Anoetic, Noetic and Autonoetic Metacognition

Janet Metcalfe & Lisa Son
Columbia University Barnard College

To appear in: The Foundations of Metacognition, M. Beran, J. R. Brandl, J. Perner, & J. Proust (Eds.) Oxford University Press: Oxford, UK.

INTRODUCTION

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

Now imagine L., who is playing a memory gambling game. He is presented with the target--a complex picture-- in a flash on the screen. The picture disappears, and 9 alternative pictures appear on the screen simultaneously. L. looks through them considering each in turn and upon seeing what he thinks is the target picture in the array, he touches it, and they all disappear. Then, though, he has to give his confidence in his answer. He can either 'pass'--choose not to wager -- or he can 'double down' -- wager big. Two betting icons appear on the screen. Nothing further will happen until he makes this retrospective decision about whether he thinks he was right or wrong. In this case, L. chooses the 'double down' icon, and he wins 3 tokens, which fall into his hopper, to be redeemed later when he has accumulated enough tokens for a prize. Had he pushed the 'pass' icon, he would have gotten only one token. But had he touched the wrong picture in the 9-option task, and then 'doubled down', he would have seen 3 tokens fly out of his hopper and disappear. L. is known, by other gamblers, as having a *serious* emotional reaction when this happens. But, fortunately, it doesn't happen often. And he does get paid off with prizes from time to time. Like other gamblers, L. is happy to play this game of making metacognitive bets on his own memory hour after hour, day after day.

Finally imagine S. trying to retrieve the name of the famous Canadian author who wrote "The Last Spike." A nagging feeling of having the answer right on the tip of his tongue plagues S. But S. cannot retrieve the answer no matter how hard he tries, and he is trying hard. His friends tell him to give up. None of them are Canadians, and they neither know the answer, nor care, to be sure. But S. refuses to listen. His mind is screaming with this impossible-to-resist emotional premonition that the answer is eminent (see Schwartz & Metcalfe, in press). And he is right, statistically, at least. When people have this feeling, they nearly always get the answer eventually. But it is hours, not moments away. Having been driven almost to distraction by this tantalizing gap in his knowledge, and knowing that the answer, oddly, is 'almost' a French name, and that the first letter of that first name is P., finally, in a flash of insight the answer--Pierre Berton--appears, seemingly unbidden out of the blue (previous intense efforts to find it notwithstanding).

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

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

The modern theorist most associated with this view is Rosenthal (2000). In advancing his 'higher order thought' (HOT) hypothesis, he argues that consciousness is essentially metacognition, which, classically (see, Nelson & Narens, 1990) entails the reflection at the metalevel upon a lower, basic, level. Rosenthal notes: "The leading idea behind the HOT hypothesis is that a mental state is conscious only if one is, in some suitable way, conscious *of* that state...A conscious state is a state one is conscious *of oneself as being in.*" (p. 231-2). Rosenthal 's HOTs involve something more than just a metalevel reflection on a basic level representation: *self*-consciousness is implied. He does not necessarily endorse an elaborate folk-theoretic notion of what self consciousness entails including being explicitly conscious of oneself as the subject, or of having all of one’s conscious thoughts and experiences come together mentally. Self-consciousness could be much more pared down: "HOTs can, instead, represent the self in some minimal way, for example, in terms simply of a distinction between oneself and everything else." But, even though minimal, some form of self-consciousness is implied. Furthermore, Rosenthal says that such consciousness can only be found in creatures; presumably, computers need not apply. But, perhaps nonhuman animals could.

Animal metacognition researchers almost invariably allude to the self-awareness aspect of metacognition in motivating their investigations of whether animals might be

[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'."

able to do metacognitive tasks. For example, Smith, Beran, Couchman, Coutinho and Boomer (2009) justify their research on animals by saying: “Metacognition is linked to self-awareness … because doubt is so personal and self-oriented. Metacognition is linked to declarative consciousness, because we can introspect and declare states of knowing. Thus, metacognition is a sophisticated capacity in humans that might be uniquely human.” (p. 40). Smith (2009) says " one of comparative psychology’s current goals is to establish whether nonhuman animals (hereafter, animals) share humans’ metacognitive capacity. If they do, it could bear on their consciousness and self-awareness too.” (p. 389). Foote and Crystal (2007), who investigated metacognition in rats, say ”People are sometimes aware of their own cognitive processes. Therefore, studies in metacognition test the hypothesis that animals behave functionally the same as an organism that is aware of its own cognitive state.” (p. 1).

And, while, if W., L., and S. were all people, we would have no qualms about admitting that the stream and quality of the metacognitive thought processes would allow us to attribute selfhood to each--they 'feel' like people-- when we realize that two of these three were not even humans, we might balk at this conclusion. And, indeed, W. in our example above, is Watson, the IBM computer who recently made front page news by beating out previous Jeopardy champions to become the new world champion. The feat is impressive, but does it imply that W. is conscious and has a self? And L. is Lashley, a rhesus monkey. S. is human, with the initial chosen for 'Self.' In that light, S.'s musings about his tip of the tongue state leave little doubt, in most people's minds, that he has mind, consciousness and self-awareness. But while, intuitively, we reject the idea that Watson might have a self, and remain agnostic about Lashley (while perhaps swayed toward the possibility by the metacognitive data), the question remains: If the evidence for self awareness is metacognition, why do we accept that evidence for Self but not for Watson? Perhaps we are merely exhibiting an anthropocentric bias, and the impressive performance on the above metacognitive tasks, by all three actors, should mean that we should, rationally, be compelled to abandon our prejudices against machine or monkey and attribute consciousness and a self to all three. One possibility, though, which we explore in this essay is that perhaps it is only *certain* metacognitive tasks, with *particular* characteristics that imply high level consciousness and selfhood. We will here endeavor to analyze tasks that have been labeled as “metacognitive” into three different levels, borrowed from Tulving’s (l985) analysis of different levels of consciousness: *Anoetic*, *Noetic*, and *Autonoetic*.

THREE LEVELS OF CONSCIOUSNESS AND METACOGNITION

Before analyzing various metacognitive tasks we will first review Tulving's (1984; Rosenbaum, Kohler, Schacter, Moscovitch, Westmacott, Black, Gao & Tulving, 2005; Wheeler, Stuss & Tulving, 1997) distinction between three different levels of consciousness.

Anoetic consciousness. At the lowest level, Tulving defines anoetic consciousness as a state that is temporally and spatially bound to the current time. Although it is a kind of consciousness, it is not one that allows escape in any way from the here and now, and so an animal functioning at this level of consciousness is stimulus bound. A judgment that refers to something in the world even though that something is interpreted through the viewer's perceptual biases and learning would, then, be anoetic. Thus, if a person were learning to discriminate between Pinot Gris and Pinot Grigio, for example, and made judgments, based on tastes of various wine samples, these judgments--being about something in the world, even though the internal percept experienced is, undoubtedly, biased by the learning mind-- would be anoetic. Note that while mental processes and past discrimination learning may interact with just what the subject perceives (we make no claim that perception is naive) the percept, itself, is bound to the moment. It is not a representation or a memory of Pinot Grigio, but rather the percept of the wine itself that is being judged (and so is neither a judgment about an internal representation nor, indeed, is it a judgment about the judgment). By some definitions (see Metcalfe & Kober, 2005; Carruthers, 2011) a judgment at this level would not be considered metacognitive at all. It would simply be a judgment about the world as perceived. But other researchers (e.g., Reder & Schunn, 1996; Smith, 2009) have labeled such judgments metacognitive. The framework specified by Nelson and Narens (1990), proposed that there are at least two levels of cognition interacting to form a metacognitive system, a basic level and a metalevel. The basic level, in this anoetic case, would not be a representation at all, however, but rather a percept, and so it is not clear that the word meta-'cognition', should be applied to judgments, such as these, concerned with percepts. They might better be called metaperceptual. But perhaps to overcome the definitional disputes about whether judgments about objects or events in the world as perceived by the subject are metacognitive, and, hopefully to forward our understanding of whether or not self-awareness is involved, we could agree to call such judgments anoetic metacognition. Anoetic consciousness, of course, makes no reference to the self. Similarly, anoetic metacognition could not be considered to involve self awareness.

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

Autonoetic consciousness. This is the highest form of consciousness and is self-reflective or self-knowing. For the first time, the self, then, is intimately involved. This level of consciousness is often, in Tulving's framework, related to human adult episodic memory, which may involve mental time travel of the self. Autonoetic consciousness is thought to be necessary for the remembering of personally experienced events, as long as the memory of those events is self-referential. An individual could not remember something that they experienced in a noetic manner, if they did not know that they had explicitly experienced it, as has been shown to be the case with certain amnesic patients, such as K.C., who are thought to lack autonoetic memory (Rosenbaum et al., 2005). But when a normal person remembers an event in which they participated, he or she is normally thought to be aware of the event as a veridical (or sometimes non-veridical) part of his own past existence, and the involvement of the self is a necessary component in this kind of consciousness. Autonoetic consciousness is not mere depersonalized knowledge. Rather, as James (1890) says: "this central part of the Self is *felt*... and no *mere* summation of memories or *mere* sound of a word in our ears. It is something with which we also have direct sensible acquaintance, and which is as fully present at any moment of consciousness in which it *is* present, as in a whole lifetime of such moments" (p.299). A normal healthy person who possesses autonoetic consciousness is capable of becoming aware of her own projected future as well as her own past; she is capable of mental time travel, roaming at will over what has happened as readily as over what might happen, independently of physical laws that govern the universe. According to Tulving (2005) only humans past infancy possess autonoetic consciousness.

Do any kind of metacognitive judgments necessarily involve autonoetic consciousness? It would seem that if the judgment makes specific reference to the self it would qualify. A metacognition at the autonoetic level might also be a judgment about one's own personal memories of one's own personal past. From the standpoint of relating metacognition to self-awareness, then, these particular kinds of metacognitions, if there are any such, are of particular importance, since it is only these that involve self-consciousness.

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

Anoetic Metacognition: Stimulus-Driven Judgments

The lowest level of metacognition is anoetic. Any judgment where the individual is evaluating an external stimulus is here categorized as anoetic. Consider the simple example when judging the value of an item, say, a mug. One could say that a mug is worth $10. One's judgment of the mug changes, though, depending on who owns the mug (Kahneman, Knetschm & Thaler, 1990). While the object is "endowed" with higher value when possessed by the individual (Thaler, 1980), as given by his or her subjective judgment, the judgment is, nevertheless, of an external stimulus rather than a representation; it is anoetic and no self awareness is involved. The judgment of the Pinot Grigio mentioned above, whether by a trained or untrained palate, also falls into this category, as do all such perceptual/categorical judgments.

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

Similarly, the classic "escape" studies in dolphins are anoetic. In one such study (Smith, Schull, Strote, McGee, Egnor, & Erb, 1995) dolphins were required to discriminate the auditory frequencies of two tones by responding with one of two responses. If a 2100-Hz tone was sounded, the dolphin was rewarded when it responded to a “2100 Hz” icon; for all lower frequencies, the dolphin was rewarded when it responded to a “<2100 Hz” icon. An error terminated the trial without reinforcement and resulted in a punishment in the form of a time out. A response to a third “escape” icon also terminated the trial, but with neither reward nor punishment. It simply acted as an expression of “I’d rather opt out of this question” and moved onto a new trial. Dolphins could do this task, and sometimes chose to escape rather than take the test. Even allowing that their doing so was a judgment, it was an anoetic judgment, and hence does not imply self-awareness. Other “escape” type studies (e.g. Shields, Smith, Guttmannova & Washburn, 2005; Smith, Shields, & Allendoerder, 1998; Washburn, Gulledge, Beran, & Smith, 2010), where the probe or percept, and not a internal representation, gives rise to the judgment, would also be included as examples of anoetic metacognition (see Terrace & Son, 2009, for a review of yet other cases of anoetic metacognition using the escape paradigm).

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

Noetic Metacognition: Judgment about an internal representation

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

Classic cue-only delayed judgments of learning are a typical case of noetic metacognitive judgments. A learning event, consisting of a cue and a to-be-learned target, is presented, and then at some later time, the person is given the cue and asked to make a judgment about whether he or she will later be able to give evidence that they know the target. If they think they will know it they give it high judgment; if not then they give a low judgment of learning. Note, if people mentally projected their selves into the future to see whether they would get the answer this judgment would be considered autonoetic. However, the data on what people actually do to make this assessment suggest that they do not so mentally time travel. The most compelling evidence for a lack of mental future projection is that people's judgments of learning do not distinguish between whether the test will be 5 minutes or 1 year hence (Koriat, Bjork, Sheffer, & Bar, 2004)--a distinction that would be large were people really mentally projecting into the future. What they appear to do instead (Son & Metcalfe, 2005; Metcalfe & Finn, 2010) is first try to recognize the cue. If they cannot do so they say that they don't know and give a fast low rating. If they do recognize it, they then attempt to retrieve the target, with judgments of learning getting lower and lower the longer it takes them to do so. Thus, the judgment is about the current retrievability of the cue and target, and hence noetic in nature.

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

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

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

The ambiguous case of Panzee the chimp: Noetic or autonoetic metacognition?

Panzee, a female chimpanzee, had been taught to use over 100 lexigrams, at the time of the 'experiment' in which one keeper hid 26 food objects and 7 nonfood objects in a large forest field, an area that Panzee knew from her past, but had not visited in 6 years (Menzel, 2005). Panzee was able to recruit the assistance of other caretakers (who knew nothing about the objects being hidden) and “tell” them where the objects were hidden. Because these new caretakers were not aware of the 'experiment' at all, let alone where the objects were hidden, when objects were found, it was thought to be the result of Panzee's "own initiative" (Menzel, 2005, p. 199). The uninformed caretaker found all 34 objects as a result of Panzee's behavior! And, furthermore, Panzee had indicated on her lexigram board 84% of the time, which particular item had been hidden in each location, and correctly identified these items at delays, for some items, of over 90 hours from the original hiding event. Evidence in support of metacognition was seen in Panzee's behavior: The caretaker noted and responded to Panzee's relative degree of excitement--a seemingly spontaneous metacognition, since it directly reflected the distance to the target. Panzee kept pointing, showed intensified vocalization, shook her arm, and bobbed her head or body as the caretaker got closer to the site (see, Menzel, 2005, p. 202). In addition, Menzel reported that Panzee seemed to do whatever it took to catch the caretaker's attention and, only once joint attention was established, touched the lexigram corresponding to the type of object hidden, pointed outdoors, sometimes went outdoors (if the caretaker followed), and continued to point manually toward the object and vocalize until the caretaker found the object. As noted by Kohler (1925), the "time in which chimpanzees live" and whether they are able to freely mentally time travel, as autonoetic consciousness requires, remains an open question, but it seems, from these data that Panzee could, at the very least, freely recall which one of at least 20 types of objects she had been shown at a distance and at a long delay, and that she was highly certain, and highly keyed up, of her own knowledge-- a feat that begins to look a lot like human autonoesis.

Autonoetic Metacognition: Self-Referential Judgments about Internal Representations

There are several kinds of metacognitive judgments that seem autonoetic. The criterion is that the judgment be specifically self referential. The three main categories of research that conform to this definition of autonoetic metacognition are source judgments, remember/know judgments, and agency judgments.

Source judgments. While there is a large literature on source judgments (see, Johnson, Hashtroudi & Lindsay, 1993; Mitchell & Johnson, 2009), most of that literature is not specifically self referential. For example, much effort has been invested in determining when and under what circumstances people are able to distinguish one person from another as the source of an utterance, but neither person is the self, or whether the original input was auditory or visual, say, or whether the background color was red or blue. Young children and older adults (Craik, Morris, Morris & Loewin, 1990; Henkel, Johnson & DeLeonardis, 1998) have especially difficulties with source judgments. But none of them qualify as necessarily being autonoetic.

However, certain source judgment are necessarily autonoetic, if the distinction the individual must make involves the self as compared to another, or the self in one form (imagining speaking, say) as compared to in another form (actually speaking). People with schizophrenia have particular difficulty with this kind of judgments (Wang, Metzak, & Woodward, 2010). Furthermore, deficits in self-other source (but note, these are often not distinguished from non-self-referential source judgments in the literature) appear to be related to positive symptoms of schizophrenia such as hallucinations and delusions.

Many of the results in the source monitoring literature focus on the details of memories of past events, and some of these studies--those that are particularly relevant for self-consciousness-- investigate the extent and manner of self-involvement in those memories. However, it could be argued that a simpler kind of metacognition-- that involving adjectival check lists, or self referential statements --is also a kind of metacognitive judgment that is also autonoetic. When a person is asked to decide whether they are warm, attractive, miserly, or intelligent, presumably these judgments are specifically referred to a representation of the self, and would need to be called autonoetic by our definition of the term. Interestingly, when one is making such judgments there is a particular area of the medial prefrontal cortex that appears to be selectively activated (Jenkins & Mitchell, 2011; Ochsner, Beer, Robertson, Cooper, Gabrieli, Kihsltrom, & D'Esposito, 2005). That area is also often found to be activated in episodic memory task that Tulving would call autonoetic in nature--a fascinating relation that deserves further research. It is conceivable that this area is, in some sense that is undoubtedly too simple but nevertheless intriguing, the seat of the self.

Remember-Know judgments. Judgments concerning whether the individual remembers that an event happened in his or her personal past, or just knows that something is familiar (Tulving, 1985; Gardiner, 1988) are metacognitive judgments proper, that, taken at face value, are specifically self-referential and hence autonoetic

(Gardiner, Richardson-Klavehn, & Ramponi, 1998; Hirshman, 1998; Yonelinas, 2002). Indeed, they have often been taken as the most quintessential of autonoetic judgments.

There is, however, dispute in the literature about exactly how the individual makes remember-know judgments. If they simply evaluate the amount of information that can be retrieved, and say that they 'remember' when they have retrieved a great deal of information, and that they 'know' when they have retrieved a lesser amount of information, then these judgments are essentially retrospective confidence judgments. As with confidence judgments detailed in the previous section, they would be noetic rather than autonoetic judgments. Some researchers have argued for such an explanation, demonstrating that many of the characteristics of remember/know judgments can be handled within a signal detection framework (Donaldson, 1996; Dunn, 2004; Wixted & Stretch, 2004). However, Yonelinas (2002) and others (e.g., Wolk, Schacter, Lygizos, Sen, Chong, Holcomb, Daffner, & Budson, 2006) have argued that two processes are involved: familiarity monitoring and recollective retrieval. These dual process theorists get closer to the original idea that there is something special and different about 'remember' judgments. But even in this dual process view, the more complex form of memory access (i.e., recollective retrieval) is not necessarily self-referential. Insofar as the judgment that one remembers *is* self-referential, then, the remember-know paradigm would appear to be an autonoetic form of metacognition, but neither model of the task emphasizes this characteristic.

Agency judgments. People are able to make fairly reliable judgments of their own agency --they can assess the extent to which they were or were not the causal agent in producing an action outcome (Metcalfe, Eich & Castel, 2010; Miele, Wager, Mitchell, and Metcalfe, in press), a clearly self-referential metacognition. However, they cannot do so infallibly. Wegner and Wheatley (1999, Wegner, 2003; Wegner, Sparrow & Winerman, 2004) have provided several fascinating experimental examples of errors in these judgments. In one study, participants, wearing headphones, with their hands at their sides, looked at a mirror image of themselves covered by a smock with the hands of a confederate protruding where their own hands would normally be seen. The participants, of course, knew that the hands that they were seeing in the mirror were not their own hands. But if a word for an object was primed (via the headphones) at just the right moment before the hands that looked like their own hands moved, people had a spooky feeling that they had reached for the object. Their judgment of agency, hence, was malleable and subject to illusion.

But while agency judgments can be distorted (as can lower level metacognitions), they are normally accurate. For example, Metcalfe and Greene (2007) showed that college students usually correctly know when they have moved a mouse to catch a target, and when noise-like interference, which distorted their own planned movements, intervened. Knoblich, Stottmeister, and Kircher (2004) showed that while typical adults can detect a distortion in their motor movements, patients with schizophrenia have great difficulty in doing so.

What about non-human animals? The data, so far, are scant but promising on this issue (Couchman, 2011). But, insofar as one component of metacognitive judgments of agency involves action monitoring non-human primates may---given their dexterity and physical competence--be excellent at it. Originally the comparator action monitoring models (Wolpert, Ghahramani, & Jordan, 1995), that form the core of most theoretical views of how people make judgments of agency, were devised as a way of understanding how it is possible for people to make nuanced and complex fast actions. The central idea is that the person has a plan of where and how to move. This plan runs off mentally in real time synchronously with their actual movement, and the feedback from the movement is collated with the expectations from the plan. If the two correspond perfectly, the action proceeds smoothly. If there is a mismatch, then an alteration is needed to correct the movement. This match /mismatch mechanism, devised for motor control, was co-opted by the metacognitive system, to allow people to make judgments of agency: if there is no discrepancy, then the person was in control. If a discrepancy occurred, though, then some outside source was distorting the correspondence between intent and action, and the person was not in full control. Presumably to accomplish acrobatic feats so common in the wild, our primate ancestors would need to have a finely tuned action monitoring system. Whether, like humans, they co-opted it to allow them to have metacognition of agency and perhaps even a concept of the self, we do not know.

CONCLUSION

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

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

How does metacognition relate to self-awareness, then? First of all, we have argued that anoetic and noetic metacognition do not imply self awareness at all. That being the case, even humans may not always be self-aware when making metacognitive judgments (e.g., Son & Kornell, 2005). But autonoetic metacognition (as long as it is not faked) suggests that the individual has self awareness, and an internalized, articulate concept of the self. Now, of course, humans may also be self aware at other times --the argument is only that anoetic and noetic metacognition provide no positive evidence.

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

References

Carruthers, P. (2011). *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford University Press.

Couchman, J. J. (2011). Self-agency in rhesus monkeys. *Biological Letters*, July 12, e-pub ahead of print.

Craik, F. I., Morris, L. W., Morris, R. G., & Loewen, E. R. (1990) Relations between source amnesia and frontal lobe functioning in older adults. *Psychology & Aging, 5*, 148-151.

Descartes, R. (1637). *Discourse on Method*. Cambridge University Press

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition, 24,* 523-533.

Dunn, J. C. (2004). Remember–know: A matter of confidence. *Psychological Review, 111*, 524-542.

Foote, A. L., & Crystal, J. D. (2007). Metacognition in the rat. *Current Biology, 17*, 551-555.

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition, 16*, 309-313.

Gardiner, J. M., Richardson-Klavehn, A., & Ramponi, C. (1998). Limitations of the signal-detection model of the remember–know paradigm: A reply to Hirshman. *Consciousness & Cognition, 7*, 285-288.

Hawkins, S. A. & Hastie, R. (1990): Hindsight: Biased judgements of past events after the outcomes are known. *Psychological Bulletin*, 107, 311-327.

Henkel, L. A., Johnson, M. K., & De Leonardis, D. M. (1998). Aging and source monitoring: Cognitive processes and neuropsychological correlates. *Journal of Experimental Psychology: General, 127*, 251–268.

Hirshman, E. (1998). On the utility of the signal detection model of the remember–know paradigm. *Consciousness & Cognition, 7*, 103-107.

Hoffrage, U., & Pohl, R. (2003). Research on hindsight bias: A rich past, a productive present, and a challenging future. *Memory, 11*, 329- 335.

Humphrey, N. (2006). *Seeing red: A study in consciousness.* Boston: Harvard University Press.

Husserl E. 1929. *Cartesian Meditations and the Paris Lectures.* The Hague: Martinus Nijhoff, 1973.

James, W. (1890). *The Principles of Psychology, Volume 1*. New York: Henry Holt.

Jenkins, A. C. & Mitchell, J. P. (2011). Medial prefrontal cortex subserves diverse forms of self-reflection. *Social Neuroscience, 6(3),* 211-218.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulleting, 114*, 3-28.

Kahneman, D., Knetsch, J., & Thaler, R. (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of Political Economy*, 98, 1325-1348.

Knoblich, G., Stottmeister, F., T.T.J. Kircher, T. T. J. (2004). Self-monitoring in patients with schizophrenia, *Psychological Medicine, 34*, 1561-1569.

Koriat A, Bjork RA, Sheffer L, Bar SK. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General, 133*, 643–656.

Kornell, N. (2009). Metacognition in humans and animals. *Current Directions in Psychological Science, 18*, 11-15.

Kornell, N., Son, L. K., & Terrace, H. S. (2007). Transfer of metacognitive skills and hint seeking in monkeys. *Psychological Science, 18,* 64-71.

Menzel, C. R. (1999). Unprompted recall and reporting of hidden objects by a chimpanzee (Pan troglodytes) after extended delays. *Journal of Comparative Psychology, 113*, 426-434.

Finn, B. & Metcalfe (2010). Scaffolding feedback to maximize long term error correction. *Memory & Cognition, 38*, 951-961.

Metcalfe, J., & Greene, M.J. (2007). Metacognition of agency. *Journal of Experimental Psychology: General, 136*, 184-199.

Metcalfe, J., Eich, T. S., & Castel, A. (2010). Metacognition of agency across the lifespan. *Cognition, 116*, 267-282.

Metcalfe, J., & Kober, H. (2005). Self-reflective consciousness and the projectable self. In H.S. Terrace & J. Metcalfe (Eds.), *The Missing Link in Cognition: Origins of Self-Reflective Consciousness* (pp. 57-83). Oxford, UK: Oxford University Press.

Miele, D. M., Wager, T. D., Mitchell, J. P., & Metcalfe, J. (in press). Dissociating neural correlates of action monitoring and metacognition of agency. *Journal of Cognitive Neuroscience.*

Mitchell, K.J., & Johnson, M.K. (2009). Source monitoring 15 years later: What have we learned from fMRI about the neural mechanisms of source memory? *Psychological Bulletin*, 135, 638-677.

Ochsner, K. N., Beer, J. S., Robertson, E. R., Cooper, J. C., Gabrieli, J. D. E., Kihsltrom, J. K., & D'Esposito, M. (2005). The neural correlates of direct and reflected self-knowledge. *Neuroimage, 28*, 797-814.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26). New York: Academic Press.

Reder, L.M. (1987). Strategy Selection in Question Answering. *Cognitive Psychology, 19*, 90-138.

Reder, L.M. & Schunn, C.D. (1996). Metacognition Does Not Imply Awareness: Strategy Choice is Governed by Implicit Learning and Memory. In Reder, L.M. (Ed.), *Implicit Memory and Metacognition*. Mahwah, NJ: L. Erlbaum, pp. 45-77.

Russell, B. (1997*). The Problems of Philosophy*. Oxford University Press, New York.

Rosenbaum, R.S., Köhler, S., Schacter, D. L., Moscovitch, M., Westmacott, R., Black, S.E., Gao, F., & Tulving, E. (2005). The case of K.C.: Contributions of a memory-impaired person to memory theory. *Neuropsychologia, 43*, 989-1021.

Rosenthal, D. (2000) Consciousness, content, and metacognitive judgements. *Consciousness Cognition, 9*, 203–214.

Schwartz, B. L., & Metcalfe, J. (in press). Tip-of-the-tongue (TOT) states: Retrieval, behavior, and experience. *Memory & Cognition*.

Shields, W. E., Smith, J. D., Guttmannova, K., & Washburn, D. A. 2005 Confidence judgments by humans and rhesus monkeys. *Journal of General Psychology 132*, 165-186.

Searle, J. R. (1992). *The Rediscovery of the Mind*, MIT Press.

Smith J. D. (2009). The study of animal metacognition. *Trends in Cognitive Science.* 13, 389–396.

Smith, J. D., Beran, M. J., Couchman, J. J., Coutinho, M. V. C. & Boomer, J. B. (2009). Animal metacognition: Problems and prospects. *Comparative Cognition and Behavior Reviews, 4*, 40-53.

Smith J. D., Schull, J., Strote, J., McGee, K., Egnor, R., & Erb, L. (1995). The uncertain response in the bottlenosed dolphin (Tursiops truncatus). *Journal of Experimental Psychology: General, 124*, 391-408.

Smith, J. D., Shields, W. E., Allendoerfer, K. R., and Washburn, W. A. (1998). Memory monitoring by animals and humans. *Journal of Experimental Psychology: General*, 127, 227-250.

Son, L. K., & Kornell, N. (2005). Meta-confidence judgments in rhesus macaques: Explicit versus implicit mechanisms. In Terrace, H.S. & Metcalfe, J. (Eds.), *The Missing Link in Cognition: Origins of Self-Knowing Consciousness*. Oxford University Press.

Son, L. K., & Metcalfe, J. (2005). Judgments of Learning: Evidence for a Two-Stage Model. *Memory & Cognition, 33*, 1116-1129.

Staddon, J. E. R, Jozefowiez, J. & Cerutti, D. T. *(2007).* Metacognition: A Problem not a Process. *PsyCrit.*

Terrace, H. S., & Son, L. K. (2009). Comparative metacognition. *Current Opinion in Neurobiology, 19*, 67-74.

Thaler, R. (1980). Towards a positive theory of consumer choice. *Journal of Economic Behavior and Organization*, *1*, 39-60.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26*, 1-12.

Tulving, E. (1984). Elements of episodic memory. *Behavioral and Brain Sciences, 7*, 223 - 268.

Tulving, E. (2005). Episodic memory and autonoesis: Uniquely human? In H. S. Terrace & J. Metcalfe (Eds.), *The Missing Link in Cognition* (pp. 4-56). New York. Oxford University Press.

Wang, L., Metzak, P.D., & Woodward, T.S. (2010). Aberrant connectivity during self-other source monitoring in schizophrenia. *Schizophrenia Research*, 125, 136 - 142.

Washburn, D. A., Gulledge, J. P., Beran, M. J., & Smith, J. D. 2010 With his memory erased, a monkey knows he is uncertain. *Biology Letters 6*, 160-162.

Wegner, D. M. (2003). *The illusion of conscious will.* Cambridge, MA: MIT Press.

Wegner, D. M., Sparrow, B., & Winerman, L. (2004). Vicarious agency: Experiencing control over the movements of others. *Journal of Personality and Social Psychology, 86,* 838–848.

Wegner, D. M. & Wheatley, T. (1999). Apparent mental causation: Sources of the experience of will. *American Psychologist,* 54, 480-92.

Wheeler, M. A., Stuss, D. T., & Tulving, E. (1997). Toward a theory of episodic memory: The frontal lobes and autonoetic consciousness. *Psychological Bulletin, 121*, 331–354.

Wixted, J. T., & Stretch, V. (2004). In defense of the signal detection interpretation of remember/know judgments. *Psychonomics Bulletin & Review, 11*, 616-641.

Wolk, D.A., Schacter, D.L., Lygizos, M. Sen, N.M., Holcomb, P.J., Daffner, K.R., Budson, A.E. (2006). ERP correlates of recognition memory: Effects of retention interval and false alarms. *Brain Research, 1096*, 148-162.

Wolpert, D.M., Ghahramani, Z. and Jordan, M.I. (1995), An internal model for sensorimotor integration, *Science, 269*,1880–1882.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory & Language, 46*, 441-517.

Zalla T., Stopin, A., Ahade, S., Sav, A. M., & Leboyer, M. (2008). Faux pas detection and intentional action in Asperger Syndrome. A replication on a French sample. *Journal of Autism and Developmental Disorders, 39*, 373-382