Chapter 6

MiniMeta: in search of minimal criteria for metacognition

Josef Perner

This is a transitory chapter linking Section I on comparative animal studies with the upcoming Section II on developmental studies with human infants and children. I am surveying the animal literature for useful methods to use with children when trying to avoid reliance on heavy metacognitive verbalization of mental states.

A sizeable research effort has evolved over the last 15 years with the aim of demonstrating metacognition in animals. The techniques involved are obviously interesting to the developmental psychologist for use on young (pre- or minimally verbal) children. Unfortunately, there are still lingering doubts (e.g. Carruthers 2008; Metcalfe 2008) as to whether these studies do show metacognition. These doubts are due to novel alternative interpretations of existing studies that have not yet been ruled out and to questions of definition. Not that there are no clear definitions, but the feeling is that these may be too restrictive for research with non-verbal organisms cutting out interesting lower level metacognitive abilities (Proust 2007, 2010; Smith 2011). My prime objective is not to arbiter between animals being or not being capable of metacognition. I want to

Fig. 6.1 Minimally meta.
look at the techniques to see whether they can, in principle, provide evidence of metacognition and thus have potential for adoption to use with young children. For this it is necessary in the first section to provide clarification on the different meanings of 'metacognition' and which one captures best its intuitive meaning. In the second section I elaborate on two pernicious problems for getting conclusive evidence for metacognition from behavioural investigations and illustrate these problems on some existing studies. In the third section a more systematic effort is made to look at the most promising methods and see how they fare in view of these interpretational problems. In the fourth section I explore the grey area between object-level cognition and metalevel cognition for a coherent way of identifying minimally metacognitive abilities: the MiniMeta project.

**Varieties of ‘metacognition’**

Even when we are after minimal criteria for metacognition, we still should have some idea before the search of what it is, for which we seek criteria. We need to clarify what we mean by 'cognition', 'meta-', and their combination. To start with *cognition*: it has traditionally been used to denote one of three kinds of mental processes that deal with how the world is as opposed to those concerned with how we want the world to be and how we feel about it (Hilgard 1980: cognition—conation—affect). In Cognitive Science the term cognition has taken on a wider meaning (Fodor 1978; Wimmer and Perner 1979). Cognitive analysis of the mind assumes that all mental processes consist of transformations of representations. So not only what we know and think but also how we want the world to be and how we feel about it are open to cognitive analysis.

The term ‘meta-‘ is Greek and means beyond or after. In this sense we would interpret ‘metacognition’ as special cognition, i.e. something that goes beyond standard cognition. This very basic meaning of ‘going beyond’ has, however, typically been applied in a recursive fashion. For instance, *metalinguage* goes beyond language in the specific sense of *language about language*, or *metamathematics* is understood as the *study of mathematics using mathematical methods* (Wikipedia). In this tradition we end up with the meaning of ‘metacognition’ as cognition about cognition. Taking thinking as the most typical cognitive state it is also often defined as thinking about thinking (e.g. Flavell 1979; Dunlosky and Metcalfe 2009).

If we take cognitions the Cognitive Science way as based on representations, then, for example, a feeling about knowledge (I feel bad about not knowing the answer), would also be a metacognition. Moreover, in this view metacognition implies metarepresentation, a term being worked hard in this volume. Equating metacognition with metarepresentation is considered by some an appropriate (Carruthers and Ritchie Chapter 5) but by others as too demanding a definition (Couchman et al. Chapter 1; Proust Chapter 14).

The clearest definition of metarepresentation is still the original one by Pylyshyn (1978) as representation of the representational relationship itself. It is however, very demanding. A representational relation relates the thing that represents (the representational vehicle— in case of a picture, the marks on the paper) with the represented, which consists of the representational target (the object or scene shown in the picture) and the representational content (the way in which the

---

1 The best known word formed with meta- is probably metaphysics: the word ‘metaphysics’ is derived from a collective title of the 14 books by Aristotle that we currently think of as making up ‘Aristotle’s Metaphysics’ Aristotle himself did not know the word. . . At least one hundred years after Aristotle’s death, an editor of his works (in all probability, Andronicus of Rhodes) entitled those 14 books ‘Ta meta ta phusika’—‘the after the physicals’ or ‘the ones after the physical ones’ (van Inwagen, 2010, p. 2).

2 Being a clear case of metacognition hinges on the use of the word ‘about’. The state we refer to with, e.g. ‘feeling good about knowing the answer’ is metacognition, while ‘feeling good because one knew the answer,’ is a debatable case. Thanks to an anonymous reviewer for alerting me to this.
picture shows the target to be: a horse in the distance as a dot; Goodman, 1976). So representing
the representational relationship would at least require some sensitivity to all the relata—least
that was my interpretation (Perner 1991). This strict version of metarepresentation admits to
true recursion, e.g. I am thinking that you are thinking that I am thinking about something. It
can, therefore, also be referred to as ‘recursive cognition’.

One common weaker understanding of metarepresentation is *representations that represent the
content of a representation* (Leslie and Roth 1993, p. 91, M-representation; Sperber 2000, p. 117).
In that case if one represents anything that does not exist (e.g. a unicorn, or a pretend scenario)
then one cannot represent anything existing (no target) only the content of someone’s represen-
tation. This can be the content of one’s own previous thought. Any train of thought about a stable
non-existing entity would then qualify as metarepresentation. In this sense metarepresentations
are not truly recursive. When I think of a unicorn no recursion is possible, since the unicorn itself
is not a representation (vehicle) with representational content.

Another conceivable weaker interpretation is to understand metarepresentation as *representation
of something that happens to be a representation*, without representing its representational relation
to what it represents—falling short of Pylyshyn’s definition. A simple and intuitive example
would be when thinking of someone thinking, where thinking is simply understood as an activ-
ity—that people in the pose of Rodin’s *Thinker* are engaged in—without any understanding that
this process can only be a process of thinking if the thinker is thinking about something, i.e. that
his thinking has representational content. Here too, metacognition in this sense is not recursive
because the thought-about thinking has no content that allows for recursive application.

In sum, we can take a narrow-scope interpretation of metacognition in the spirit of the classical
trilogy of mental states (Hilgard 1980) as concerning only epistemic (cognitive) mental states or
take a wider interpretation including all mental states in the spirit of Cognitive Science. I will stick
to the wider usage. We can also take different views on the effects of the prefix ‘meta-’ and how it
produces at least three different wider scope meanings of ‘metacognition’.

1. Special Cognition: cognition beyond standard cognition (on the way to metacognition):
   a. Vehicle Reference: referring to something that happens to be a cognition.

---

3 Minimally only the existence of the relata has to be acknowledged, not any precise form of them, as Leslie
and Roth (1993) once thought was being claimed. That is, in the case of mental representation, one need
not—and we typically do not—represent the fact that their vehicles are neural states, and do not represent
whether the content is in form of linguistic or analogue pictorial form, etc.

4 Pylyshyn’s definition implies an understanding that the represented representation is a representation,
which in turn implies some minimal understanding of what a representation is, namely as characterized by
its representational relationship to what it represents. This led me to define metarepresentation as *repre-
sentation of a representation as a representation* (Perner 1991). In analogy, metacognition would consist of
cognition about cognitions as cognitions, which implies some understanding of what makes cognitions
what they are, namely their representational content.

5 The case of pretence seems to contradict this conclusion. For, I can pretend that my pretend character is pre-
tending something (metapretence). My claim would be that in this case I need to understand the pretended
pretence as a representation, i.e. as something that has content and not just something that is content.

6 This is similar to what Proust (2010 p. 7; chapter 14 this volume) means with ‘de re’ in connection with
how epistemic emotions refer to the epistemic state.

7 Perner (1991) gave an illustration in terms of metalinguistics, of illiterate workmen referring to objects in
the shape of B, A, and R, which they have to mount atop the entry to a bar, as letters. Although their con-
versations about these objects refer to linguistic entities (letters) their discourse is ‘metalinguistic’ only in
this weak sense discussed here.
b. Content Reference: referring to the content of a cognition.

2. Recursive Cognition: cognition about cognition (as cognition):
   c. Representational Reference: cognition about cognition as cognition.

What makes the claim that animals or young children are able of metacognition is the idea that they are able of recursive cognition (cognition about cognition as cognition). For, if it were just an ability to recognize being in a state of thinking (cognition about cognition), without any concern for the intentional content of the thinking, then this ability would be similar to recognizing that one is in a state of digesting—which would also be an interesting ‘reflective’ ability but not one for which the term ‘metacognition’ would be most natural.

Nevertheless it is useful to be aware of the looser meanings of ‘metacognition’ outlined above. They may explain why some behaviour feels intuitively metacognitive even though cognitive analysis shows no need to be based on recursive cognition. As shown in Fig. 6.2 these cases could be considered to lie on the slope from ordinary object-level cognition to full blown recursive metacognition at the higher level. In this sense they can be considered minimally metacognitive or ‘minimeta’. Fig. 6.2 also shows additional steps to be discussed later.

A pair of pernicious problems

Inferring metacognitive abilities from behavioural data has proven to be less than straightforward. Here I point out some of the deeper methodological and conceptual problems, which make it very difficult to infer metacognition—understood as recursive cognition—on the basis of behavioural indicators. I intend to illustrate these problems on existing techniques, which have been almost exclusively developed in the comparative animal literature.

I want to emphasize that my goals are to find techniques that can overcome these problems for potential use with pre- or low-verbal children and to draw awareness to these general problems affecting our theory of mental representation. In any case, I do not want to conclude my survey with denying animals any ability for metacognition. However, in checking whether experimental paradigms surpass my problems or not I have to check particular studies. Here details of the studies often matter, and one can be easily accused of ‘cherry picking’, i.e. picking only on the easy targets. This would be counterproductive for my enterprise; I should pick the hard targets (‘inverse cherry picking’). They are more likely to lead me to promising methods. Reviewers of an early draft have already directed me to some tougher targets. I also hope that my arguments will
be taken seriously enough to be controlled for in future research, which is a generally desirable feature (Shettleworth 2010).

**BEING in a state versus KNOWING that one is in this state**

To see the depth of this problem a brief (I promise!) excursion into why Cognitive Psychology needs to posit mental representations will help. Psychology attempts to explain intelligent or adaptive behaviour, which consists of movements that change the world in the service of reaching a goal by adapting to the given circumstances. So, goals and circumstances determine the movement. For this to be possible there must be some causal link from goals and circumstances to an animal’s movement. Since there is no direct physical cause discernible, we need to posit some ‘internal’ entities that correspond to external situations and goal states, i.e. neural states that represent these external causes. These neural states do actually exist (unlike non-existing goals) and, thus, are able to cause behaviour.

Now we go one step up considering adaptive responses to other sentient beings. If I want to predict whether you will come to our meeting I have to know whether you know that it takes place (a clearly recursive representation). So if I can make a correct prediction we have some basis for claiming that I must have a metarepresentation of your knowledge. Now we go one step in and consider knowledge of one’s own knowledge state (the central case of metacognition). If I need to predict whether I will go to the meeting I just have to know that it takes place. I do not need to know that I know it takes place. This would be superfluous. Hence, from my correct prediction of what I will do we cannot convincingly infer metacognition as a recursive concern about my own knowledge.

Whence this asymmetry? Proust (2007) made much of this asymmetry to conclude that metacognition differs essentially from metarepresentation. My explanation is very simple. Your knowledge state cannot take causal influence on my predictions unless I metarepresent it by knowing that you know about the meeting. In contrast, my own knowledge about the meeting is an integral part of my cognitive system involved in generating predictions. It can, therefore, causally effect the prediction of my behaviour directly, without need to metarepresent my knowledge.

One reason why we find it difficult to distinguish knowing from the recursive knowing that one knows may be our Cartesian intuition, where our mind is considered to be transparent to itself (Churchland 1984). This also explains the reluctance with which the idea of unconscious knowledge (knowledge of which one does not know that one has it) was accepted in scientific psychology. Clearly, with such strong common sense intuitions it is tempting to equate an effect of BEING in a state of knowing with KNOWING that one knows. Yet, keeping these two things distinct is essential. Every behaviour that takes environmental conditions into account depends on knowledge of the environment. If knowing were the same as knowing that one knows then just about every behaviour would be evidence for metacognition.

Now, Call and Carpenter (2001) claimed to have demonstrated metacognitive abilities by showing that chimpanzees as well as very young children reacted differently depending on being knowledgeable or ignorant about the location of a reward. Participants could see one of two tubes being baited. In one condition (full knowledge) they saw which tube this was, in the other (partial ignorance) they did not. Then they had to point to a tube. If it contained the bait they got it, otherwise they did not. Chimpanzees as well as 2½ year old children were able to adapt their behaviour to their knowledge or partial ignorance. They looked first through the tubes to see which one

---

8 In particular, goal states do not even exist at the time behaviour is caused. So how could they possibly have causal efficacy.
was baited more often when they were partially ignorant than when they knew where the bait was 
(superfluous looking). This behavioural difference between the knowledge condition (not look- 
ing) and the partial ignorance condition (looking into the tubes) is interpreted as evidence of 
metacognition (see Call Chapter 4, this volume, for more details on this test and other related 
tests). How can that be? Chimpanzees, who know where the bait is, go for it and do not look 
around; when they do not know (total ignorance) where it is they look around; when they know 
roughly where it is (partial ignorance) they go there and then look around within that region. This 
behaviour can be governed by the degree of the chimpanzee’s knowledge/ignorance without any 
recursive cognitions about his degree of knowledge. In fact, what would it help the chimp to know 
that he knows that the bait is in one of tubes and to know that he does not know in which one? So 
there is no compelling evidence for recursive cognition (metacognition) in these experiments. 

One question though is whether the evidence can be firmed up by other features of the chimpan-
zee’s’ behaviour (see ‘Information seeking’ section). Another question that remains is why we, or 
many of us, think or feel that metacognition is involved. This question will be taken up in the 
penultimate section on MiniMeta.

Internal—external: representing a state of the world versus 
representing an inner state (representation) caused by 
the external state.

Representing a state of the world is standard object-level cognition. Representing an internal state 
caused by the external state smacks of potential metacognition. Is this impression warranted? 
Let me illustrate the problem with a little thought experiment. When I enter an overheated 
room I will get hot. This internal physical state is not a cognitive state. It is, however, necessary for 
giving rise to two kinds of cognitions: feeling hot and realizing that this is a hot room. In either 
case I might open the window. From my opening the window an external observer could not tell 
whether I do so because I feel hot (a cognition about my inner state) or because I realized the room 
is too hot (a cognition about an external state of the world). That is, the observer cannot determine 
whether my cognitive system is concerned with my inner state or with the state of the room. 
This indeterminism, I argue, affects also most experimental demonstrations of metacognition, 
in particular, those based on the so-called ‘opt-out’ paradigm. I use Smith et al.’s (1997) classic 
psychophysical pixel density test (see Couchman et al. Chapter 1, this volume) to illustrate the 
problem. In their study rhesus monkeys were trained to make one response (‘d’) when pixel den-
sity was dense and make another response (‘s’) when pixel density was sparse. These response 
options had task dependent outcomes. For correct responses monkeys got a food reward, for incor-
correct responses they had to endure a time out period. Monkeys learned this task to perfection at the 
clear density values but made errors close to the border line of sparse and dense. After reasonable 
mastery a third, task independent response option (‘o’) was introduced. Making this response 
resulted in a standard outcome (independent of pixel density) of medium value.9 Both monkeys 

9 Since this payoff structure was used in many studies with quite different kinds of rewards or non-rewards 
I try to give a more abstract characterization that, in my view, does justice to all variants. When the correct 
option is taken the subjective expected utility (SEU) of the outcome is high, if the wrong option is taken 
it is low. Clearly this must be so, because if animals subjectively valued a food pellet less than a time out, 
they would never learn to respond correctly. Similarly, they must be sensitive to the probability with 
which responding with ‘s’ or ‘d’ results in the valued outcome, hence subjective expected utility. The 
critical element of all opt-out paradigms is an additional response option with a task independent out-
come. I denote this option with ‘o’ (for opt-out). Its outcome is first of all independent of what the set task 
is (e.g. the density discrimination) and it must have an SEU below that of the correct response or else the
learned to use the o-response appropriately in the region of objective highest uncertainty, i.e. most often at those density values at which their use of the other two responses was equi-frequent. The authors’ ‘metacognition’ interpretation of this result is that the monkeys learned to use the o-response in response to detecting their own uncertainty or to feeling uncertain. And many have followed suit with this interpretation.

Unfortunately, this interpretation suffers from the same indeterminism as outlined above with ‘being hot’. Being uncertain is, like being hot, an inner state. It is, unlike being hot, also a cognitive state—but it is not a metacognitive state (see earlier section ‘BEING in a state versus KNOWING that one is in this state’). This inner state is necessary for two types of further cognitions required to guide a learned response: realizing that I am uncertain (a clear metacognition), or realizing that this is a difficult problem (a cognition about an external state of the world). Either one of them can be the basis for my adaptively emitting the o-response. Consequently, an observer of my o-response cannot determine whether my cognitive system is concerned with my inner state of uncertainty or with the difficulty of the task because the behavioural responses afforded to the monkeys are insufficient to distinguish these two possibilities.

A common objection comes with the intuition that judging the difficulty of a task may implicitly already be metacognitive (one could call difficulty a ‘metacognitive’ property). My Anonymous Reviewer objected to my analysis that difficult is a subjective notion and therefore it is hard to see how an animal could assess difficulty without any consideration of its own mental states. It is true, tasks in themselves do not come in degrees of difficulty. Their difficulty depends on how they are managed by the person trying to solve them. But that kind of subjectivity holds for a lot of properties, e.g. relative size (bigger than me, I run; smaller than me, I stay), being hot, or being disgusting (it is disgusting, leave it; it is ok, eat it). Being hot and being disgusting are similar to being difficult in that they depend on evaluation of an external object or event in relation to an internal ability or reaction to it. The maggots on a piece of rotting meat make me feel disgust and I judge the maggots and the whole scene disgusting. Similarly, I read a question on a test, no answer comes to mind, and I judge it difficult and turn the page. If the subjectivity of that judgement is a reliable indicator of metacognition then my quest is at an end: we simply look when children become able to tell when a room feels subjectively hot or when they become able to reject food they deem disgusting. Few would find such a demonstration convincing. So why would learning to skip difficult test items be evidence for metacognition?

Michael Beran in his comments wondered why students’ ability to strategically skip multiple choice items is considered a metacognitive skill in all of the traditional metacognition literature in educational science. I surmise that this is so because we have a common sense theory of why test items are difficult. They are difficult because they exceed our knowledge. We tacitly assume that all students share that theory and when they judge an item difficult (not intrinsically metacognitive) they are aware that this is because they don’t know the answer (which makes animals would always choose ‘o’ (except for erratic exploration). Its SEU must also lie above the average SEU of correct and all incorrect options (randomly choosing any option), or else it would not be chosen on a regular basis. Last but not least, the animal has always the option to not choose any of the above and do nothing (or one of the causally irrelevant activities). Importantly, the SEU of the other responses must be higher than the SEU of doing nothing. I take it that this characterization of the task is a simple logical consequence of the assumption that animals prefer actions with higher SEU over actions with lower SEU outcomes.

The decision to skip can be based on a sense of difficulty not being able to produce anything easily. There is an interesting link to the developmental evidence (Kloo and Rohwer Chapter 10, this volume). Younger children pin their understanding of the word ‘know’ on being able to give an answer to a question. So they mistake easy guesses for knowing.
it metacognitive). The educational literature is not tuned to asking the foundational issues, which
are the focus of this volume, of when one can infer metacognition from task performance. It
works under the plausible premise that students metacognitively understand the link between
their knowledge and task difficulty, i.e. that they understand or should understand that if the test
is difficult then they should have learned harder for it.

Now let me build up a contrasting intuition with an imagined blindsight patient. On each trial
we show him in his blind field an X or an O, or sometimes nothing. He is instructed to say ‘X’ or
‘O’ or press a button to move to the next trial. When he calls out the same letter as is displayed he
receives a nice chime, otherwise a grating scratch, unless he presses the move-on button. From his
subjective point of view he makes responses totally unrelated to the display where he can’t see
anything anyway, but feels a natural tendency to say either ‘X’ or ‘O’ (because the unconsciously
perceived stimulus primes these responses). However, sometimes this tendency is missing
(because nothing is presented and no priming takes place) and then he presses the move-on but-
ton. If our patient performs well in this task, can we infer that he is having metacognitions of a
minimal kind: ‘I saw an X’ or ‘I knew there was an X’? Blindsight patients evidently have informa-
tion (unconscious knowledge) about the stimuli, or else they could not respond contingently.
What they lack is awareness of their mental state (metacognition) with which they behold that
information. So it seems that from the stimulus-contingent behaviour we cannot, and in this case
should not, infer any metacognitive insights.

This undecidability of whether a representation of an inner mental state or of an outer worldly
state drives our behaviour is a problem of even wider significance than my examples might sug-
gest: a really pernicious problem. Evans (1982) proposed the notion of an ascent routine that
enables us to ascend from what we (in our judgement, from our perspective) consider a fact to the
mental realm and attribute with logical impunity a corresponding belief to ourselves. For instance,
if I take for a fact that Obama was born in the USA, then I must believe that he was born in the
USA. Gordon (1995) extended this idea in the service of simulation theory to all mental states.11
This correspondence between own mental states and external properties makes it a serious prob-
lem of ever knowing the cognitive basis of behaviour that occurs consistently with a particular
mental state. Does it depend on recursive cognition about that mental state or on object-level
cognition about the external fact from which one can ascend to that mental state?

Summary

I have pointed out two problems for showing metacognition (recursive cognition) on a behav-
ioural basis, and illustrated these problems with examples from classical metacognition experi-
ments. I now go through some of the more recent evidence and show that all (the toughest nuts I
could find) founder in their metacognitive claims on at least one of these problems.

11 Whatever is a fact in my judgement is also (unfailingly) something I believe to be so; whatever is in my
judgement frightening (as opposed to in my judgement frightening to others) is also something I am
frightened of; whatever is in my judgement desirable is something I desire, etc. A good intuition check for
these connections is an extension of Moore’s paradox for belief (Moore 1942; Gordon 2007): The claim ‘I
believe that p and not p’ seems contradictory without being a logical contradiction. By extension a similar
non-logical contradictoriness seems to adhere to: ‘I desire that p and p is (in my judgement) not desira-
ble,’ and ‘I am frightened by x and x is (in my judgement) not frightening’.
That ascent routines exist for every mental state is a strong claim. A problematic case seems to be ‘hope’
(Goldman cited by Gordon 2007). I maintain that an ascent routine for hope is possible provided one
admits conditional statements about the world. Given that ‘it would be a good thing if our CO₂ emissions
will be reduced,’ I can ascend to attributing to myself: ‘I hope that CO₂ emissions will be reduced’.
Recursive cognition in animals: which is the best method to detect it?

Information seeking

Call and Carpenter (2001) introduced the information seeking paradigm used for chimpanzees and very young children that I described in the earlier ‘BEING in a state versus KNOWING that one is in this state’ section. Members of both species differentiated between trials where they knew in which of several tubes a bait was hidden and trials where they only had partial knowledge that the bait was in one of the tubes but not which. They looked through the tubes more frequently in the partial than full knowledge condition in order to find out where the bait was before commit-ting themselves to a final choice.

My analysis concluded that there was no evidence for recursive cognition in these data. When an animal wants food it looks around to find some. When it knows where food is it goes straight there. Nobody would claim metacognition being involved. The tubes experiment only shows that animals can restrict their exploratory behaviour to the region within which they know (partial knowledge) that food can be found.

There are now several follow-up studies. Call (chapter 4 this volume) agrees that random search exploratory behaviour does not need metacognition to be explained. But he argues that metacognition is required to explain the very specific behaviour that animals show in pursuit of knowledge in these new studies. So this is the cherry to pick because it is the toughest nut to crack.

Krachun and Call (2009) showed that animals not only look inside the relevant containers, but they are also very adept at positioning themselves in relation to the differently shaped containers so that they could look inside: ‘The crucial aspect of this study was that owing to the containers’ diverse geometry and their position on the platform, subjects had to position themselves in different locations depending on the container to spy the food’ (Call Chapter 4, this volume). This is an impressive cognitive feat, but how does it relate to metacognition? Complexity of cognition does not make it metacognitive, as Penn and Povinelli (2007) have emphasized in the context of theory of mind. In fact, my worries go beyond the question of whether this sophisticated behaviour needs metacognition to be explained. I wonder whether metacognition could play any helpful role at all. Let me go back to first principles.

An animal’s behaviour is guided by a goal and knowledge of how to achieve that goal. So if an animal wants to get to the food it will go where it knows the food is. If there is no such knowledge the desire in combination with its ignorance will trigger search. This can be just random looking around. No metacognition need be involved as Call agrees. If the animal has partial knowledge of where the food is the random search will be constrained by that knowledge. Still no metacognition required.

Now the question is how can we explain animals that do not just look around randomly but who engage in strategically locating themselves in relation to containers in order to get a look inside. I suspect the critical assumption here is that random looking can be understood as a ‘response’ to not knowing where the food is, while guided looking needs to be understood as instrumental behaviour in order to achieve a goal (getting a good look inside). So we need a goal. Evidently, wanting the food in combination with insufficient knowledge about its location generates a goal of wanting to get a look at the food. The animal must have general knowledge how to achieve this goal. For instance, it has to know that it needs to look inside every corner, niche, or container within the relevant search space and it needs knowledge about how to position itself in order to get a look inside containers, etc.
This seems an explanation without reference to metacognition. Is there a gap in this explanation that needs to be filled by metacognition? One such gap may be the triggering of the new goal to get a look at the food. The claim is that this can be triggered by wanting the food and by not knowing where it is, which are intentional states but not metacognitive states (see section ‘BEING in a state versus KNOWING that one is in this state’). In fact, for explaining how lack of knowledge triggers the desire for more information I can see no explanatory advantage of resorting to metacognition. We would have to assume that wanting food and not knowing where it is first leads the animal becoming metacognitively aware that it does not know where the food is, then that metacognition triggers the desire to get a look at the food. Why would that explanation be more elegant or more complete than the original?12

There may be another gap residing within my account just given: the goal of getting a good look includes a metacognitive understanding that a good look will provide a particular visual experience. Krachun and Call’s (2009) frequent reference to ‘visual perspective taking’ suggests such a tacit assumption. Certainly, the point of the looking is to get a new visual perspective and with it a new experience and knowledge. But does the animal have to understand this? Is it not enough to aim for a good look rather than a good visual experience? The deeper reason for the good look need not be apparent to the animal, only to evolution, which provided the animal with the desire for getting good looks.

Another critical gap in my argument may be the explanation of how the animal can position itself correctly in relation to a container in order to get a good look inside. For this ability the animal needs first of all intricate knowledge of how to achieve a good look. I cannot see how any metacognitive knowledge can add anything.

In sum, the intricacies of animals’ guided search call for a goal to get a good look inside all relevant places in the search space and intricate knowledge about how to achieve such a good look under specific circumstances. How metaknowledge can help here remains mysterious. So my argument is not that an alternative explanation can be found for guided search by cherry picking case specific explanations, but that reference to metacognition does not contribute in any way to an explanation for the animals impressive search behaviour.

Beran and Smith (2011) elaborated on a quite intricate information seeking procedure used unsuccessfully on pigeons before (Roberts et al. 2009). Monkeys were shaped to choose (operating a joystick) one icon of two that revealed a sample stimulus. They then had to choose the other icon, upon which the sample disappeared and three test stimuli appeared. One of them was the same as the earlier sample. For choosing the matching stimulus a reward was given. Both species, rhesus macaques as well as capuchin monkeys, succeeded on this part of the test which pigeons solidly had failed. In a series of further sessions the other three possible variations of occluded/visible sample and occluded/visible test stimuli were successively introduced. The most difficult final trial block was one where all four versions were presented intermixed. Four of eight macaques had failed. In a series of further sessions the other three possible variations of occluded/visible sample and occluded/visible test stimuli were successively introduced. The most difficult final trial block was one where all four versions were presented intermixed. Four of eight macaques were able to reach this final phase and learn the optimal response for the last version added, which required to reveal the sample while the test options were already visible (Occluded Sample—Revealed Comparisons; experiment 2). In contrast, not one of seven capuchins managed that level (experiment 3).

Mastery of this condition is intricate. However, the intricacy lies first of all in realizing that one needs to choose the test item that is the same as the sample and for that one needs the sample.

12 Importantly, I question only how metacognition could possibly improve my explanation, i.e. make it more watertight. I am not questioning that metacognition could improve the animal’s way of optimizing its search for more information or other ways of dealing with this situation.
When the sample is or was not present then one needs to produce it by clicking the ‘produce sample’ icon. But this does not involve any (clearly metacognitive) recursive cognitions like ‘Where is information about the sample?’ or ‘Where can I see the sample?’. It is not even clear why such cognitions, if animals were capable of them, were of any help to them beyond the basic cognition: ‘What is the sample?’.

An interesting question here is why pigeons are apparently incapable of looking for a sample when they are presumably perfectly able to look for food. A plausible reason would be that looking for a sample in order to find a matching item later is of course a greater intellectual challenge than looking for food to consume. Being capable of greater cognitive complexity does, however, not make for metacognition.

Another interesting question concerns the source of the difficulty in the final phase of this study, which some rhesus macaques mastered but no capuchin monkey was able to. Beran (pers. comm.) sees the critical aspect in the need for monitoring: ‘Thus, there is a monitoring component to this, not just a goal cognition, because information seeking behaviour (or, in the condition where all information is presented already—the lack of information seeking behaviour) is driven by not just goal cognitions but also assessment of the current environment against that goal cognition’.

I agree with this assessment, but it strikes me that it is a description of the most basic practical reasoning we ascribe to any organism whose movements we call ‘behaviour’.

In any case I can see no real need for metacognitive monitoring (monitoring one’s mental states over and above monitoring the environment). For instance, in the two conditions in which the test stimuli are visible the animal wants to press the stimulus that matches the sample. If the sample is visible the animal knows what it is and clicks on the matching item. If the sample is not visible the animal generates the subgoal of producing the sample by clicking on the ‘get sample’ icon. The sample appears, the animal knows what the sample is, and it clicks on the matching test item. Evidently the animal has to go through several cognitive states of wanting to get something, being ignorant of some things and knowing other things. But what help could be provided by additional metacognitions: the animal knowing that it wants to produce the sample, knowing that it does not know what the sample is, knowing that it sees the sample, etc.? In conclusion: there is no convincing evidence from this approach that recursive cognition must be involved.

Call (2010) showed with the tubes set up that great apes (chimpanzees, gorillas, orangutans, and bonobos) checked the tubes more often when partially ignorant than when knowledgeable. In addition (experiment 3: ‘passport effect’), the apes checked more often when the bait was particularly attractive (grape) than when it was not (carrot). A small difference of less than 10%, but it occurred under partial ignorance as much as under full knowledge (when apes had seen which tube the reward was put inside). This effect (not the effect of ignorance vs. knowledge) could be due to a simple preference for looking at attractive rewards more than at less attractive ones. So at this point the results are not very telling. But let us assume this preference factor can be excluded, e.g. they know the bait is not for them but for the caretaker. They still check what the caretaker gets but check as often for high- as for low-quality food. In that case the difference in looking would get closer to the theoretical significance that Call was aiming for. The ‘superfluous’ checking would seem triggered by being afraid of getting it wrong.

However, is the reason for checking really the fear of getting it wrong, or rather fear of it not being there anymore? My intuition about my checking my luggage for my passport is that I check

---

13 Josep Call refers to the ‘superfluous’ checking whether the object is where you know it is as his ‘passport effect’, because when he travels he packs the passport the night before and then keeps checking every so often to reassure himself that it is still there even though he very well knows it is there.
that I have put it in there. I am not checking whether my assumption (belief) that it is in there may be wrong. Or at least, my intuition is not clear enough to tell between these two reasons. And my intuition as to why the animal checks the food in the tubes is even less clear. Does it check the fact that the bait is indeed in the respective tube or does it check its knowledge of this fact? As the discussion of Internal-external (ascent routines in section 'Internal–external') has made clear this question cannot be easily answered and not on existing data.

**Opt-out**

I have described the basic opt-out paradigm (version by Smith et al. 1997) in section 'Internal–external' as an illustration of how the internal–external problem affects interpretation of these data. In particular it makes it difficult to tell whether the animal responded to being uncertain or to being faced with a difficult task. I should point out that the basic paradigm is also affected by my other pernicious problem ('BEING in a state versus KNOWING that one is in this state' section). Even if animals do respond to *being uncertain*, it would not be evidence for recursive metacognition, for it would be a response caused by the animal *BEING* uncertain and not by the animal KNOWING that it is uncertain.

Interestingly, the being-versus-knowing-that-one-is-problem takes a slight, but relevantly different role in the opt-out paradigm than in the information seeking case discussed in the 'BEING in a state versus KNOWING that one is in this state' section. Lack of knowledge, presumably—or so I have argued, triggers automatically a desire for engaging in information gathering activity. For this to happen, the animal does not need to know that it lacks sufficient knowledge. This is different in the opt-out experiment where uncertainty (lack of knowledge) does not trigger anything that would lead to an o-response. The o-response has to be learned and become associated with the state of being uncertain. Perhaps one could argue that mental states like uncertainty can only cause innately specified effects (looking around randomly) but an animal cannot learn to associate a novel behaviour with it directly. This can only happen when the animal is aware of being uncertain.

I do not know whether there is anything to this idea; but it is interesting. So let me assume for argument’s sake that conditioning of a response to an inner state requires awareness (knowledge) of being in that state. Davidson (1987, 1993) has looked at conditioning of fear responses to inner states of food deprivation in rats. For instance, rats were food deprived for either 23 or only 6 hours, and then received a shock depending on condition after 23 or after 6 hours deprivation. This shock was always at least 6 hours after last food intake to prevent conditioning to recent memory of food. After 24 trials rats differentiated between deprivation levels by their differential rate of freezing in anticipation of the shock. Then to test whether freezing was conditioned to the inner state of food deprivation rather than external stimulus aspects rats were either given an intubation of high-calorie food or a sham intubation. Relative to sham intubation, the high-calorie load increased freezing for rats previously shocked under 6-hour food deprivation and decreased freezing for rats previously shocked under 23-hour food deprivation. So, clearly freezing in anticipation of shock can be conditioned to the inner state of being hungry. If, by hypothesis, conditioning to an inner state requires knowledge of the inner state then these results show that rats can be aware of their state of hunger. If, in contrast, we agree that behaviour can be conditioned directly to an inner state (without the animal knowing that it is in that state) then no reflective abilities can be claimed for rats. But then the findings from the opt-out experiments that training animals to use the o-response when being uncertain do not seem to require any metacognitive ability either.

The basic opt-out paradigm as described in its version used by Smith et al (1997) has undergone great evolution since its conception. So the hope arises that the more refined recent versions might overcome my pernicious problems pair.
Smith et al. (2006) introduced deferred feedback to ‘make it impossible for the uncertainty response [i.e., o-response] to be conditioned by feedback signals, responsive to reinforcement history, or based in low-level associative cues’ (p. 289). The o-response was trained in groups of four trials consisting of a mix of easy and difficult problems. The animal’s response and deserved outcome was recorded but the outcome for each trial was only paid out after each block of four trials: first all the rewards for correct responses were paid out and then the sum of penalty time-outs for wrong responses had to be endured. The o-response counted neither for rewards nor for a time out. Smith et al. (2008) showed that modelling animals’ performance with an associative model assuming a baseline tendency to use the o-response provided a good fit to the data under transparent feedback but not to data gained under deferred feedback. It was concluded that animals must have used their state of uncertainty for learning when to use the o-response. However, this conclusion still leaves open whether the o-response was conditioned to KNOWING that they were uncertain (metacognition) or just to BEING uncertain (no metacognition: Problem of section ‘BEING in a state versus KNOWING that one is in this state’) and whether it was conditioned as a response to being uncertain (a mental state) or to encountering a difficult trial (an external condition).

Washburn et al. (2006) showed that use of the o-response transfers immediately to novel discriminations (see also Kornell et al. 2007; Couchman et al. 2010). Such transfer provides important evidence that the o-response could not have been conditioned to the middle range of the discriminative (density) stimulus dimension (e.g. if semi-dense, take o-response; known as the ‘middle stimulus argument’). Unfortunately transfer to novel dimensions does not help solve my pernicious pair of interpretation problems. The transfer can be explained by both, conditioning to the state of being uncertain (‘BEING in a state vs. KNOWING that one is in this state’) or to difficult trials (‘Internal–external: representing a state of the world versus representing an inner state (representation) caused by the external state’). It is hard to see what a metacognitive account of animals being conditioned to knowing that they are uncertain would add.

Hampton (2001) introduced the memory version of the basic opt-out paradigm. Monkeys were first presented with a sample, and then after a delay with a choice between opting-out (low reward: pellet) or opting-in followed by an array of stimuli containing one identical to the initial sample. Correct choice of the stimulus that matched the sample led to high reward (peanut), error choice to a 15-second timeout.

Again this can be explained by conditioning to having a memory without need for monkeys to be aware of the quality of their memory as the metacognitive interpretation has it. What monkeys learn is that if there was no sample then they should opt out, otherwise they should opt in and subsequently press the icon corresponding to the sample. Since there was always a sample but the monkey’s knowledge of it depended on its memory, the longer the delay between sample and test the more likely that no memory could be retrieved. The important point is that the monkey need not monitor the quality of its memorial process (a metacognitive act) but only monitor what had happened in the past, which is presented through memorial processes.

Washburn et al. (2010) used transcranial magnetic stimulation (TMS) to interfere with a monkey’s early processing of visual images. The monkey’s recognition accuracy declined and its use of the o-response increased. Again this result does not help decide whether the monkey was monitoring the past and used the o-response when—from its point of view—there had been no sample stimulus or whether it was monitoring the quality of its memory and used the o-response when judging that its memory has faded or gone.

Kornell et al. (2007) introduced yet another twist to this paradigm: the retrospective opt-out (confidence judgements) and combined it with the transfer to completely novel tasks. Two rhesus macaques were trained to indicate the large dot in a field of many slightly smaller dots. Difficulty of
the task depended on the difference in size. After one of the dots had been indicated two icons appeared: HR (high risk) and LR (low risk). Choosing the HR-icon resulted in a large reward (+3 tokens) if the large (correct) dot had been indicated beforehand, but resulted in a loss (−3 tokens) if any other (wrong) dot had been indicated. If the LR-icon was chosen a small reward (+1 token) accrued regardless of what had gone before. Monkeys learned to press the HR icon after easy trials and the LR icon after difficult trials where they were more likely to have chosen a wrong dot. The metacognitive interpretation of these data is that monkeys realized when they might have made a mistake and learned to signal their uncertainty or low confidence by choosing the LR icon. However, I can think of at least three alternative explanations that do not bring metacognition into play.

(1) Monkeys’ behaviour in this basic task can be explained by assuming they learn to press the largest dot if they detect one, otherwise they pick any dot. The likelihood that they detect a largest dot is higher for easy trials, where the largest dot is easy to see, than for difficult trials. They also learn that after having chosen the largest dot, provided there was one different enough for them to see, it pays to follow this choice by choosing the HR icon; but if—from their point of view—there was not a largest dot to press (or there was but they chose the wrong dot) they should follow it up by choosing the LR icon. Problems for this explanation arise from the transfer findings. After the dot discrimination task the monkeys were trained on a recognition memory task. After a sequence of about 6 pictures a test slide showing nine pictures simultaneously was presented. Animals had to choose the picture that had been presented before in the sequence of pictures. After mastery of this memory task the familiar LR- and HR-icons were presented after each choice of test picture. Amazingly, right from the beginning (first 4 test sessions) the animals showed a significant tendency to follow correct selection of the old picture with choice of the HR-icon and wrong choices of new test items with the LR-icon. This result would not be possible under my explanation. Since there was no largest dot to press in the memory task they should always opt for the LR icon.

This problem for my explanation can be rectified by assuming that stimuli promising high payoff are imbued with a value (the largest dot is a good thing). And they get conditioned to choosing the HR icon after having chosen a good thing, the LR icon otherwise. Under this description of the stimuli the response options can easily transfer to the memory task. In the memory task they learn that the test picture that was the same as the one that occurred in the series before is a really good thing. When you choose it you get a large reward. So this makes them prefer to press HR after choosing the good item and LR if they could not find a good item among the array. The assumption that animals tend to rely more on the appetitive value of stimuli (good, bad) rather than their physical description (largest dot, same as sample, etc.) is underlined by the standard position on ultra quick processing of the emotional significance of stimuli (LeDoux 1996; Zajonc 1980).

There are two more alternative explanations that keep closer to the metacognitive interpretation. (2) The choice of LR depends on being uncertain but it is the state of uncertainty that makes the animal prefer the LR icon and not a metacognitive process of the animal realizing that it is uncertain. This approach will directly transfer to the memory task. (3) The animal distinguishes between easy and difficult trials, chooses to follow easy trials with HR and difficult trials with LR, and does the same in the first discrimination task and later in the memory task. None of these explanations requires clearly recursive metacognition.

Kiani and Shadlen (2009) taught rhesus monkeys to move their gaze to a target to the right or to the left depending on a left/right movement stimulus at fixation point. One of the target points was presented in the receptive field of neurons in lateral interparietal (LIP) cortex whose activity was recorded. Task difficulty was varied by the clarity of the movement stimulus. On some trials
an 'opt-out target was shown above fixation point about ½ sec after extinction of the motion. Then the fixation cross extinguished telling the monkey to move his eyes to one of the targets. Looking at a target in the direction of motion led to full drink reward, looking at the opposite target led to a time out, and looking at the opt-out target to 80% of drink reward'. As expected, accuracy went up and frequency of looking at the opt-out target went down with clarity of the motion stimulus. What is new is that the LIP neurons not only predicted at which of the presented targets (in direction of motion or opposite) the animal would later look. They also predicted whether the animal would look at the opt-out target before it even knew whether such a target would be available or not. The authors present a model to explain their data and mention that their model makes metacognitive explanations for certainty monitoring unnecessary (Kiani and Shadlen 2009, p. 763). What I would like to know is what the neural response had to be so that a metacognitive explanation would be needed.

First- to third-person perspective transfer: awareness of seeing

There is an interesting methodological approach in theory of mind to showing that animals and infants understand other people’s looking as a mental state of seeing. Its core argument draws on the involvement of metacognition. For instance, experiments by Hare et al. (2000) show that chimpanzees can anticipate a conspecific’s likely action depending on whether the other can look at the bait behind a transparent screen or cannot look at it because it is shielded by an opaque screen. The question of theoretical significance is what this shows about their understanding of the mind. The ‘mentalist’ explanation (e.g. Tomasello, Call, and Hare 2003) assumes that chimpanzees really understand that looking at the object in its hiding place (looking being a purely observable external event) leads to seeing the object and where it is (a mental state with subjective content), which leads to knowing where the object is (a mental state that enables to adapt behaviour to the observed event). The alternative proposal (Povinelli and Vonk 2003; Penn and Povinelli 2007; Perner 2010) refers to ‘behaviour rules’ that inferentially link observable behaviour (looking) with observable behaviour (go for the food) without a mediating chain of inferences involving mental states like seeing and knowing. That is, chimpanzees understand that if a conspecific looks or has looked at the place with food then he will go for it when able to do so.

One longstanding proposal for deciding this issue is to use a method pioneered by Novey (1975) with infants to test for their ability to infer from their first person experience with transparent and opaque goggles that another person can or cannot see when wearing these goggles (Heyes 1998). The central idea behind the proposed investigation is that there could be no behaviour rule relating the personal experience of seeing things when wearing these goggles to the experience of seeing another person wearing these goggles and directing his head towards a target object or event. Melzoff and Brooks (2008) found that 1-year-old infants can use this information and are more prone to follow an adult’s head direction with their gaze when the adult wears a blindfold that they have experienced as being transparent than when wearing one they had experienced as opaque. Teufel et al. (2010) used goggles and reported that by 2½ years children can also verbally indicate through which goggles the other could see and through which he could not. This suggests — so the argument goes — that such young infants must be aware, not just of what they see, but of the fact that they can see, a clearly metacognitive awareness, or else it would be inconceivable how they could possibly make a link between their experience of seeing/not seeing things and the ability of others to see.

14 Smith et al (2008) would not count this study as a serious contender because by rewarding the use of an opt-out response opens the gate for behaviourist alternative explanations.

15 Tomasello and Call (2006) changed their position on this second claim.
seeing with the goggles to what another person can or cannot experience when wearing these
goggles.

This procedure has been unsuccessfully used with chimpanzees. Before someone rushes into
improving the technique for a repeat, a word of caution: these data with infants may not show
what people take them to show. We have known for some time that chimpanzees can distinguish
between transparent and opaque screens (see results from Hare et al. 2000) and they seem to be
particularly sophisticated at understanding what transparency/opaqueness affords (see the inform-
al observations photographically documented in Povinelli 1996). No one has claimed that this
is evidence for metacognition in chimpanzees. Going along with this intuition I conclude that the
ability to distinguish transparent from opaque screens in infants is no evidence for metacognition
either. We need to ask how we determine whether something is transparent or opaque. Presumably
in the same way we determine whether something is red or blue, or square or round, by using our
own personal experience with it: if it looks blue it is blue . . . if it looks transparent it is transparent.
The goggles and blindfolds used in the infant experiments differ from the transparent/opaque
screens used with chimps in that one cannot see from a distance what they are— one has to bring
them close to one’s eyes. As a consequence, when observing another person wearing the goggles
one cannot simultaneously see whether they are transparent or opaque. One can only know from
memory. But apart from this difference the studies with goggles require the same inferences as
those with screens. That is they do not exclude behaviour rules. One has to look through the gog-
gles to determine whether they are translucent or opaque as much as one has to look through the
screen to determine whether it is transparent or opaque. Once one knows this one can anticipate
the other person’s actions and abilities on the basis of behaviour rules: if there is a transparent
object (goggles or screen) between his eyes and the target he will behave adaptively towards the
target, otherwise not.

Interestingly, this further step does not come so easily to 2½-year-old children as shown in the
third experiment by Teufel et al. (2010). Although they could indicate correctly which goggles one
could see through, they showed no sign of understanding that transparency made a difference to
the wearer’s knowledge. While the adult was wearing the glasses the child saw a sticker being
put inside one of two containers. The children requesting help opening the container made as
many pointing gestures to the container regardless of which goggles the adult had been wearing
during hiding. Only when children were given direct experience of the adult being unable to act
sensibly when wearing the opaque glasses did they adjust their requesting behaviour accordingly
(experiment 3).

So there is no mileage in goggles for deciding the issue plaguing theory of mind in infants and,
thus, no mileage for demonstrating metacognition of perception, unless one wants to claim that
the ability to distinguish transparent from opaque screens itself requires metacognition. In that
case the evidence is already in from chimpanzees’ sophistication with transparent objects and no
goggle experiment is needed.

**MiniMeta**

**The project**

My analysis of some of the most impressive ‘metacognition’ studies with animals leaves me with-
out a single clear behavioural test, from which one can infer metacognition in young children
without having to rely on sophisticated language use. This makes me wonder why many people,
me included, have the initially unquestioned impression that each of these paradigms do demonstrate metacognition\textsuperscript{17}—until one engages in a strict cognitive analysis of the phenomena. Of course, one answer to this question may be that our intuition is simply misguided by uncritical application of our folk psychology. The mistake is to think that the behaviour shown in these tasks can only occur for the reasons that we would give when asked to justify or explain our own behaviour in those situations. This tendency is reinforced by our awareness of the studies’ objective to assess metacognition, and so our intuitive understanding of the tasks is already framed in metacognitive terms.

Yet, there may be a more objective fact underlying our intuitions. The distinction between object-level and metalevel may not be as dichotomous as these labels suggest. There may be a more continuous slope leading from the lower to the higher level (see Fig. 6.2). Our tendency to see metacognition in ‘metacognition’ tasks stems from the fact that they require cognitions that are not needed for ordinary object-level cognition but are typical ingredients of metacognition. One could say that they are cognitions that go beyond (one meaning of meta-) ordinary cognitions in the direction of full blown recursive metacognition.

The point of this enterprise is not primarily to explain our intuitions about applying the term ‘meta-’ but to learn more about the nature of metacognitive tasks and what animals and young preverbal children can do in this direction. To safeguard against panmetacognition (seeing metacognition everywhere) I want to adhere to the following three MiniMeta Check Criteria:

1. **Necessity:** is the component cognition that makes behaviour intuitively ‘meta-’ necessary for the behaviour to occur?

   Demonstrations of metacognition in non-verbal creatures are based on showing behaviour under conditions in which the behaviour could allegedly not be shown unless some special cognitive (minimally metacognitive) processes were involved. So we need to look first whether the behaviour could occur with only patently ordinary cognitive processes. Only if the special cognition turns out to be necessary for this behaviour can further checks provide evidence for MiniMetaCognition.

2. **Directionality:** MiniMetacognition is cognition that goes beyond ordinary object-level cognition in the direction of recursive metacognition.

   This criterion should ensure that not just any unusual cognition be classified as MiniMeta. Only cognitions that have some affinity with standard metarepresentational metacognition should qualify. For instance, Carruthers and Ritchie (Chapter 5, this volumes) questioned whether opt out tasks testing for knowledge of uncertainty require metarepresentational understanding of uncertainty and suggested that they may require a certain feeling generated by uncertainty. Although the queasiness caused by indecision may not be proper recursive cognition of feeling queasy about one’s uncertainty, even the first-order state of simply feeling queasy has affinity with queasiness about uncertainty because it was caused by uncertainty.

3. **Exclusivity\textsuperscript{18}:** MiniMetacognition should only be needed for behaviour that is intuitively metacognitive. It should not as well be needed for behaviour that has never been claimed to

\textsuperscript{17} I follow here a certain Principle of Understanding: ‘Never think you’ve understood something unless you’ve also got a good explanation for why others before you kept getting it wrong!’ (Perner 1991, footnote p. 58).

\textsuperscript{18} This could also be dubbed the ‘why more experiments?’ argument. If the analysis turns up that long known findings provide equally good evidence of metacognition, then the question arises: Why those additional metacognition experiments?.
demonstrate metacognition (unless the new discovery leads to a convincing re-evaluation of the implications of the original findings).

For instance, Proust (2007) has argued that regulatory processes of monitoring and control should count as metacognitive. One could counter that established models of even the most simple action control posit that a corollary discharge from the motor command is used to project the intended movement (forward model), which is then compared with somatosensory feedback of the actual movement, and any registered deviations are used to correct the future movement path in advance (Wolpert et al., 1995). Since such monitoring and control is so common it would make just about all behaviour metacognitive (Carruthers and Ritchie Chapter 5, this volume).

In the following subsections I illustrate the MiniMeta approach with but one fully argued example but indicate other intuitively promising venues.

**Implicit awareness of ignorance: information search under partial ignorance.**

In the section 'BEING in a state versus KNOWING that one is in this state', I described the experiment by Call and Carpenter (2001). Chimpanzees looked more often inside a tube to check where the bait was before committing themselves to a definite choice if they only knew that a bait was in one of the tubes but not which (partial knowledge) than when they knew the precise tube (full knowledge). This has been taken as evidence for metacognition. To see whether this interpretation is warranted I contrasted this result to the typical exploratory behaviour shown by an animal that does not know where the food is (no knowledge). One can accommodate this finding easily within regular object-level cognition: content of knowledge determines behaviour. If the animal knows that the bait is in location *x* it will retrieve it from *x*. If it knows it is in location *y* it will retrieve it from *y*. If the animal knows nothing about the bait’s location it will engage in exploratory behaviour, looking around. So far, there is no intuition that metacognition would be involved. Now, if the animal knows that the bait is in one of the tubes, but not in which one, object-level cognition would lead the animal to restrict its exploration to the tubes. So, why does the partial ignorance case raise the spectre of metacognition?

Here is one MiniMeta idea. In the full knowledge and no knowledge case there either is information about the bait’s location or there is no information about it. Whereas in the partial ignorance case the animal has to represent a disjunctive state of affairs: the bait is either in tube 1 or in tube 2. Ordinary cognition in animals uses perceptual input to keep an updated mental model of where things are. A disjunction requires alternative mental models and is an ‘implicit’ way of representing one’s ignorance about the actual location. This would provide an objective feature for our intuition that pointed search under partial knowledge involves metacognition. Now let us see whether this suggestion passes our checks for MiniMeta.

1. Necessity: are alternative models of reality necessary for showing the observed behavioural differences in the knowledge and partial ignorance task? We do not know. The animal might just have a single model that specifies the region of the tubes and, consequently, the animal explores the tubes and not anywhere else. Call and Carpenter (2001) noticed that some individuals used the particularly efficient search of seeing an empty tube and then choosing the other tube without looking inside it first (see also Call, this volume). This speaks for alternative models because if one model is ruled out then only one alternative is left and needs no checking. Whereas, if search is limited to a region then the animal would tend to search until it caught a glimpse of the bait.
The recent evidence by Krachun and Call (2009, see section 'BEING in a state versus KNOWING that one is in this state') that animals place themselves adaptively into the right position for looking inside relevant containers does, however, not help decide the issue. The adaptive self placement requires sophisticated knowledge of how to best explore a container, but it does not require a disjunctive representation 'it could be in one of these containers'. The animal just looks inside every container within the restricted search space.

Nevertheless, although we have no clear evidence for the animals entertaining alternative models, we have a promising line of research to pursue.

2. Directionality: entertaining alternative models is not just any kind of unusual cognition. It clearly points to uncertainty. Hence, although it is not metacognitive in the metarepresentational sense (i.e. the animal knowing that it does not know in which tube the bait is), knowing that the bait could be either here or there is a clear step in that direction (an 'implicit' admission of ignorance).

3. Exclusivity: our MiniMeta evidence consists of animals engaging adaptively in information seeking behaviour because they represent alternative models (implicit knowledge of their partial ignorance). To my limited knowledge of the animal literature I think that there is no evidence that animals can do this in other situations where one would not get the intuitive impression of metacognition being involved.

In conclusion, experiments to show that animals base their information seeking behaviour on alternative models of reality are worth their money. Beck et al. (chapter 11, this volume) use a similar approach to children's understanding of uncertainty.

'Metacognitive' properties

In the section 'Internal–external' I pointed out that many types of mental states are caused by particular types of situations or objects: disgust is caused by disgusting objects, uncertainty is caused by difficult problems, etc. This makes it difficult to decide whether behaviour shown under these conditions is indicative of metacognition about the type of mental state or of plain cognition about the type of situation. There is, though, a strong intuition that these cases differ. For instance, when judging that something is disgusting we would not feel that this requires a metacognitive concern about one's feeling disgusted by the object. In contrast, when a task is judged difficult because one does not know the answer, one does have the feeling that this judgement involves some metacognitive awareness of one's lack of knowledge.19

Another property of objects that smacks of 'metacognition' is transparency (see section 'First- to third-person perspective transfer: awareness of seeing'). Yes, I can classify objects as transparent or opaque, but to judge something as transparent don't I need some metacognitive awareness of the fact that I can see through it? That is the intuition, but we do not know what really goes into such a judgement. Finding the critical difference between judging something as disgusting and judging something as difficult or transparent would be an important advance for the MiniMeta project. All I can contribute here is to point out the difficulties.

Notions like 'subjectivity' and 'involvement of mental states' (as my Anonymous Reviewer suggested) do not go far enough. Take for example size constancy in visual perception. The task is to judge whether two objects presented at different distances from the observer are the same or different in size. Our visual system can be fooled but by and large can do this quite accurately. When we try to explicate how it can do this we get quickly entrenched in very metacognitively sounding

19 An intuition I share with my Anonymous Reviewer.
arguments: The visual system knows that a more distant object projects a smaller retinal image than the same size closer object. The system takes this into account by judging the size of objects in relation to their estimated distance. Does this make size a metacognitive property because its judgement relies on knowledge about perception? The strong intuition is that it is not. But then, what makes properties like difficulty and transparency minimally metacognitive?

Conclusion

I have set out in search of methods for investigating metacognition in children that do not rely on language. I formulated two rather pernicious problems for any attempt to infer metacognition from behavioural data. Unfortunately, search of the rich repertory of methods developed in the comparative animal literature showed that no single method—of those that I thought would be most promising—could overcome these two problems.

Other interpretation problems with these tasks also were discussed. So, the comparative literature has increasingly taken to arguing with cross species consistency (Couchman et al. 2012). Species which show information seeking also learn the opt-out response, like chimpanzees and rhesus macaques, while others fail to show these behaviours, such as pigeons, rats, and capuchin monkeys. The split between these groups of species is not completely clear cut. For instance, capuchins can learn to ask for information which pigeons cannot (Beran and Smith 2011; also see 'Information seeking' section). Nevertheless, this separation into two groups is seen as confirmation of the view that one group is capable of metacognition and the other is not, because the intended common denominator of the different tasks is supposed to be metacognition. Extending this approach to developmental investigations with children is costly. It demands employment of several different tasks and hope for a cross age consistency. And then there is still the possibility that the common feature of the so-called metacognition tasks that separates groups of species is not metacognition but some other kind of cognitive complexity.

Another promising indirect way of strengthening the evidence for metacognition was recently reported by David Smith and his colleagues (Smith 2011). They used their pixel density opt-out task together with a concurrent executive task. Their reasoning was that metacognition makes executive demands and, therefore, a concurrent executive task should interfere with the opt-out response but not with the primary choices (sparse/dense). Indeed with human participants a concurrent variation of the number Stroop task interfered specifically with the opt-out response to the degree that it eliminated the opt-out response altogether. They now also have data from monkeys. A concurrent delayed matching-to-sample task interferes to some degree with the opt-out response but not with the primary response options. These are impressive confirmations of a risky prediction from the metacognition stance.

Yet, it still leaves us with the not implausible possibility that the hallmark of the opt-out response is an executive demand and not necessarily metacognition. As a matter of wild speculation this executive command could be the ability to distance oneself from rash responding. That makes animals which have this ability more likely to gather more information before responding, and not take blind risks with the potentially best paying response. Rhesus macaques and great apes have this ability, and perhaps capuchins, pigeons, and rats lack it.

Although the methods used with animals do not provide unambiguous evidence for metacognition under cognitivist scrutiny, intuitively these tasks do feel metacognitive. One resolution of this contradiction would be if these tasks require abilities that lie between object-level and meta-level cognitions, i.e. they are ‘minimally metacognitive’. This would explain the intuition and also why the cognitivist analysis does not admit them as metacognitive in the sense of recursive cognition. In my MiniMeta programme I gave some guidelines of how to identify whether a task
is minimeta. I only managed to outline the strategy and, thus, only presented information search under partial knowledge as a fully argued example. I came to consider the ability to entertain alternative mental models as an implicit way of acknowledging partial ignorance. I also looked at ‘metacognitive’ properties, properties that can only be detected with the use of metacognition. I was unable to pursue the potential of this line lacking the relevant insights how to proceed. There are very likely many other minimeta lines to explore. A particularly interesting case of cognition between object- and metalevel may be the case of conditional reasoning as elaborated by Johannes Leitgeb in Chapter 14.

So, in the end I cannot be of much help to my colleagues contributing to the next section. My search for the most telling paradigm across the rich repertory of methods in the comparative animal literature did not come up with a clearly satisfactory result. The best suggestion I can pass on is to look for developmental consistency across different ‘metacognition’ paradigms. Unfortunately this is a costly approach. Beyond that I hope that my pair of pernicious problems will provide a good measurement bar against which to assess the developmental findings.

Acknowledgements

I thank my friends and colleagues who have given me critical pieces of advice and criticism on an earlier version of this chapter: Michael Beran, Johannes Brandl, Josep Call, Tony Dickinson, Frank Esken, Ulrike Klossek, Sara Shettleworth, and last but not least The Anonymous Reviewer. Financial support came from the European Science Foundation and the Austrian Science Fund (FWF Project I93-G15) ‘Metacognition of Perspective Differences’.

References


Novey, M. S. (1975). The development of knowledge of other’s ability to see. Unpublished Doctoral dissertation, Department of Psychology and Social Relations, Harvard University.


SECTION I: METACOGNITION IN NON-HUMAN ANIMALS


