are warm, attractive, miserly, or intelligent, presumably these
 41 judgements are specifically referred to a representation of the self, and would need to be called
 42 auto-noetic by our definition of the term. Interestingly, when one is making such judgements,
 43 there is a particular area of the medial prefrontal cortex that appears to be selectively activated
 44 (Ochsner et al. 2005; Jenkins and Mitchell 2011). That area is also often found to be activated in
 45 episodic memory task that Tulving would call auto-noetic in nature—a fascinating relation that

1 deserves further research. It is conceivable that this area is, in some sense that is undoubtedly too
2 simple but nevertheless intriguing, the seat of the self.

3 Remember-know judgements

4 Judgements concerning whether the individual remembers that an event happened in his or her
5 personal past, or just knows that something is familiar (Tulving 1985; Gardiner 1988) are meta-
6 cognitive judgements proper, that, taken at face value, are specifically self-referential and hence
7 auto-noetic (Gardiner et al. 1998; Hirshman 1998; Yonelinas 2002). Indeed, they have often been
8 taken as the most quintessential of auto-noetic judgements.

9 There is, however, dispute in the literature about exactly how the individual makes remember-
10 know judgements. If they simply evaluate the amount of information that can be retrieved, and
11 say that they ‘remember’ when they have retrieved a great deal of information, and that they
12 ‘know’ when they have retrieved a lesser amount of information, then these judgements are essen-
13 tially retrospective confidence judgements. As with confidence judgements detailed in the previ-
14 ous section, they would be noetic rather than auto-noetic judgements. Some researchers have
15 argued for such an explanation, demonstrating that many of the characteristics of remember/
16 know judgements can be handled within a signal detection framework (Donaldson 1996; Dunn
17 2004; Wixted and Stretch 2004). However, Yonelinas (2002) and others (e.g. Wolk et al. 2006)
18 have argued that two processes are involved: familiarity monitoring and recollective retrieval.
19 These dual process theorists get closer to the original idea that there is something special and
20 different about ‘remember’ judgements. But even in this dual process view, the more complex
21 form of memory access (i.e. recollective retrieval) is not necessarily self-referential. Insofar as the
22 judgement that one remembers *is* self-referential, then, the remember-know paradigm would
23 appear to be an auto-noetic form of metacognition, but neither model of the task emphasizes this
24 characteristic.

25 Agency judgements

26 People are able to make fairly reliable judgements of their own agency—they can assess the extent
27 to which they were or were not the causal agent in producing an action outcome (Metcalf et al.
28 2010; Miele et al. 2011), a clearly self-referential metacognition. However, they cannot do so infal-
29 libly. Wegner and Wheatley (1999; Wegner 2003; Wegner et al. 2004) have provided several fas-
30 cinating experimental examples of errors in these judgements. In one study, participants, wearing
31 headphones, with their hands at their sides, looked at a mirror image of themselves covered by a
32 smock with the hands of a confederate protruding where their own hands would normally be
33 seen. The participants, of course, knew that the hands that they were seeing in the mirror were not
34 their own hands. But if a word for an object was primed (via the headphones) at just the right
35 moment before the hands that looked like their own hands moved, people had a spooky feeling
36 that they had reached for the object. Their judgement of agency, hence, was malleable and subject
37 to illusion.

38 But while agency judgements can be distorted (as can lower-level metacognitions), they are
39 normally accurate. For example, Metcalf and Greene (2007) showed that college students usually
40 correctly know when they have moved a mouse to catch a target, and when noise-like interfer-
41 ence, which distorted their own planned movements, intervened. Knoblich et al. (2004) showed
42 that while typical adults can detect a distortion in their motor movements, patients with schizo-
43 phrenia have great difficulty in doing so.

44 What about non-human animals? The data, so far, are scant but promising on this issue
45 (Couchman 2012). But, insofar as one component of metacognitive judgements of agency
46 involves action monitoring non-human primates may—given their dexterity and physical

1 competence—be excellent at it. Originally the comparator action monitoring models (Wolpert
 2 et al. 1995), that form the core of most theoretical views of how people make judgements of
 3 agency, were devised as a way of understanding how it is possible for people to make nuanced and
 4 complex fast actions. The central idea is that the person has a plan of where and how to move.
 5 This plan runs off mentally in real time synchronously with their actual movement, and the feed-
 6 back from the movement is collated with the expectations from the plan. If the two correspond
 7 perfectly, the action proceeds smoothly. If there is a mismatch, then an alteration is needed
 8 to correct the movement. This match/mismatch mechanism, devised for motor control, was co-
 9 opted by the metacognitive system, to allow people to make judgements of agency: if there is no
 10 discrepancy, then the person was in control. If a discrepancy occurred, though, then some outside
 11 source was distorting the correspondence between intent and action, and the person was not in
 12 full control. Presumably to accomplish acrobatic feats so common in the wild, our primate ances-
 13 tors would need to have a finely tuned action monitoring system. Whether, like humans, they
 14 co-opted it to allow them to have metacognition of agency and perhaps even a concept of the self,
 15 we do not know.

16 Conclusion

17 Is it conceivable that a non-human animal or a computer could exhibit auto-noetic metacogni-
 18 tion? So far, to our knowledge, no computer has ever done any truly self-referential task. But
 19 typically, computers are not programmed to remember their past or project into their future. Nor
 20 are they programmed to take particular account of things they themselves did. But there seems to
 21 be no ‘in principle’ reason why this could not be programmed into them. It is imaginable that a
 22 computer-robot could be programmed to encode the visual scenes that occurred from their per-
 23 spective while they moved around in the environment and use those ‘personal’ records in later
 24 encounters, tagging particular knowledge as specific to them. Watson, too, could be programmed
 25 to tag his own answers and those of the other participants such that he could later ‘remember’ the
 26 source of the answers. But if that were done would it mean that Watson would have auto-noetic
 27 metacognition?

28 One argument against this is that, although such noting and tagging would allow him to give
 29 answers that mimicked those of a person who had a self, the records of the computer would com-
 30 prise a pseudo self. Humphrey (2006) has made a fascinating case that the internalized concept
 31 of a self developed in animals because it bestowed evolutionary advantages on those who had it.
 32 The advantage accrues because the self as an embodied and encapsulated concept results in an
 33 individual who both has a mind, and has a concept of its own physical body and, thereby, strives
 34 to preserve and foster it. If one compared an animal with a self to one without, the former would
 35 be more motivated to protect its physical body. And, of course, protecting one’s body is evolu-
 36 tionarily advantageous. If the ‘real’ self is necessarily linked to some such creature-based evolu-
 37 tionary account, then even if Watson could access the digital records taken from his perspective,
 38 or could answer Watson versus other source questions correctly, he would not thereby manifest
 39 a ‘real’ self. The deep and meaningful characteristics of what self-reference means to humans and
 40 to their survival would not follow from answering such questions correctly. In short, the answers
 41 to the questions directed at determining whether the answerer has auto-noetic consciousness
 42 could be faked.

43 How does metacognition relate to self-awareness, then? First of all, we have argued that auto-
 44 noetic and noetic metacognition do not imply self-awareness at all. That being the case, even humans
 45 may not always be self-aware when making metacognitive judgements (e.g. Son and Kornell
 46 2005). But auto-noetic metacognition (as long as it is not faked) suggests that the individual has

1 self-awareness, and an internalized, articulate concept of the self. Now, of course, humans may
 2 also be self-aware at other times—the argument is only that anoetic and noetic metacognition
 3 provide no positive evidence.

4 At present, we know almost nothing about self-awareness in non-human primates and other
 5 animals. The question has not yet been posed. But, if someone were able to convincingly devise a
 6 method of asking a monkey whether he was the agent or someone else was, he might be able
 7 to answer it correctly. And, it would not be too far fetched to suppose that—in the complex social
 8 world in which primates in the wild live, in which keeping track, over time, of exactly who
 9 did what to whom might enhance one's chances of survival—a self might be a valuable thing
 10 to have.

11 References

- 12 Carruthers, P. (2011). *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford: Oxford
 13 University Press.
- 14 Couchman, J. J. (2012). Self-agency in rhesus monkeys. *Biological Letters*, 8(1), 39–41.
- 15 Craik, F. I., Morris, L. W., Morris, R. G., and Loewen, E. R. (1990). Relations between source amnesia and
 16 frontal lobe functioning in older adults. *Psychology & Aging*, 5, 148–51.
- 17 Descartes, R. (1637). *Discourse on Method*. Cambridge University Press.
- 18 Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition*,
 19 24, 523–33.
- 20 Dunn, J. C. (2004). Remember-know: A matter of confidence. *Psychological Review*, 111, 524–42.
- 21 Finn, B. and Metcalfe (2010). Scaffolding feedback to maximize long term error correction. *Memory &*
 22 *Cognition*, 38, 951–61.
- 23 Foote, A. L. and Crystal, J. D. (2007). Metacognition in the rat. *Current Biology*, 17, 551–5.
- 24 Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition*, 16, 309–13.
- 25 Gardiner, J. M., Richardson-Klavehn, A., and Ramponi, C. (1998). Limitations of the signal-detection
 26 model of the remember-know paradigm: A reply to Hirshman. *Consciousness & Cognition*, 7, 285–8.
- 27 Hampton, R. R. (2001). Rhesus monkeys know when they remember. *Proceedings of the National Academy*
 28 *of Sciences of the United States of America*, 98(9), 5359–62.
- 29 Hawkins, S. A. and Hastie, R. (1990). Hindsight: Biased judgements of past events after the outcomes are
 30 known. *Psychological Bulletin*, 107, 311–27.
- 31 Henkel, L. A., Johnson, M. K., and De Leonardi, D. M. (1998). Aging and source monitoring: Cognitive
 32 processes and neuropsychological correlates. *Journal of Experimental Psychology: General*, 127, 251–68.
- 33 Hirshman, E. (1998). On the utility of the signal detection model of the remember-know paradigm.
 34 *Consciousness & Cognition*, 7, 103–7.
- 35 Hoffrage, U. and Pohl, R. (2003). Research on hindsight bias: A rich past, a productive present, and a
 36 challenging future. *Memory*, 11, 329–35.
- 37 Humphrey, N. (2006). *Seeing red: A study in consciousness*. Boston, MA: Harvard University Press.
- 38 Husserl E. (1929). *Cartesian Meditations and the Paris Lectures*. The Hague: Martinus Nijhoff, (1973).
- 39 James, W. (1890). *The Principles of Psychology, Volume 1*. New York: Henry Holt.
- 40 Jenkins, A. C. and Mitchell, J. P. (2011). Medial prefrontal cortex subserves diverse forms of self-reflection.
 41 *Social Neuroscience*, 6(3), 211–18.
- 42 Johnson, M. K., Hashtroudi, S., and Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*,
 43 114, 3–28.
- 44 Kahneman, D., Knetsch, J., and Thaler, R. (1990). Experimental tests of the endowment effect and the
 45 Coase theorem. *Journal of Political Economy*, 98, 1325–48.
- 46 Knoblich, G., Stottmeister, F., and Kircher, T. T. J. (2004). Self-monitoring in patients with schizophrenia.
 47 *Psychological Medicine*, 34, 1561–9.

AQ: Please give
 the initial.

- 1 Kohler, W. (1925). *The Mentality of Apes*. New York: Routledge and Kegan Paul.
- 2 Koriat, A., Bjork, R. A., Sheffer, L., and Bar, S. K. (2004). Predicting one's own forgetting: The role of
3 experience-based and theory-based processes. *Journal of Experimental Psychology: General*, 133, 643–56.
- 4 Kornell, N. (2009). Metacognition in humans and animals. *Current Directions in Psychological Science*,
5 18, 11–15.
- 6 Kornell, N., Son, L. K., and Terrace, H. S. (2007). Transfer of metacognitive skills and hint seeking in
7 monkeys. *Psychological Science*, 18, 64–71.
- 8 Menzel, C. R. (1999). Unprompted recall and reporting of hidden objects by a chimpanzee
9 (*Pan troglodytes*) after extended delays. *Journal of Comparative Psychology*, 113, 426–34.
- 10 Menzel, E. (2005). Progress in the study of chimpanzee recall and episodic memory. In H. S. Terrace and
11 J. Metcalfe (Eds.) *The missing link in cognition*, pp. 188–224. Oxford: Oxford University Press.
- 12 Metcalfe, J. and Greene, M. J. (2007). Metacognition of agency. *Journal of Experimental Psychology: General*,
13 136, 184–99.
- 14 Metcalfe, J. and Kober, H. (2005). Self-reflective consciousness and the projectable self. In H. S. Terrace and
15 J. Metcalfe (Eds.) *The Missing Link in Cognition: Origins of Self-Reflective Consciousness*, pp. 57–83.
16 Oxford: Oxford University Press.
- 17 Metcalfe, J., Eich, T. S., and Castel, A. (2010). Metacognition of agency across the lifespan. *Cognition*, 116,
18 267–82.
- 19 Miele, D. M., Wager, T. D., Mitchell, J. P., and Metcalfe, J. (2011) Dissociating neural correlates of action
20 monitoring and metacognition of agency. *Journal of Cognitive Neuroscience*, 23(11), 3620–36.
- 21 Mitchell, K. J. and Johnson, M. K. (2009). Source monitoring 15 years later: What have we learned from
22 fMRI about the neural mechanisms of source memory? *Psychological Bulletin*, 135, 638–77.
- 23 Ochsner, K. N., Beer, J. S., Robertson, E. R., et al. (2005). The neural correlates of direct and reflected
24 self-knowledge. *Neuroimage*, 28, 797–814.
- 25 Nelson, T. O. and Narens, L. (1990). Metamemory: A theoretical framework and new findings.
26 In G. Bower (Ed.) *The psychology of learning and motivation* (Vol. 26), pp. 125–41. New York:
27 Academic Press.
- 28 Reder, L. M. (1987). Strategy selection in question answering. *Cognitive Psychology*, 19, 90–138.
- 29 Reder, L. M. and Schunn, C. D. (1996). Metacognition does not imply awareness: Strategy choice is
30 governed by implicit learning and memory. In L. M. Reder (Ed.) *Implicit Memory and Metacognition*,
31 pp. 45–77. Mahwah, NJ: L. Erlbaum.
- 32 Russell, B. (1997). *The Problems of Philosophy*. New York: Oxford University Press.
- 33 Rosenbaum, R. S., Köhler, S., Schacter, D. L., et al. (2005). The case of K.C.: Contributions of a
34 memory-impaired person to memory theory. *Neuropsychologia*, 43, 989–1021.
- 35 Rosenthal, D. (2000) Consciousness, content, and metacognitive judgements. *Consciousness Cognition*, 9,
36 203–14.
- 37 Schwartz, B. L. and Metcalfe, J. (2011). Tip-of-the-tongue (TOT) states: Retrieval, behavior, and
38 experience. *Memory & Cognition*, 39(5), 737–49.
- 39 Shields, W. E., Smith, J. D., Guttmanova, K., and Washburn, D. A. (2005). Confidence judgments by
40 humans and rhesus monkeys. *Journal of General Psychology*, 132, 165–86.
- 41 Searle, J. R. (1992). *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.
- 42 Smith J. D. (2009). The study of animal metacognition. *Trends in Cognitive Science*, 13, 389–96.
- 43 Smith J. D., Schull, J., Strote, J., McGee, K., Egnor, R., and Erb, L. (1995). The uncertain response in the
44 bottlenosed dolphin (*Tursiops truncatus*). *Journal of Experimental Psychology: General*, 124, 391–408.
- 45 Smith, J. D., Shields, W. E., Allendoerfer, K. R., and Washburn, W. A. (1998). Memory monitoring by
46 animals and humans. *Journal of Experimental Psychology: General*, 127, 227–50.
- 47 Smith, J. D., Beran, M. J., Couchman, J. J., Coutinho, M. V. C., and Boomer, J. B. (2009). Animal
48 metacognition: Problems and prospects. *Comparative Cognition and Behavior Reviews*, 4, 40–53.

- 1 Son, L. K. and Kornell, N. (2005). Meta-confidence judgments in rhesus macaques: Explicit versus implicit
 2 mechanisms. In H. S. Terrace and J. Metcalfe (Eds.) *The Missing Link in Cognition: Origins of*
 3 *Self-Knowing Consciousness*, pp. 296–320. New York: Oxford University Press.
- 4 Son, L. K. and Metcalfe, J. (2005). Judgments of learning: Evidence for a two-stage model. *Memory &*
 5 *Cognition*, 33, 1116–29.
- 6 Staddon, J. E. R., Jozefowicz, J., and Cerutti, D. T. (2007). Metacognition: A problem not a process. *PsyCrit.*
- 7 Terrace, H. S., and Son, L. K. (2009). Comparative metacognition. *Current Opinion in Neurobiology*, 19,
 8 67–74.
- 9 Thaler, R. (1980). Towards a positive theory of consumer choice. *Journal of Economic Behavior and*
 10 *Organization*, 1, 39–60.
- 11 Tulving, E. (1984). Elements of episodic memory. *Behavioral and Brain Sciences*, 7, 223–68.
- 12 Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, 26, 1–12.
- 13 Tulving, E. (2005). Episodic memory and auto-noesis: Uniquely human? In H. S. Terrace and J. Metcalfe
 14 (Eds.) *The Missing Link in Cognition: Origins of Self-Knowing Consciousness*, pp. 4–56. New York:
 15 Oxford University Press.
- 16 Wang, L., Metzack, P. D., and Woodward, T. S. (2010). Aberrant connectivity during self-other source
 17 monitoring in schizophrenia. *Schizophrenia Research*, 125, 136–42.
- 18 Washburn, D. A., Gullledge, J. P., Beran, M. J., and Smith, J. D. (2010). With his memory erased, a monkey
 19 knows he is uncertain. *Biology Letters*, 6, 160–2.
- 20 Wegner, D. M. (2003). *The illusion of conscious will*. Cambridge, MA: MIT Press.
- 21 Wegner, D. M. and Wheatley, T. (1999). Apparent mental causation: Sources of the experience of will.
 22 *American Psychologist*, 54, 480–92.
- 23 Wegner, D. M., Sparrow, B., and Winerman, L. (2004). Vicarious agency: Experiencing control over the
 24 movements of others. *Journal of Personality and Social Psychology*, 86, 838–48.
- 25 Wheeler, M. A., Stuss, D. T., and Tulving, E. (1997). Toward a theory of episodic memory: The frontal
 26 lobes and auto-noetic consciousness. *Psychological Bulletin*, 121, 331–54.
- 27 Wixted, J. T. and Stretch, V. (2004). In defense of the signal detection interpretation of remember/know
 28 judgments. *Psychonomics Bulletin & Review*, 11, 616–41.
- 29 Wolk, D. A., Schacter, D. L., Lygizos, M., et al. (2006). ERP correlates of recognition memory: Effects of
 30 retention interval and false alarms. *Brain Research*, 1096, 148–62.
- 31 Wolpert, D. M., Ghahramani, Z., and Jordan, M. I. (1995). An internal model for sensorimotor integration.
 32 *Science*, 269, 1880–2.
- 33 Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research.
 34 *Journal of Memory & Language*, 46, 441–517.
- 35 Zalla T., Stopin, A., Ahade, S., Sav, A. M., and Leboyer, M. (2008). Faux-pas detection and intentional
 36 action in Asperger syndrome. A replication on a French sample. *Journal of Autism and Developmental*
 37 *Disorders*, 39, 373–82.