Acta Anal
https://doi.org/10.1007/s12136-018-0340-0

Getting It Together: Psychological Unity and Deflationary Accounts of Animal Metacognition

Gary Comstock[1] · **William A. Bauer**[2]

Received: 8 September 2017 / Accepted: 3 January 2018

Abstract Experimenters claim some nonhuman mammals have metacognition. If correct, the results indicate some animal minds are more complex than ordinarily presumed. However, some philosophers argue for a deflationary reading of metacognition experiments, suggesting that the results can be explained in first-order terms. We agree with the deflationary interpretation of the data but we argue that the metacognition research forces the need to recognize a heretofore underappreciated feature in the theory of animal minds, which we call *Unity*. The disparate mental states of an animal must be unified if deflationary accounts of metacognition are to hold and untoward implications avoided. Furthermore, once *Unity* is acknowledged, the deflationary interpretation of the experiments reveals an elevated moral standing for the nonhumans in question.

Keywords Metacognition · Psychological unity · Animal minds · Brainets · Moral standing of animals · Uncertainty test

1 Introduction: Metacognition and Consciousness in Animals

I have metacognition, *the power to think about my thoughts*. If you are asking yourself whether you believe this claim, you are metacognizing because you are reflecting upon your beliefs—your first-order thoughts—using second-order thoughts. Metacognition provides humans with a measure of monitoring and control not available in non-

✉ William A. Bauer
wandbauer@gmail.com; wabauer@ncsu.edu

Gary Comstock
gcomstock@ncsu.edu

[1] Department of Philosophy and Religious Studies, North Carolina State University, 458 Withers Hall, Campus Box 8103, Raleigh, NC 27695, USA

[2] Department of Philosophy and Religious Studies, North Carolina State University, 434A Withers Hall, Campus Box 8103, Raleigh, NC 27695, USA

Published online: 20 January 2018

metacognitive beings and is an important factor in our moral standing insofar as it indicates executive control processes overseeing subconscious processes, thereby revealing a sophisticated, complex mind. Should some nonhuman animal (hereafter, simply animal or animals) have metacognition, then that fact, were it a fact, ought to bring with it some high level of moral standing for the animal, too.

We argue that the experiments into animal metacognition have not shown conclusively that animals metacognize, but they have revealed an important logical constraint that those writing about the theory of animal minds have largely ignored: the necessity that any account of an individual animal's beliefs, desires, and emotions must recognize the unity of these diverse mental states. Any theory of animal minds that lacks the principle we call *Unity* is not only deficient philosophically. It is also vulnerable to false impressions of the animals' importance; if one thinks animal minds are simple and lack deep moral significance, one might further think animals should not "be allowed to get in the way of any morally-serious objective," as one philosopher once concluded (Carruthers 1989: p. 514). Call this latter view of the moral insignificance of non-metacognizing animal minds neo-Cartesianism.

Holding a very strict kind of higher-order theory of consciousness, neo-Cartesians would argue that whereas animals have first-order, world-directed thoughts, these thoughts are not conscious because they are not available for access by second-order, self-directed thoughts. If one holds that being accessible for recall is what makes a thought a conscious thought, then an animal's inability to metacognize may be evidence that it may not feel like anything *at all* to be that animal. In a thought experiment now well-rehearsed in the animal minds community, Carruthers (1989: pp. 505–506) asks his readers to imagine how it feels when they are speeding down a highway, their conscious attention wholly abstracted from what they are doing. Suppose they are daydreaming about their next vacation and completely unaware of where they are or how fast they are going.[1] Suddenly, they "come to," regain their senses and see to their horror a truck immediately in front of them. They slam on their brakes, narrowly averting a collision. What was the driving experience like for them in those minutes when they were on autopilot? It was like nothing at all. They have no memory of landmarks they passed, glances into their rear view mirror, or any other behaviors in which they engaged. Such experiences are nonconscious experiences and, neo-Cartesians argue, all animal experiences are like that, thus showing why Nagel (1974) is wrong to identify phenomenological subjectivity with experience *simpliciter*. Animals surely have nonconscious experiences and can control their behaviors and yet doing so does not (ever) feel like anything for them. They have experiences but not phenomenological subjectivity. (These supposedly "nonconscious experiences" might, contrary to the assumptions in this paragraph, in fact represent some kind of non-phenomenological consciousness. That is to say, the driving example raises questions concerning whether there are types of consciousness, questions we will address shortly.)

[1] Armstrong (1993: p. 93) also discusses this type of case. Discussing a case that parallels the driving case, Tye (2003: p. 2) imagines a distracted philosopher walking home thinking about her latest theory. She later realizes that she was not aware of any of her perceptions on the walk home. She sees (in some sense) the sidewalk and the trees (otherwise she would trip and bump into things), but lacks what Tye (2003: p. 5) calls "introspective consciousness" which seems to require a metacognitive ability. The distracted philosopher and the daydreaming driver are not introspectively aware of their perceptions (Tye 2003: p. 5). Yet they are conscious in some sense, as we discuss below.

We think the neo-Cartesian's high standard for consciousness is probably false, yet framing the issue like this is dialectically useful. For if comparative psychologists have shown that some animals are metacognizing, then those animals *can* bring "lost" moments to mind and consciously analyze their beliefs and desires. That is, evidence of metacognition would almost surely be evidence of phenomenological consciousness. Indeed, over the last 20 years, evidence has appeared that seems to show that some animals—notably some monkeys (rhesus macaques, not capuchins), dolphins, rats, and baboons—do metacognize (Smith et al. 1997; Washburn, Smith, and Shields 2006; Hampton 2009; Kornell, Son, and Terrace 2007; Hampton 2001; for dolphins: Smith et al. 1995; for rats: Foote and Crystal 2007; for baboons: Malassis, Gheusi, and Fagot 2015). They can stop in their tracks, reflect on what they have seen and done, and even decide how much weight to give to their confidence that they do (or do not) know the answer to a question. If some animals are capable of pausing to survey their memories and to assess the state of items in their knowledge inventory, then we have powerful evidence against the idea that all animal experience is nonconscious and, moreover, against the idea that animal experience is less than morally significant. These broader issues and their moral significance have cultivated our interest in metacognition.

However, thus far, we have overlooked some complexities of consciousness, which sorting out will help to make our intentions and later analysis clear.[2] First, while metacognition may very well be sufficient for consciousness (phenomenal consciousness or otherwise), it is not obvious at all that metacognition is necessary for consciousness. So it seems that a creature could have some kind of consciousness without having metacognition. Intuitively, there are various types or aspects of consciousness, some of which may be possessed without the others. Tye (2003: pp. 5–11), for example, introduces four types of consciousness: (1) introspective consciousness (consciousness of one's percepts, when you see things and are aware of seeing them; this is metacognitive), (2) discriminatory consciousness (ability to recognize perceived items and distinguish them from others), (3) responsive consciousness (processing and responding to information from the world; this can come in degrees), and (4) phenomenal consciousness (the subjective, qualitative state of awareness; what it is like). These types of consciousness may not always be distinct (e.g., having discriminatory consciousness may necessitate having responsive consciousness).[3]

Let us correlate Tye's distinctions with a broader distinction between access consciousness (having access to information for use in guiding one's reasoning and behavior) and phenomenal consciousness (same as (4) above) due to Block (1995). For Block (1995: p. 233), these can come apart (e.g., philosophical zombies and blindsight cases) though usually do not. It seems that discriminatory and responsive consciousness need not (but can) have a phenomenal component, so are essentially types of access consciousness; this is what appears to be going on with the driver from above, as well as the distracted philosopher (see footnote 1). In these cases, information is available to guide behavior; the subject discriminates objects and responds appropriately without deliberately attending to these perceptions or metacognizing.

[2] Thanks to an anonymous reviewer for suggesting that we discuss in greater depth the notion of consciousness.

[3] We think it is better to categorize consciousness into types, as with Tye (2003), rather than levels. Despite the utility of the concept of levels of consciousness in cognitive science, it faces conceptual difficulties and problems to do with properly ordering different types of global states (see Bayne et al. 2016).

Given these points, we maintain that animals of the kind used in the metacognition experiments, and perhaps, many other kinds of animals are at least conscious in this way: under normal, healthy conditions, they have discriminatory consciousness and responsive consciousness; more than likely, they have phenomenal consciousness, but not introspective consciousness. Since they can respond to external world stimuli, and discriminate among stimuli (involving a kind of first-order judgment), they certainly have access consciousness.

Returning to the question of metacognition, we maintain that monkeys and other animals have not yet been shown to have metacognition or introspective consciousness. Like human infants, animals do not "consciously think things to themselves" (Carruthers 1992: p. 184). We maintain that a deflationary, anti-metacognitive explanation of the results of the animal experiments is essentially correct if couched in terms of first-order beliefs and desires, as argued, for example, by Carruthers (2008) and as amended to include affective states (Carruthers and Ritchie 2013). We focus on Carruthers' work because it is the most complete and powerful interpretation of the animal metacognition data that we know of. We defend his schema while arguing that it is incomplete because it ignores the fact that the subject animals' beliefs, desires, and emotions must be psychologically unified in order to generate the behaviors observed in the experiments. By ignoring the principle we develop, *Unity*, first-order accounts are subject to two problems: they are unfalsifiable and metaphysically promiscuous.

To entice the reader, here is a snapshot of our metaphysical worry. Without *Unity*, first-order accounts of animal cognitive behavior are free to posit minds where none exist. Consider technological arrangements involving what Miguel Nicolelis calls a "Brainet." Here, three or more animal brains are interfaced and tasked collectively to solve simple problems (Ramakrishnan et al. 2015). First-order accounts of technological Brainets involving three animals admit, curiously and implausibly, explanatory appeals to a fourth, brainless mind, a result we find metaphysically untenable. While we sign on to the first-order explanation, we show how the problem of "mind-creation" and other unwelcome possibilities can be avoided. We do so by focusing on the fact that the cognitive, conative, and emotional states of the healthy, typically developing monkeys used in the experiments are integrated. By drawing attention to this psychological fact, we elude critical problems that otherwise threaten to subvert first-order explanations of purported animal metacognition.

Here is our roadmap for the rest of the paper. After further explaining what metacognition is, Section 2 describes the empirical evidence for metacognition in animals and the first-order interpretation of it. Section 3 presents our main argument for a necessary qualification—concerning the unity of psychological states—to first-order, deflationary accounts of metacognitive behavior. Section 4 offers a refined statement of the main principle, *Unity*, and contrasts it with other notions of unity. Section 5 explains *Unity*'s significance and demonstrates that untoward metaphysical and neuroscientific implications await those who ignore it. Lastly, Section 6, after briefly recapping the main arguments of the paper, explains why the principle of *Unity* plausibly elevates the moral standing of monkeys and similarly sophisticated animals.

2 The Uncertainty Test and Metacognition in Humans and Animals

How do we humans know we are metacognizing? One way is the feeling that we have when thinking about our thoughts. When thinking about one's beliefs, especially a

belief critical to achieve a desire, one typically hesitates and deliberates, searching to verify the belief's reliability. While hesitating, one is metacognizing.

We assume that upon observing behaviors (verbal, bodily) in others that are associated with metacognition in our own case, we have some reason to infer that the others are also metacognizing. Should we make the same inference in the case of animals? If an animal challenged with the same sorts of cognitive puzzles that cause metacognizing in humans responds with the same sorts of behaviors as the behaviors with which humans respond to those puzzles, it seems the burden of proof is on those who would deny that the animals are metacognizing. The problem is that we cannot confirm the inference as we would do with children test subjects, for instance, by asking them, for animals lack the linguistic capacity to understand the question. However, comparative psychologists have devised various empirical tests to detect metacognition, including the so-called uncertainty test.[4] Below, we describe one version of this test and the results for both humans and animals.

The uncertainty test is meant to gauge a person's ability to tell when they know the correct answer to a question and when they do not. Subjects are given a range of trials in which they must discriminate between computer screens packed with illuminated pixels and screens with very few pixels. The initial trials fall into two categories, sparse and dense, but eventually screens are added that fall somewhere in the middle. Subjects predictably have a difficult time deciding whether to push the "D" (dense) or "S" (sparse) key in response to these displays. As subjects receive a reward for each correct answer they give and a timeout for each incorrect answer, they are motivated not to answer incorrectly. Therefore, when given the chance to decline to hazard a response by pressing the "U"[5] key—an action that immediately initiates a new trial with a guaranteed, modest, reward—subjects may have learned to express their belief that they do not know the right answer by using the third, "don't know" option. Subjects become more reluctant to wager—and more likely to press "U"—as the number of pixels on the screens comes closer and closer to the median point between dense and sparse (henceforth, "ambiguous" screens). They have effectively used metacognition not only to recognize difficult trials but to avoid risking responses that to them are likely to result in less than optimal consequences (so argue those researchers who believe the test detects meta-representational self-reflective thought). These results are shown in Fig. 1.

Researchers believe (and we have no reason to object) that individuals not capable of employing "U" appropriately probably lack metacognition. But if one knows that one's informational state does not include the item required by the question, one may deselect actions with a low percentage of being rewarded handsomely and favor instead the response with a high percentage of being rewarded modestly (that is, the "U" key). Only metacognizers will be able to use the "U" key appropriately because only they understand what they know and do not know (argue those who believe the experiments demonstrate metacognition).

It is an open question whether nonhuman animals have metacognition. Various experiments seem to indicate that some species do have metacognition, but the results

[4] Another test, for instance, is the so-called "false belief" test, constructed to ascertain when a child first learns to understand that the child's beliefs are her own and may be different from others' beliefs. We will focus exclusively on the uncertainty test.

[5] In the experiments, the relevant key is marked "?" but we change the designation here only to aid our reader in interpreting Figs. 1 and 2.

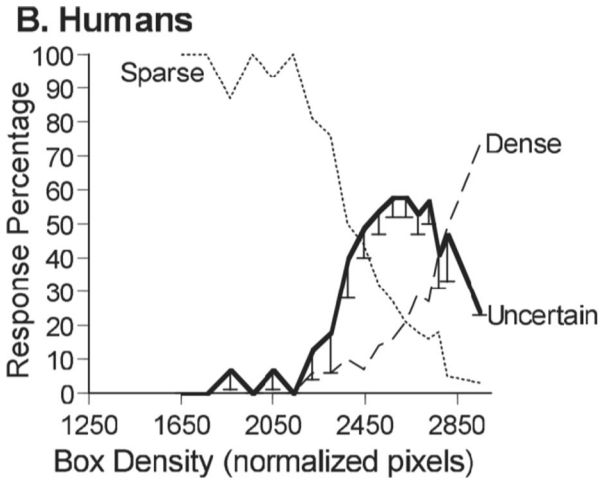

Fig. 1 "The performance of seven humans in the dense-sparse task. The dense response was correct for boxes with exactly 2,950 pixels – these trials are represented by the rightmost data point for each curve. All other boxes deserved the sparse response" (Smith, Shields, and Washburn 2003: p. 322). [We derive Fig. 1 from the upper rightmost panel, panel "B. Humans," in Fig. 3 of Smith, Shields, and Washburn 2003, *Behavioral and Brain Sciences*, 2003: p. 322. Copyright 2003 by Cambridge University Press. Reprinted with the permission of Cambridge University Press] (For further discussion, see Beran et al. (2010))

are controversial. Peter Carruthers (2008) argues, in a paper cited by philosophers and psychologists alike,[6] that the evidence can be plausibly interpreted entirely in non-metacognitive, first-order terms. Let us look at the data in one of these experiments, an experiment in which rhesus monkeys seem to exhibit second-order thoughts. In our presentation of these experimental results and parts of the subsequent discussion, we focus on monkeys for illustrative purposes. However, our core philosophical and normative claims should apply to dolphins and other mammals that have performed similarly on metacognition experiments.

For two decades, beginning with Smith et al. (1995), researchers have been experimenting with animals to see whether any nonhumans are capable of metacognition. Figure 2 shows the results of the dense-sparse test (Smith, Shields, and Washburn 2003) given to a rhesus monkey.

Observe that the monkey's responses (Fig. 2) track the humans' responses closely (Fig. 1). When the monkey is shown a familiar screen, it gets the right answers, as do humans. When the monkey is shown a more difficult screen, it is less likely to gamble, again mirroring the humans' behavior. Is the animal's reluctance to wager an indication that it is surveying its state of knowledge and deciding that it does not know the right answer? Is it uncertain, wanting to seek more information?

Many think so, and they include psychologists (Couchman et al. 2010), animal behaviorists (Rosati and Santos 2016), and philosophers (Gennaro 2009; DeGrazia 2009; Proust 2009, 2010). For example, Gennaro (2009: pp. 186–193) argues that many animals (particularly primates, but some others too) have the concepts required for "I-

[6] References to Carruthers (2008) made by philosophers include Jacobson (2010) and Proust (2009, 2010), and references made by psychologists include Beran and Smith (2011), Couchman et al. (2009), and Smith (2009).

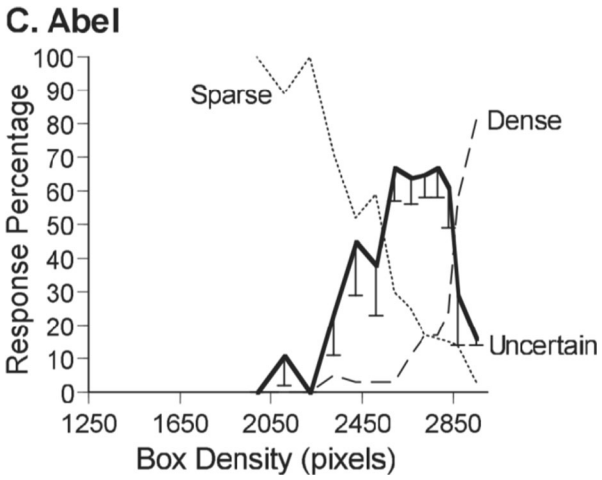

Fig. 2 "The performance of monkey Abel in the dense-sparse discrimination depicted in the same way [as the performance of the seven humans is depicted in Fig. 1]" (Smith, Shields, and Washburn 2003: p. 322). [Reprinted from Fig. 3 of Smith, Shields, and Washburn, *Behavioral and Brain Sciences*, 2003: p. 322. Copyright 2003 by Cambridge University Press. Reprinted with the permission of Cambridge University Press]

thoughts" (higher-order thoughts) on the empirical grounds of uncertainty tests like the one described above. Some, however, think not. Those we call "deflationists" argue—properly, we believe—that one can explain the animal responses, in the above and related experiments, in simpler, because first-order and associational, terms (Le Pelley 2012; Hampton 2009). The uncertainty-monitoring behaviors, we note, occur only after a period of extensive training during which time the monkey probably comes to associate use of the opt-out response with aversive qualities (for a behavioral economic model of this interpretation, see Jozefowiez, Staddon, and Cerutti 2009). In making their cases, both animal metacognitivists and deflationists typically assume that both folk psychology[7] and Morgan's canon[8] hold, and we do not bring these assumptions into question.

[7] Folk psychology consists of the pre-theoretical assumptions people make about their own and others' minds (Goldman 1993). Scientific progress is possible using folk psychology. The folk understand their own uncertainty in terms of conflicts between and among beliefs and desires. The subjects in the experiments, for example, desire to answer all questions in the way that brings reward but sometimes they do not know the right answer. When confronted with an ambiguous figure, the subject is unable to react quickly because of a paralyzing mismatch between their beliefs and desires. In folk psychology, therefore, the typical explanation of uncertainty is to say that the subject does not know on which belief they ought to act.

[8] Morgan's canon is a methodological principle used to guide the study of comparative psychology: "In no case may we interpret an action as the outcome of the exercise of a higher psychical faculty, if it can be interpreted as the outcome of the exercise of one which stands lower in the psychological scale" (Morgan 1894: p. 53). This requires that "the most general cognitive mechanism" (Karin-D'Arcy 2005: p. 182)—presumably not metacognition in the experiments discussed in this paper—be used to explain animal behavior. However, in a revised statement of the principle, Morgan cautions that the canon should not prohibit one from attributing more complex psychological processes *if* independent evidence suggests animals do undergo the more advanced process in question (for discussion, see Karin-D'Arcy 2005: p. 182). Grounding his view in Morgan's canon, Carruthers (2008) holds that attributing more complex psychological processes is not necessary to explain animal behavior in the uncertainty tests since his first-order explanation is sufficient. We agree, provided the qualifications we advance in this paper. For cautions about the proper interpretation of Morgan's Canon, see Sober (2009), Fitzpatrick (2008), Andrews and Huss (2014), and Andrews (2012).

Carruthers (2008) does *not* hold that animals' uncertainty behaviors can be explained merely in terms of electrical inputs and outputs, stimuli and responses, or even associative learning mechanisms. Holding that the animals have *simple* minds—not *no* minds—he grants the appropriateness of employing intentional concepts (beliefs and desires) to explain the behaviors. He goes further, noting that his 2008 explanation is deficient because it neglects the role of emotion (Carruthers and Ritchie 2013). Still, Carruthers' theory is incomplete, even when the cognitive and conative account (Carruthers 2008) is supplemented with the animals' emotional states (Carruthers and Ritchie 2013). To make good on the claim that the animals' behaviors can be parsed in first-order terms which appeal "only to states and processes that are world-directed rather than self-directed" (Carruthers 2008: p. 59), the theory must include reference to a principle we call *Unity*.

We agree with Carruthers that when the animals hesitate before betting, when they pause and seemingly try to figure out whether they know if an ambiguous figure is dense or sparse, their responses do not require second-order, meta-representational terms. More generally, we think that in comparing two psychological accounts, A and B, of some target phenomenon T, if account A is simpler than B in relevant ways (e.g., avoiding appeal to second-order terms), yet retains equivalent explanatory and predictive power for T, then A should be favored (see footnote 7). So, the explanation of animal behavior in the metacognition experiments need not, and should not, appeal to propositional attitudes indexed to self-reflective subjects (e.g., "I don't believe that I know whether this pattern is sparse or dense and I desire more information before pressing S or D"). The explanation of animals' behaviors in the uncertainty experiments does not require metacognitive concepts. So Carruthers argues, and so far we agree. However, something critical is still missing.

3 *Unity*: an Amendment to Carruthers' First-Order Schema

We contend that any first-order, deflationary explanation of animal uncertainty behavior is deficient if it neglects the unity of the individual's beliefs, desires, and emotions. That is, the relevant mental states (in any kind of mind, be it animal, human, or artificial) must be properly integrated in order to generate the various behaviors—pushing the question mark, for instance—as seen in uncertainty experiments. Mental states must be unified both synchronically and diachronically, across relevant sets of neurons, in order to generate the relevant behavior.

On one hand, if the neural correlates of two beliefs are not synchronized, that is, not part of a unified neural set existing in active, interconnected nets at specific locations stretched across the brain, then any given belief-instantiated-in-a-neural net cannot strengthen (or weaken) another (similarly neurally instantiated) belief with which it is supposed to cooperate (or compete), nor may it play a role in explaining the behavior it is supposed to cause (or inhibit). On the other hand, if two beliefs are not unified over time, then they are, again, not bound together in the necessary way. If the relationship between the concept of "one" and the concept of "two" is not established—if, for example, they belong to unrelated sets—then performing as simple an operation as counting from one to two is impossible.[9] Humans have the capacity to *attribute* the

[9] Kant (1998: p. A103) discusses the need for the unity of consciousness in the process of counting.

unity of beliefs and desires to themselves or others, but that is not what we have in mind here; attribution capacity may involve metacognition, and we are not claiming that. What we are claiming is that, for particular kinds of behavior (e.g., uncertainty responses), the relevant beliefs and desires must in fact be unified.[10]

Cognition of the sort in the uncertainty tests needs unity because a judgment is made by the individual animal, occurring in response to a tension involving competing interests. What we have in mind is the following condition, implicitly assumed in folk psychological attributions:

> Unity: A set of mental states is unified just in case all of the relevant beliefs and desires are related to each other *in the right way*.

The "right way" will be clarified, upon further discussion, by a more precise statement of *Unity*. For now, we note that the importance of *Unity* to understanding mind and cognition goes back at least to Kant's discussion of the transcendental unity of apperception in *Critique of Pure Reason* and continues up through recent work found in Bayne (2010) and others.

Why is *Unity* needed? Consider the following example, based on Carruthers' (2008: pp. 63–6) interpretation of data from uncertainty tests as discussed in, e.g., Smith et al. (1997). We take the example from Carruthers' analysis noting that he does not address the matter of the set's unity. He uses the numbers in brackets to identify weak ($_w$) or strong ($_s$) first-order representational states. Then, in a move that foreshadows our principle of *Unity*, he postulates a gatekeeper mechanism at the forefront of the mind (Carruthers 2008: p. 67). The mechanism acts as a switch and adjudicates between which beliefs and desires gain entry. Contradictory beliefs of equal weight (e.g., a belief that *P* and an equally strong belief that *not-P*) will cancel each other out. Strong beliefs about a matter ($BELIEF_s$) will outweigh contrasting weak beliefs ($BELIEF_w$) about that same matter and so the strong beliefs will gain entry while the weak beliefs idle.

(1) $BELIEF_s$ [if the pattern is dense and D is pressed, then food results]
(2) $BELIEF_w$ [the pattern is dense]
(3) $DESIRE_s$ [food]
(4) $BELIEF_s$ [if the pattern is sparse and D is pressed, then a time out results]
(5) $BELIEF_w$ [the pattern is sparse]
(6) $DESIRE_s$ [no time out]

The first three states taken together, Carruthers writes (2008: p. 64), "generate (7), a weak desire to press D in order to obtain food [. . .] But states (4) through (6) likewise create (8), a weak desire not to press D, in order to avoid time out."

(7) $DESIRE_w$ [press D]
(8) $DESIRE_w$ [do not press D]

[10] Insofar as an animal's beliefs and desires are unified in the relevant situations, this fact, if it is a fact, at least *suggests* that basic cognition feels to an animal no differently than basic cognition feels to us, when we are not metacognizing.

Note that the beliefs must arrive at the gatekeeper at the same time or else they will not function in the way required by the first-order explanation. This observation forms the basis for P2, below. We can explicate Carruthers' argument as follows:

P1: Representational states (1)–(8) exist.
P2: If (a) representational states (1)–(8) exist and (b) are unified in animal A, then A will press a question mark.
C: Animal A will press a question mark.

As is apparent, the argument is valid only if we add the following premise which cannot be explicitly found in Carruthers' analysis:

P3: Representational states (1)–(8) are unified in animal A.

The conjunct (b) in the antecedent of P2 is attributed to Carruthers' explanatory schema as an implicit assumption; only if P3 is supplied (or P1 is appropriately modified) will the argument succeed in showing that this set of beliefs and desires entails conflict, in a particular animal, about which key to press. This conflict, generated by the unified set (1)–(8), may be either a first-order (non-metacognitive) belief or feeling of the animal; we think the most plausible hypothesis is that it is a feeling of conflict or uncertainty, as Carruthers suggests in later work (Carruthers and Ritchie 2013). But, our point here is that the eight states must be appropriately unified in an individual for the argument to be valid.

4 *Unity*, Once Again

We are now in position to give a more precise formulation of the principle of *Unity*.

> *Unity* [refined]: A set, S, of two or more mental states [(1), (2)…], in a neurologically healthy individual animal, A, is unified just in case the states in S are colligated synchronously and diachronically in A's neural structure such that all relevant contents of S can be accessed by A for a sufficient (but not unlimited) period of time, $t_1 \ldots t_n$, for A to perform a cognitive operation (such as an inference from S, or a judgment about possible actions in response to S).

Several clarifications are in order. First, the restriction to an individual animal is necessary to rule out worrisome metaphysical implications (see below, Sections 5.1–5.2). Second, the designation of a specific period of time is necessary to ensure our analysis remains empirically falsifiable and sensitive to future neuroscientific discoveries (see below, Sections 5.3–5.4). Third, the individual must be neurologically healthy to rule out, for instance, individuals having dissociative identity disorder (formerly called multiple personality disorder) in which S might be unified in one of A's identities but not in another of A's identities. Fourth, *Unity* being satisfied does not imply that A is self-aware of S; however, it does imply that A experiences S. Fifth, if Carruthers and Ritchie (2013) are right that a first-order account requires positing

feelings, and feelings are quintessentially phenomenal states, then such states should also be unified and thus our principle should be interpreted to include phenomenal states. Our principle is consistent with all of the relevant states being phenomenal, or having a phenomenal aspect, but does not require it. Lastly, we do not contend to know where or how the unity of S occurs in the neural structure of the brain, a question best addressed by neuroscience.[11] We argue only that *Unity* must be present for the gatekeeping mechanism posited by Carruthers to be triggered and thus for uncertainty responses to occur.

So our contention is this. Animal and human behaviors in situations of uncertainty are psychologically possible only if they are properly unified. *Unity* is a necessary condition of any causal, folk psychological explanation of uncertainty responses.

Using Carruthers' example, a set of beliefs and desires is unified if the following obtains:

Animal A: (1), (2), (3), (4), (5), (6), (7), (8), [1–8]

The unity of (1) through (8) is represented as [1–8]. Not only do states (1) through (8) exist, but they are unified in animal A. Without acknowledging [1–8], Carruthers and other deflationists cannot explain why *this* animal hesitates in front of the ambiguous screen. With *Unity* in hand, however, the way is clear to conclude that first-order states explain A's behavior.

To further clarify our principle, we will situate it against other conceptions of unity.[12] For example, Tye (2003: pp. 11–15) and Bayne and Chalmers (2003: pp. 24–27) provide taxonomies of unity. They each posit several types of unity (with some differences in their characterizations): object unity (multiple features represented as part of one object), spatial unity (multiple features represented as being in one space), neurophysiological unity (cognitive powers correlated with one neurological mechanism or a specific neural set), subject unity (psychological or conscious states occurring together within one subject), and more.

Broadly speaking, our *Unity* is a kind of subject unity (the unity required for a single subject having multiple states simultaneously or diachronically), which we assume should be grounded in neurophysiological unity. We note that our principle only requires, as made explicit in the formulation, that A have access consciousness (Block 1995)—associated with responsive and discriminatory consciousness, as discussed by Tye (2003) and mentioned in Section 1—not phenomenal consciousness (though our principle is compatible with phenomenal consciousness). Therefore, *Unity* is also a kind of *access unity*, a concept established by Bayne and Chalmers (2003: p. 29): the relevant mental states must be "jointly accessible" (accessible together at once) for judgment and guiding action. What matters, they note, is that the states are accessible, not that they are actively accessed; thus, it is a dispositional conception of access unity.[13] Whereas access unity corresponds with access consciousness,

[11] Whether *Unity* means that all contents of the metacognitive state must be globally broadcast in the brain or whether it applies only to the contents of specific modules is a matter we do not take up here.

[12] Thanks to an anonymous reviewer for encouraging us to compare our notion of unity to others.

[13] Not only must the two or more states be disposed to be accessed at time t_1 (or between t_1 and t_n, for diachronic cases), but the subject must be disposed to access them at t_1 (or during t_1 through t_n), in order for access to be possible.

phenomenal unity corresponds with phenomenal consciousness, when two or more mental states "are jointly experienced," i.e., "there is something it is like to be in both states at once" (Bayne and Chalmers 2003: p. 29).[14,15]

With our principle clarified and situated against some other conceptions of unity, we move to challenges to *Unity*. Although Bayne and Chalmers (2003: pp. 31–32) criticize access unity taken as a *general* account of unity, their worries are relevant to whether our principle—interpreted as a kind of access unity—is sufficient to meet the job we set for it. Bayne and Chalmers (2003: p. 32) justifiably doubt that all possible sets of access-conscious states can be access unified at the same time. For, as they claim, there could be millions of states, creating a complex conjunction of states very implausible for an agent to access at once in order to make judgments and guide behavior. However, they doubt even a weaker requirement according to which any two access-conscious states be access-unified at a time (Bayne and Chalmers 2003: p. 32). Access unity for two states, P and Q, can break down (Bayne and Chalmers 2003: pp. 35–37) in cases where P and Q are individually accessible, but P and Q together are not accessible, perhaps due to an "access bottleneck" at a specific time (Bayne and Chalmers 2003: p. 35). For instance, an experiment conducted by Sperling (1960) and discussed by Bayne and Chalmers (2003: p. 35) seems to show that joint access to a complex set of informational states can be inhibited while access to individual bits of that information is not. Subjects are briefly shown a matrix of three rows with four letters each; when queried, they can report more accurately on the contents of any of the three individual rows, but much less accurately on the contents of the entire matrix. The oddity is that they do seem to "know" reasonably well what information each row contains, but they cannot access the information of all three rows simultaneously. In response to this worry, as already noted, Bayne and Chalmers (2003) want their account of unity to be generalizable so that *any* two access-conscious states should be access-unified.

We agree that access can be inhibited in some cases. Yet, in others, it will hold. All we claim is that for any first-order schema like Carruthers' to work, any two states in the *relevant kinds* of cases (not all cases, thus not just *any* two possible states) must be accessible at once. In other words, only some kinds of states, at specific periods of time, need to be access-unified on our account. Although access unity must be present in the animals in the experiments if a first-order explanation is to succeed in accounting for their behavior without positing metacognition, notice that even if the animals *are* metacognizing, the relevant states need to be access-unified. This is because metacognizing creatures must-have access to two or more states that they are thinking about (the targets of the higher-order state). This is an important point because it shows that our invocation of a kind of access unity in the first-order, non-metacognitive interpretation is not something a higher-order interpreter should dispute.

In response to our claim that *Unity* is required for the first-order explanatory schema to work, a critic might object in one of two ways. First, one could resist *Unity*, contending that "All that is needed is for the relevant mental states to be processed in the right way, that is, in a step-by-step algorithmic order. Once the final step is reached,

[14] The taxonomy of unity becomes more complicated if we combine the broader concepts of access unity and phenomenal unity with the different kinds of unity discussed above (object unity, spatial unity, etc.). These variations are not central to our argument.

[15] Bayne and Chalmers (2003: p. 33, 46) are primarily interested in subsumptive phenomenal unity, in which a set of phenomenally conscious states are subsumed under a single phenomenal state.

conflict is generated and the uncertainty response results." This objection, however, misses our point. What is needed is not just a series of states algorithmically arranged, but an integrated collection of states held by a particular animal at a particular time capable of causing uncertainty behaviors. *Unity* is required for first-order schemas like Carruthers' to work.

Second, one could downplay the significance of *Unity*. Yes, says this critic; *Unity* is required for Carruthers' explanatory schema to work but this fact is so obvious that it hardly bears mentioning, as if it were important to add that the first-order explanations require that the animal be alive. We agree that philosophers are not obligated to mention every one of their assumptions. And, even though some philosophers have deemed the unity of consciousness worthy of careful analysis, as we have mentioned, we agree that its significance here may not be immediately apparent. We devote the remainder of this paper, therefore, to addressing this criticism. We do so by describing two cases, each one consistent with Carruthers' explanatory schema sans *Unity*. Each case has undesirable philosophical results—results which our principle excludes.

5 Why First-Order Accounts Cannot Do without *Unity*

We provide two arguments for *Unity* (Sections 5.1 and 5.3) and show why each one matters (Sections 5.2 and 5.4).

5.1 Too Many Animals

For starters, here is a logical possibility that Carruthers' explanatory schema cannot discount. The six states, (1)–(6) may be spread over two animals without any cognitive connection and thus without giving rise to (7) and (8), as this arrangement shows:

Animal B: (1), (2), (3)
Animal C: (4), (5), (6)

Rather, in this case, animal B will automatically and unreflectively press D; and animal C will, similarly, non-problematically press S. In Too Many Animals, the way in which desires and beliefs are present fails to produce the conjunction of (7) and (8), which are nonetheless displayed in the hesitating behavior of the test animals. Therefore, Carruthers' first-order explanation does not show what causes the gatekeeping mechanism to launch into action.

We contend that the production of (7) and (8) requires a unified set of beliefs and desires, (1)–(6)—in keeping with *Unity*—in which the contents of the individual beliefs and desires are responsive to, and modifiable by, each other. This is a metaphysically significant requirement, as we next argue.

5.2 Brainets: Why Too Many Animals Matters

Unity is important not only to avoid a logical conundrum but to keep Carruthers' metaphysics in check. Without *Unity*, minds begin to multiply.

Suppose we have three monkeys: m1, m2, and m3. Each individual has a distinct mind or brain (let us assume some version of materialism about the mind is correct), and its beliefs and desires are its own. Now suppose that m1–m3 enter a sort of "mind-meld" as a result of scientists connecting their brains in what Miguel Nicolelis calls a "Brainet." Brainets are groups of three or more animal brains (mice in some experiments, monkeys in others) accomplishing a common task by "cooperating and exchanging information in real time through direct brain-to-brain interfaces" (Pais-Vieira et al. 2015). Brainets have been implemented in experiments in which three monkey brains are joined by computers and tasked to move an avatar arm in a cursor-moving task.[16] In the Brainet, particular beliefs and desires in the minds of m1–m3 "collaborate" to undertake an action A (e.g., moving a cursor). The monkeys are not aware that they are collaborating. Each animal is located in a separate room and does not know that its brain is connected to two other brains.

In a Brainet, the first-order beliefs and desires of the three animals work together to achieve an end sought independently by each animal. The three animals do not create a fourth entity that seeks the end. Such a fourth entity would require us to say what is implausible, namely that the technology unifies the three animals' beliefs and desires into a new mind. Were a fourth entity being created, one might be excused for enthusing that the new mind "*self-adapts* to achieve a common motor goal" (Ramakrishnan et al. 2015, emphasis added). But the implicit claim that there is a new mind and the explicit claim that the new mind is autonomous would only be plausible were the Brainet forming its own beliefs and desires based on input it is receiving independently of m1–m3. It is not. Rather, three brains, each working on its own, is cooperating with the others to do A. They are not being united into a fourth mind to do A.

We reject the spooky idea that Brainet technology has the power to create minds. And yet, if Brainet researchers adopted Carruthers' first-order cognitive architecture while leaving *Unity* behind, they could endorse this counterintuitive conclusion. Without *Unity*, three animals playing a computer game together may be explained by postulating the creation of a new mind, m4. We illustrate this mind on the right side of the figure below. As can be seen, m4 consists of six first-order beliefs and desires, each one borrowed from the animal indicated, plus two new beliefs, (7) and (8), produced by the conjunction of (1)–(6):

Animal D: (1), (2)
Animal E: (3), (4) ➔ m4: (1), (2), (3), (4), (5), (6), (7), (8) [1–8]
Animal F: (5), (6)

Notice that the contents of m4's mind are the same as animal A. However, since the assumptions with which we are working include only three animals (D, E, and F) and exclude animal A, it must be that m4 is a *new* mind. Have Ramakrishnan and fellow Brainet researchers taken the mental states distributed across three monkeys and used them to create a mind unattached to any of the three brains? And is this new disembodied mind self-adaptively generating (7) and (8) to produce [1–8]? No. Rather,

[16] If we can do this with monkeys, we can do it with humans. Hirstein (2012) argues for the possibility of mind-melding between humans, resisting the claim that the mind is necessarily private. The question arises whether such mind-melding creates an additional mind, a conclusion that seems virtually impossible to reach once *Unity* is firmly in place.

m1–m3 are working independently of each other, each one separately obeying *Unity*. In so doing, they are (involuntarily) cooperating to achieve a shared goal, not collectively generating a mind. *Unity* being absent, Carruthers' schema risks ontological profligacy by opening the door to the case of Too Many Animals.

We turn now to our second argument for *Unity*.

5.3 Temporal Disjunction

Suppose that an animal at t_1 has (only) these four beliefs:

Animal G at t_1: (1), (2), (3), and (6)

Then, 4 s later at t_2, the animal has lost these states and now, instead, holds (only) these two:

Animal G at t_2: (4) and (5)

Note that states (1)–(6) are unified if we take as our temporal frame times t_1 through t_2, but disunified if we look at t_1 and t_2 separately. If Carruthers' conflict between (7) and (8) is to be generated in a way that produces an uncertainty response, however, (1)–(6) must all be present *at a time*—that is, either at t_1 or at t_2. But this is impossible in the scenario envisioned above; the conflict cannot exist at t_1 because at t_1, (7) has been generated by the conjunction of (1)–(3) and yet its partner in crime, (8), does not exist because (4) and (5) do not (yet) exist. Nor can the conflict exist at t_2. For at t_2, the subject's mental states generate neither (7) nor (8).

5.4 Unfalsifiable Claims: Why Temporal Disjunction Matters

Whether or not the brain is a massively asynchronous, parallel system, the timing of events in it is critical to understand its workings (Zeki 2015). Many animal beliefs and desires come and go quickly; they do not persist in the animal's brain for a long period of time. Whenever two such countervailing evanescent beliefs arrive at the gatekeeper mechanism and cause the mechanism to refuse admittance to either belief, it must be true that both beliefs reach the mechanism at the same time. Or, to be precise, roughly the same time (see below). For if they arrive at separate times, the mechanism will admit first one and then the other. The beliefs, in other words, must be temporally unified for any first-order explanation to work. Suppose that an animal at t_1 possesses (2):

(2) BELIEF_{w} [the pattern is dense]

Three seconds later, at t_2, the animal has lost (2). But it now possesses (5):

(5) BELIEF_{w} [the pattern is sparse]

The temporal gap between the two beliefs means that (5) plays no role in producing hesitation at t_1. In the absence of (5), belief (2) would ordinarily determine the animal's

behavior; we expect it to press D because the gatekeeping mechanism has not been triggered and it is free to act on (2). But Temporal Disjunction concerns a different case. Here, (5) is not present at t_1 and yet, contrary to expectations, the animal does not act on (2). What causes the suppression of the impulse to act on (2) in Temporal Disjunction? The explanation cannot be first-order because the gatekeeping mechanism is not in play. How then do we explain *this* case of uncertainty?

A number of possibilities suggest themselves. Perhaps the animal's limbic system is preoccupied with some other matter, or has entered the so-called default "resting" mode, or is in some other way incapable of responding appropriately to (2). If the animal's affective systems must be enlisted to act on (2), then problems in the limbic system might cause hesitation. (Imagine that a conspecific has shrie-ked in another room, hijacking the experimental animal's emotions.) Or, perhaps the animal's motor cortex is unprepared. If the motor cortex must be enlisted to act on (2), then problems in this area might cause hesitation. (Imagine that the computer screen in front of the animal has begun to shake and the animal is busy steadying it.) If the motor cortex must reach a specific readiness potential state to sponsor acting on (2) and this state is not present, then this fact might cause hesitation. Both of these possibilities are plausible candidates to explain hesitation to act on (2) in the absence of (5). Here is a third possibility, more to our point. Perhaps a metacognitive process, running concurrently in the background at t_1, has been on the lookout for (2) and now becomes aware of it. The metacognitive process puts the brakes on (2). Suppose the metacognitive process has been primed by previous experience to be ready to override (2) when it appears. Perhaps, it has reasons of its own to be suspicious of (2) and so it steps in quickly and short circuits the desire to act on (2).

Let us flesh out this third picture with a few more details. Macaques and humans can subconsciously identify objects after seeing pictures of them for as little as 13 ms (Potter et al. 2014). We will call these sorts of beliefs *proximal*. Proximal beliefs are unconscious and evanescent, coming and going without the agent's awareness of them and rarely enduring for more than tens of milliseconds. Contrast proximal beliefs with *standing* beliefs. Standing beliefs endure for at least a half a second and, in some cases, for a lifetime. Consequently, they are available to us; we can become aware of them. Metacognitive beliefs are standing beliefs; for example, it takes at least 300 ms for us to become conscious of—to be able to report our having seen—a stimulus flashed on a screen (Cul, Baillet, and Dehaene 2007). If the stimulus is a picture of oneself flashed for 13 ms on a screen, one might form a proximal belief that one has seen oneself but not be able to say why one has this belief. However, if the stimulus is a picture of oneself and visible for 500 ms, one may be able to point to the screen and say "That was me." Only in the second case would metacognition be at work.

How are proximal and standing beliefs relevant to first-order explanations of purported metacognition? Metacognition operates on a continuum with sub-personal cognitive processes. Metacognition may be set in motion when an evanescent proximal belief "times out," that is, leaves the area of the gatekeeping mechanism. The gate-keeping mechanism, a domain specific module dedicated to one specialized task, takes mostly proximal beliefs as inputs. When it is flooded with such beliefs, however, it cannot produce a result before the proximal beliefs expire. In this case, metacognitive processes, if the individual has them, must take over using standing beliefs it can

consciously retrieve. And so the individual hesitates, moves its head around, squints at the screen, and seeks more information—exhibiting the behaviors experimentalists observe in cases of human metacognition.

Call the point at which the gatekeeping mechanism gives way to metacognition the point of critical self-cognitive load, or t_c. At t_c, the mechanism fails and throws its results into (what Bernard Baars calls) a global workspace. Here, working memory can broadcast its contents to various domains and profit from the work of multiple serial operations using meta-representations (Baars 1988, 2005). When cognitive load is light—say, for example, that all of the screens to be interpreted in the monkey experiments are clearly sparse or dense, or the pictures of myself in our self-recognition thought experiment are unproblematically *me*—the gatekeeper is able to handle all chores and, in these cases, first-order explanations suffice. But as the screens become harder to identify and the animal begins using more and more time to bind together the various inputs—or as I have a harder and harder time deciding whether the picture is really of me and not my brother—the subject's brain will be challenged with increasingly large sets of data points and increasingly urgent deadlines. The demands require keeping a very large number of proximal beliefs and desires in active memory. If the number multiplies further, the mechanism may become overloaded. Eventually, the sheer magnitude of inputs either shuts down the mechanism or, if second-order reflective capacities are available, triggers help from the higher order. The higher order, consisting in part of active memory and standing beliefs, is able to go back in time and retrieve the standing beliefs required for the task.

If this picture is correct, our theory must be on the lookout for any instance when the gatekeeper is overrun with proximal beliefs; for at this point, either the non-metacognizing animal is frustrated and simply stops what it is doing altogether, or, if the animal has metacognition, it employs that resource, turning its attention to its standing beliefs and desires.

Where is the point in time where metacognition begins? Answering this neuroscientific question is not within the scope of this paper. We want only to insist on the importance of *Unity* to the matter. For if our theory of animal cognition does not decisively rule out the case of Temporal Disjunction, then the various mental inputs into a monkey's mind can always be stretched out over time to mechanistic processes occurring much later. But this is cheating. As we have seen, inputs must arrive at the gatekeeper mechanism at the same time. First-order explanations cannot justifiably appeal to mental states that occur distally in time because if the second of two proximal beliefs arrives at the gatekeeper mechanism after the first proximal belief has expired, then the gatekeeper mechanism will not be able to consider both beliefs at the same time. And, in this case of temporally unrelated unconscious beliefs, the gatekeeper mechanism will not explain anything.

Here is our worry re-stated. Deflationary theories of metacognition excluding *Unity* seem unfalsifiable insofar as they make possible first-order explanations of animal behavior that appeal to beliefs and desires wherever they may be found. With such ample, albeit unrelated resources in hand, the skeptical explainer of animal behavior need not worry that animal behaviors—no matter how slow, effortful, or meditative—will add up to something that requires an appeal to a metacognitive process. Our point is that, without the constraint of unity over time, analysts may make false negative mistakes, so-called type II errors of incorrectly believing a false null hypothesis is true.

In the case of the unwary deflationist, this would mean missing signs of metacognition and drawing the conclusion "no metacognition present" when metacognition is present (cf. Sober 2005; Andrews and Huss 2013).

6 Concluding Remarks: *Unity* and the Moral Standing of Monkeys and Similar Animals

Some nonhuman animals may think about their thoughts even if no experimental data yet show they do. We have argued that first-order schemas, such as Carruthers (2008), supplemented by our principle, *Unity*, explain the existing data (Section 3). These theories must recognize *Unity* or risk two unwelcome implications: postulating minds where there are none (Sections 5.1 and 5.2), and allowing the generation of scientific models of animal minds that are impervious to falsification (Sections 5.3 and 5.4). Keeping *Unity* clearly in view will help experimentalists identify signs of metacognition should they appear in future studies.

Moreover—to offer a more speculative, anti-neo-Cartesian conclusion—it seems that the conditions identified in *Unity* (plus beliefs, desires, and emotions) may be all that is required for conscious experience of a relatively sophisticated, complex sort. Assume for the sake of argument that the integrated information theory (IIT) of consciousness (Tononi 2008) is correct. According to Tononi (2008: p. 217), consciousness is "integrated information." Represented as ϕ, integrated information is defined as "the amount of information generated by a complex of elements, above and beyond the information generated by its parts" (Tononi 2008: p. 216). According to IIT, consciousness is a measurable phenomenon and the degree of consciousness correspondingly increases with the degree of integrated information. Even very simple animals on this view have a degree of consciousness, so it stands to reason that monkeys will have a relatively high degree of consciousness. Although IIT is controversial (for example, see Cerullo 2015), it does retain attractive features: in quantifying consciousness, it makes it more measurable and objective. We only employ IIT here as a model of how information in the brain and consciousness might be correlated.[17]

If the myriad, information-carrying mental states in monkeys and relevantly similar animals are unified—as must be the case if they are to perform as they do on the types of metacognition tests discussed in this paper—then they must be integrated in such a way that together they can do things (generate cognitive results) that no subgroup of them can do individually. Provided the arguments for *Unity* in this paper, the assumption of IIT, and that the amount of integrated information in animal brains is relatively high owing to their unification during key cognitive episodes—such as states of uncertainty—it follows that consciousness in these animals is assuredly more complex than deflationary accounts of metacognition, such as Carruthers' account, let on. Although these animals may or may not be metacognitive, they have sophisticated, unified minds. Given at least a rough correspondence between cognitive complexity

[17] Although there are reasons to doubt the existence of levels of consciousness as a conceptual necessity or neuroscientific reality (Bayne et al. 2016), as mentioned in footnote 2, this need not contradict talk of "degrees" or "amount" of consciousness (or, conscious contents) for a specific type of consciousness (e.g., phenomenal, discriminatory, responsive), or to the total information processing occurring in the mind.

and moral worth, we contend that monkeys and similar animals have a higher moral standing than fellow deflationary theorists may allow.

Acknowledgements We thank Dorit Bar-On for getting us started on this topic; Peter Carruthers for helpful criticisms when we presented at the 2017 University of Connecticut ECOM conference on "Human and Nonhuman Animals: Minds and Morals;" Irina Mikhalevich for comments at the 2016 Central APA meeting, as well as participants there; participants at the 2015 North Carolina Philosophical Society meeting, the 2014 Towards a Science of Consciousness Conference at the University of Arizona, and the 2013 Pacific University Northwest Philosophy Conference; and several anonymous reviewers.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

Andrews, K. (2012). *Do apes read minds?: toward a new folk psychology*. Cambridge: MIT Press.

Andrews, K., & Huss, B. (2013). Assumptions in animal cognition research. Proceedings of the CAPE International Workshops, Part II presented at the CAPE philosophy of animal minds workshop, Kyoto University, February 12. https://repository.kulib.kyoto-u.ac.jp/dspace/bitstream/2433/203241/1/capes_1_152.pdf.

Andrews, K., & Huss, B. (2014). Anthropomorphism, anthropectomy, and the null hypothesis. *Biology and Philosophy, 29*(5), 711–729. https://doi.org/10.1007/s10539-014-9442-2

Armstrong, D. (1993). *A materialist theory of the mind*. New York: Routledge. (Originally published in 1968).

Baars, B. J. (1988). A cognitive theory of consciousness. In *Cambridge [England]*. New York: Cambridge University Press.

Baars, B. J. (2005). Subjective experience is probably not limited to humans: the evidence from neurobiology and behavior. *Consciousness and Cognition, 14*(1), 7–21. https://doi.org/10.1016/j.concog.2004.11.002

Bayne, T. (2010). *The unity of consciousness*. New York: Oxford University Press.

Bayne, T., & Chalmers, D. J. (2003). What is the unity of consciousness? In A. Cleeremans (Ed.), *The unity of consciousness: binding, integration, dissociation* (pp. 23–58). New York: Oxford University Press.

Bayne, T., Hohwy, J., & Owen, A. M. (2016). Are there levels of consciousness? *Trends in Cognitive Sciences, 20*(6), 405–413.

Beran, M. J., Couchman, J. J., Coutinho, M. V. C., Boomer, J., & David Smith, J. (2010). Metacognition in nonhumans: methodological and theoretical issues in uncertainty monitoring. In A. Efklides & P. Misailidi (Eds.), *Trends and prospects in metacognition research* (pp. 21–35). New York: Springer http://www.springerlink.com/content/u74731956121t3l2/

Beran, M. J., & Smith, J. D. (2011). Information seeking by rhesus monkeys (Macaca mulatta) and capuchin monkeys (Cebus apella). *Cognition, 120*(1), 90–105. https://doi.org/10.1016/j.cognition.2011.02.016

Block, N. (1995). On a confusion about the function of consciousness. *Behavioral and Brain Science, 18*, 227–247.

Carruthers, P. (1989). Brute experience. *The Journal of Philosophy, 86*(5), 258–269. https://doi.org/10.2307/2027110

Carruthers, P. (1992). *The animals issue: moral theory in practice*. Cambridge [England]: Cambridge University press.

Carruthers, P. (2008). Metacognition in animals: a skeptical look. *Mind & Language, 23*(1), 58–89. https://doi.org/10.1111/j.1468-0017.2007.00329.x

Carruthers, P., & Brendan Ritchie, J. (2013). The emergence of metacognition: affect and uncertainty in animals. In M. J. Beran, J. Brandl, J. Perner, & J. Proust (Eds.), *Foundations of metacognition* (pp. 76–93). Oxford: Oxford University Press. http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199646739.001.0001/acprof-9780199646739-chapter-005.

Cerullo, M. (2015). The problem with phi: a critique of integrated information theory. *PLoS Computational Biology, 11*(9), 1–12. https://doi.org/10.1371/journal.pcbi.1004286

Couchman, J. J., Coutinho, M. V. C., Beran, M. J., & Smith, J. D. (2009). Metacognition is prior. *Behavioral and Brain Sciences, 32*(02), 142–142. https://doi.org/10.1017/S0140525X09000594

Couchman, J. J., Coutinho, M. V. C., Beran, M. J., & Smith, J. D. (2010). Beyond stimulus cues and reinforcement signals: a new approach to animal metacognition. *Journal of Comparative Psychology, 124*(4), 356–368. https://doi.org/10.1037/a0020129

Del Cul, A., Baillet, S., & Dehaene, S. (2007). Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS Biology, 5*(10), e260. https://doi.org/10.1371/journal.pbio.0050260

DeGrazia, D. (2009). Self-awareness in animals. In R. W. Lurz (Ed.), *The philosophy of animal minds* (Vol. 201). Cambridge: Cambridge University Press.

Fitzpatrick, S. (2008). Doing away with Morgan's canon. *Mind & Language, 23*(2), 224–246. https://doi.org/10.1111/j.1468-0017.2007.00338.x

Foote, A. L., & Crystal, J. D. (2007). Metacognition in the rat. *Current Biology, 17*(6), 551–555. https://doi.org/10.1016/j.cub.2007.01.061

Gennaro, R. J. (2009). Animals, consciousness, and I-thoughts. In R. W. Lurz (Ed.), *The Philosophy of Animal Minds* (pp. 184–200). Cambridge: Cambridge University Press.

Goldman, A. I. (1993). The psychology of folk psychology. *Behavioral and Brain Sciences, 16*(01), 15–28. https://doi.org/10.1017/S0140525X00028648

Hampton, R. (2001). Rhesus monkeys know when they remember. *PNAS, 98*(9), 5359–5362.

Hampton, R. R. (2009). Multiple demonstrations of metacognition in nonhumans: converging evidence or multiple mechanisms? *Comparative Cognition & Behavior Reviews*, *4*(January), 17–28.

Hirstein, W. (2012). *Mindmelding: consciousness, neuroscience, and the mind's privacy*. New York: Oxford University Press.

Jacobson, H. (2010). Normativity without reflectivity: on the beliefs and desires of non-reflective creatures. *Philosophical Psychology, 23*(1), 75–93. https://doi.org/10.1080/09515080903532282

Jozefowiez, J., Staddon, J. E. R., & Cerutti, D. T. (2009). Metacognition in animals: how do we know that they know? *Comparative Cognition & Behavior Reviews, 4*. https://doi.org/10.3819/ccbr.2009.40003.

Kant, I. (1998). *Critique of pure reason*. Translated by Paul Guyer and Allen W. Wood. 1781, 1st ed./1787 2nd ed. Cambridge; New York: Cambridge University Press.

Karin-D'Arcy, M. R. (2005). The modern role of Morgan's canon in comparative psychology. *International Journal of Comparative Psychology, 18*(3) http://escholarship.ucop.edu/uc/item/3vx8250v#page-2

Kornell, N., Son, L. K., & Terrace, H. S. (2007). Transfer of metacognitive skills and hint seeking in monkeys. *Psychological Science, 18*(1), 64–71. https://doi.org/10.2307/40064579

Le Pelley, M. E. (2012). Metacognitive monkeys or associative animals? Simple reinforcement learning explains uncertainty in nonhuman animals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(3), 686–708. https://doi.org/10.1037/a0026478

Malassis, R., Gheusi, G., & Fagot, J. (2015). Assessment of metacognitive monitoring and control in baboons (Papio Papio). *Animal Cognition, 18*(6), 1347–1362. https://doi.org/10.1007/s10071-015-0907-8

Morgan, C. L. (1894). *An introduction to comparative psychology*. London: Walter Scott.

Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review, 83*(4), 435.

Pais-Vieira, M., Chiuffa, G., Lebedev, M., Yadav, A., & Nicolelis, M. A. L. (2015). Building an organic computing device with multiple interconnected brains. *Scientific Reports* 5 (July). https://doi.org/10.1038/srep11869.

Potter, M. C., Wyble, B., Hagmann, C. E., & McCourt, E. S. (2014). Detecting meaning in RSVP at 13 ms per picture. *Attention, Perception, & Psychophysics, 76*(2), 270–279. https://doi.org/10.3758/s13414-013-0605-z

Proust, J. (2009). Overlooking metacognitive experience. *Behavioral and Brain Sciences, 32*(2), 158–159.

Proust, J. (2010). Metacognition. *Philosophy Compass, 5*(11), 989–998. https://doi.org/10.1111/j.1747-9991.2010.00340.x

Ramakrishnan, A., Ifft, P. J., Pais-Vieira, M., Byun, Y. W., Zhuang, K. Z., Lebedev, M. A. & Nicolelis, M. A. L. (2015). Computing arm movements with a monkey Brainet. *Scientific Reports* 5 (July). https://doi.org/10.1038/srep10767.

Rosati, A. G., & Santos, L. R. (2016). Spontaneous metacognition in rhesus monkeys. *Psychological Science, 27*(9), 1181–1191. https://doi.org/10.1177/0956797616653737

Smith, J. D. (2009). The study of animal metacognition. *Trends in Cognitive Sciences, 13*(9), 389–396. https://doi.org/10.1016/j.tics.2009.06.009

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*(4), 391–408. https://doi.org/10.1037/0096-3445.124.4.391

Smith, J. D., Shields, W. E., Schull, J., & Washburn, D. A. (1997). The uncertain response in humans and animals. *Cognition, 62*(1), 75–97. https://doi.org/10.1016/S0010-0277(96)00726-3

Smith, J. D., Shields, W. E., & Washburn, D. A. (2003). The comparative psychology of uncertainty monitoring and metacognition. *Behavioral and Brain Sciences, 26*, 317–373.

Sober, E. (2005). Comparative psychology meets evolutionary biology: Morgan's canon and cladistic parsimony. In L. Daston & G. Mitman (Eds.), *Thinking with animals: new perspectives on anthropomorphism* (pp. 85–99). New York: Columbia University Press.

Sober, E. (2009). Parsimony and models of animal minds. In R. W. Lurz (Ed.), *The philosophy of animal minds* (Vol. 257). Cambridge: Cambridge University Press.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied, 74*(11), 1–29.

Tononi, G. (2008). Consciousness as integrated information: a provisional manifesto. *Biological Bulletin, 215*, 216–242.

Tye, M. (2003). *Consciousness and persons: unity and identity*. Cambridge, MA: MIT Press.

Washburn, D. A., Smith, J. D., & Shields, W. E. (2006). Rhesus monkeys (Macaca Mulatta) immediately generalize the uncertain response. *Journal of Experimental Psychology: Animal Behavior Processes, 32*(2), 185–189. https://doi.org/10.1037/0097-7403.32.2.185.

Zeki, S. (2015). A massively asynchronous, parallel brain. *Philosophical Transactions of the Royal Society B, 370*(1668), 20140174. https://doi.org/10.1098/rstb.2014.0174