Neuroscience and Biobehavioral Reviews 71 (2016) 167–175

Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

journal homepage: www.elsevier.com/locate/neubiorev

Review article

Neural correlates of metacognition: A critical perspective on current tasks

Andrea Insabato[a,*,1], Mario Pannunzi[a,1], Gustavo Deco[b]

[a] Universitat Pompeu Fabra, Theoretical and Computational Neuroscience, Center for Brain and Cognition, Roc Boronat, 138, 08018 Barcelona, Spain
[b] Institució Catalana de Recerca i Estudis Avançats (ICREA), Universitat Pompeu Fabra, Theoretical and Computational Neuroscience, Center for Brain and Cognition, Roc Boronat, 138, 08018 Barcelona, Spain

ARTICLE INFO

Article history:
Received 13 May 2016
Received in revised form 17 August 2016
Accepted 26 August 2016
Available online 30 August 2016

Keywords:
Metacognition
Decision-making
Confidence
Monitoring
Animal cognition

ABSTRACT

Humans have a remarkable ability to reflect upon their behavior and mental processes, a capacity known as metacognition. Recent neurophysiological experiments have attempted to elucidate the neural correlates of metacognition in other species. Despite this increased attention, there is still no operational definition of metacognition and the ability of behavioral tasks to reflect metacognition is the subject of debate. The most widely used task for studying metacognition in animals, the uncertain-option task, has been criticized because it can be solved by simple associative mechanisms. Here we propose a broad perspective that generalizes those critiques to another task, post-decision wagering. Moreover, we extend this critical view to account for recent neurophysiological evidence. We argue these tasks are simple enough that any animal could solve them using very simple mechanisms such as sensory-motor associations. In this case, it is impossible to know whether all animals are metacognitive, or if the tasks are simply not appropriate. Therefore, we suggest using better defined concepts until a suitable task for metacognition is available.

Contents

* Corresponding author.
E-mail address: andrea.insabato@upf.edu (A. Insabato).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.neubiorev.2016.08.030
0149-7634/

1. Introduction: An unsolved question about metacognition

Since Aristotle's *De Anima*, metacognition has interested philosophers and has recently become a popular topic in psychology and in neuroscience (see Section 2 and Fig. 1). Despite this increased attention, the definition of metacognition is a matter of debate (see Section 2) and there is still no operational definition.

While metacognition has long been thought to distinguish humans from other animals, the current consensus is that other animals might also have metacognition. Indeed research on animal models is expected to bring significant advances in understanding the neural correlates of metacognition. Nevertheless, it is still unclear which species have metacognition (Inman and Shettleworth, 1999; Sutton and Shettleworth, 2008; Kepecs et al., 2008; Foote and Crystal, 2007; Smith et al., 1995; Call and Carpenter, 2001; Komura et al., 2013; Kornell et al., 2007; Perry and Barron, 2013).

A central theme of this debate is the ability of behavioral tasks to selectively reflect metacognition and avoid confounding by simpler cognitive functions (see Section 2). Hence, the major challenge is to find an adequate test of metacognition for non-linguistic animals.[2]

The most widely used task for studying metacognition in animals, the uncertain-option task (Smith et al., 1997), has been criticized on the basis of modelling studies (Smith et al., 2008; Jozefowiez et al., 2009; Le Pelley, 2012). Here we argue that these critiques can be extended to new neurophysiological evidence made available in recent years. Moreover we propose a similar critical perspective of behavioral and neurophysiological evidence of metacognition in another task, post-decision wagering (Son and Kornell, 2005). We note that these tasks can be used both as tests for metacognitive abilities and as measures of a given metacognitive variable, e.g. uncertainty or confidence. The critiques of current tasks for metacognition concern those tasks as tests for metacognition (however a poor test is also a poor measure).

1.1. A critical perspective on animal metacognition studies

To introduce our perspective, consider a fictitious scenario: an illness with elusive and highly variable symptoms, for which no diagnostic procedure exists yet. One day a doctor develops a new test for this disease. She wants to estimate the specificity and selectivity of the test, so she begins testing many supposedly ill and supposedly healthy subjects. In the end it turns out that all results are positive. With this finding the doctor is unable to tell whether the test is 100% specific and everybody is ill or whether some positive results are false and there are both healthy and ill people. We believe that this situation is similar to that of current tasks for metacognition. As for the fictitious test for the illness, we do not know if these tasks are testing metacognition or some related phenomenon (e.g. estimation of difficulty). We could consider that a good behavioral task that truly reflects metacognition could distinguish animals that have metacognition from those that do not.

Currently it is not clear which species would pass this filter, and this discussion is beyond the scope of this article (Le Pelley, 2012; Perner, 2012). However we can be sure that a task any animal can solve would be an all-pass filter and, as such, not useful. This case is similar to that of the fictitious doctor: We would be unable to tell whether all animals have metacognition or whether the task is giving some false positives.

For the remainder of this article we will analyze results related to decision confidence, i.e. the belief that a choice is correct, as this is considered a key manifestation of metacognition (Carruthers, 2008; Metcalfe, 2008; Smith et al., 2003). While various tests and procedures have been developed over the past 25 years (see Section 3), we will focus on the two tasks that have been most widely used to explore the neural correlates of metacognition (Kepecs et al., 2008; Kiani and Shadlen, 2009; Komura et al., 2013; Fetsch et al., 2014; Lak et al., 2014): the post-decision wagering (Son and Kornell, 2005) and the uncertain option tasks (Smith et al., 1997).

We now briefly summarize our perspective, to highlight our assumptions and ideas:

1. All animals are able to make perceptual decisions.
2. If an animal is able to make perceptual decisions, then it has at least a simple decision mechanism.
3. All animals have at least a simple decision mechanism.
4. Post-decision wagering and uncertain option tasks can be solved using simple decision mechanisms.
5. All animals are able to solve post-decision wagering and uncertain option tasks.

Statement 1 does not need to be strictly proven for every animal. Instead, using an evolutionary argument, we may assume it to be true for all animals more complex than e.g. *C. Elegans*. Statement 3 comes from 1 and 2, and thus applies to all animals referred to in statement 1. Statement 4 is the one we will advocate for throughout this article. This statement is based on the notion of a "simple mechanism", i.e. mechanisms that use only stimulus representations, and decisions based on those representations.

Note that the problem with statement 5 does not come from the assumption that only some animals have metacognition (although it may be reasonable to assume that many simple animals have no metacognition, such as nematodes, fruit flies, or some fish). As for the fictitious diagnostic test, statement 5 is problematic because it makes it impossible to know if post-decision wagering and uncertain option are selectively testing for metacognition or perhaps some related phenomenon.

Given the lack of a formal definition, it would be useful to have at least a minimal requirement for metacognition, a necessary even if not sufficient condition for metacognition. We may propose monitoring as such a minimal requirement, given that all accounts of metacognition are based on some form of monitoring (Nelson and Narens, 1990). Such necessary condition could be useful because those processes that do not fulfill the requirement cannot be considered metacognitive. This reasoning may help the reader to follow our discussion, since the reviewed models do not implement monitoring, but our general argument is independent of the concept of monitoring (see Section 6 for a discussion of the concept of monitoring).

In the following two sections we give a partial overview of significant studies on animal metacognition during the past 25 years, highlighting global trends that we believe to have arisen and a concise summary the tasks used to study confidence in animals.

2. A quarter century of animal metacognition studies

In this section, we briefly present the most important results on animal metacognition. While we are aware of the complexity of scientific opinion on metacognition, we wish to guide the reader through the last 25 years of research on animal metacognition. This is a prospective, biased and non-exhaustive view of the literature, in the sense that some studies are simplified, and deserve a more extensive description that is out of scope here, while other studies have been excluded from our report entirely. Evidence on animal metacognition during the last quarter century can be seen globally as a path going at times towards the discovery of animal

[2] Note that this need can also be applied to pre-verbal human infants: A reliable test for metacognition could be used to study the emergence of metacognitive abilities during development.