The highs and lows of theoretical interpretation in animal-metacognition research

J. David Smith1,*, Justin J. Couchman2 and Michael J. Beran3

1Department of Psychology and Center for Cognitive Science, The University at Buffalo, State University of New York, Buffalo, NY 14260–4110, USA
2Department of Psychology, State University of New York at Fredonia, Fredonia, NY 14063, USA
3Language Research Center, Georgia State University, Decatur, GA 30034, USA

Humans feel uncertain. They know when they do not know. These feelings and the responses to them ground the research literature on metacognition. It is a natural question whether animals share this cognitive capacity, and thus animal metacognition has become an influential research area within comparative psychology. Researchers have explored this question by testing many species using perception and memory paradigms. There is an emerging consensus that animals share functional parallels with humans’ conscious metacognition. Of course, this research area poses difficult issues of scientific inference. How firmly should we hold the line in insisting that animals’ performances are low-level and associative? How high should we set the bar for concluding that animals share metacognitive capacities with humans? This area offers a constructive case study for considering theoretical problems that often confront comparative psychologists. The authors present this case study and address diverse issues of scientific judgement and interpretation within comparative psychology.

Keywords: metacognition; uncertainty monitoring; metamemory; comparative cognition; decision-making

1. INTRODUCTION

Humans know when they do not know or remember. They respond well to uncertainty by deferring response and seeking information—for example, they Google. Humans’ responses to uncertainty ground the literature on metacognition [1–6]. Metacognition is defined to be the monitoring and control of basic perceptual and cognitive processes. The theoretical assumption is that some minds can deploy a cognitive executive that oversees and optimizes thought and problem-solving. Researchers assess these metacognitive functions in humans by collecting judgements of confidence, feelings of knowing and tip-of-the-tongue experiences.

Humans’ metacognitive capacity is linked to sophisticated aspects of mind. Metacognition can reveal a hierarchical structure to cognition, because often metacognitive processes regulate lower-level cognitive processes [7]. Metacognition reveals humans’ awareness of their cognition [8,9], because humans often reflect consciously on their cognitive states and declare them to others. Metacognition may also reveal humans’ self-awareness [10], because states like uncertainty are often imbued with self feelings (e.g. I don’t know).

Metacognition’s sophistication raises the question of whether it is uniquely human, and one of comparative psychology’s current goals is to establish whether non-human animals (hereafter, animals) share this capacity [11–13]. If they do, it could bear on their consciousness and self-awareness [14], and it would affect many theoretical debates within comparative psychology. Given the question’s importance, Smith and his colleagues inaugurated research on animal metacognition [15–19]. This area has been reviewed [11–13,20], and active research continues in this area [21–39].

To explore animal metacognition, one cannot just adopt the usual human measures like feelings of knowing and tip-of-the-tongue states. Animals have no way to declare these states and feelings. For the same reason, the usual human measures are not suitable for young humans [40]. Instead, comparative researchers have built behavioural tasks that have two components. First, researchers make some trials difficult, to stir up something like an uncertainty state in animal minds. Second, researchers give animals a response apart from the task’s primary discrimination responses that lets them decline to complete any trials they choose. This uncertainty response lets animals manage uncertainty and declare it behaviourally/observably. If animals are metacognitive and monitor internal uncertainty states or internal assessments of the probability of responding correctly, they should recognize difficult
trials as doubtful or error-causing and decline those trials proactively and adaptively.

We will illustrate this approach to studying animal metacognition, introducing readers to the area and noting features of the experiments that raise issues of high- and low-level theoretical interpretation. Above all, we will have to deal with the question of whether animals employ metacognitive or non-metacognitive strategies to avoid difficult trials. In the inaugural study [16], a dolphin (Tursiops truncatus) made a ‘high’ response to a 2100 Hz tone or a ‘low’ response to any lower tone (1200–2099 Hz). The frequency (Hz) of low trials was adjusted to constantly challenge the animal’s psychophysical limit and to maximize difficulty and uncertainty within the task. The animal could respond ‘uncertain’ to decline the trials he chose. Figure 1a shows that he assessed correctly when he was at risk for error, selectively declining the difficult trials near threshold. His uncertainty responses peaked near 2086 Hz, 0.11 semitones from the standard high tone. Humans perform similarly in this task, and humans say their uncertainty responses reflect their high-level, conscious, metacognitive states of uncertainty.

The dolphin said nothing, but produced distinctive uncertainty behaviours. We carried out a factor-analytic study of his ancillary behaviours on trials of different pitch. His hesitation-wavering behaviours peaked at his perceptual threshold, too (figure 1b). These hesitation behaviours could be additional behavioural symptoms of uncertainty. Tolman [41] appreciated these hesitation behaviours—that he called ‘lookings and runnings back and forth’—because he thought they might operationalize animal consciousness for the behaviourist. That is a provocative way to frame this field’s main question: do we accept Tolman’s definition, and grant the dolphin high-level, metacognitive uncertainty in this perceptual task, or not?

This is a difficult problem of scientific inference. Analogous questions have attended the study of animals’ counting, language, timing, self-awareness, theory of mind and so forth. In fact, the animal-metacognition literature is a good case study in this inference problem. It raises issues that generalize constructively to other comparative research domains. We present this case study here.

2. TESTING LOW-LEVEL INTERPRETATIONS OF ANIMAL METACOGNITION

The dolphin met the criteria described by Hampton [29] for a metacognitive performance. There was an observable behaviour (high and low responses) that could be scored as (in)correct. There was variation in the accuracy of primary responding across trial levels so that accuracy could be correlated to the use of the secondary, metacognitive response. There was a secondary, observable behaviour (the uncertainty response) that might reflect monitoring processes overseeing the animal’s primary responding. This secondary behaviour was strongly (negatively) correlated across trial levels with the accuracy of the primary responses.

Yet, one is hesitant to immediately credit animals with high-level metacognitive capacities.

Animal-metacognition research, like all comparative research, bears an interpretative burden given the tradition of explaining animals’ behaviour at the lowest psychological level [42]. Therefore, even given possibly metacognitive performances by some species, one must ask whether they might be explained using low-level, associative mechanisms. In fact, the possible low-level bases for uncertainty responses by animals—that is, the possibilities that these responses were elicited by stimuli or entrained by reinforcement contingencies—was the principal theoretical issue through the first decade of animal-metacognition research [29,34,43]. To the extent that animals’
uncertainty responses are triggered reactivly by stimulus cues, by reinforcement histories, and the like, one would conclude that animals' uncertainty systems present a weaker analogue to human uncertainty, and that animals are not metacognitive. To the extent that animals' uncertainty responses turn out to be more highly cognitive—more executive, more controlled, more deliberate, perhaps even conscious—one would conclude that animals' uncertainty systems present a stronger analogue to human uncertainty, and that animals are metacognitive. Consequently, research has focused sharply on the low-level cues and processes that animals might use to achieve metacognitive performances, and on whether these can be disconfirmed as the behavioural cause of those performances.

In this section, we describe this decade of research, including the theoretical concerns raised and the empirical answers offered. Some aspects of the resulting low-level/high-level dialogue represent comparative psychology at its best. The associative criticisms were disciplined and testable. They provoked new paradigms. They produced consensual answers and theoretical development in the field.

(a) Reinforced uncertainty responses

One associative concern was that animals sometimes received food rewards or tokens for uncertainty responses [27,28,30,36,37,44–46]. This approach could make the uncertainty response attractive solely for its reward properties, independent of any metacognitive role it plays in a task. This approach made it difficult to rule out low-level interpretations or to affirm metacognitive interpretations.

To address this concern, researchers removed the reward contingency for that response [19,22,35]. In one case [22], macaques judged whether arrays were less or more numerous than a session-specific numerosity. Numerosities nearer the boundary value were more difficult to classify. The uncertainty response only cleared the old trial and brought the next, randomly chosen trial. It offered no food reward, food token, trial hint or easy next trial for its use. But monkeys still made uncertainty responses selectively for the trials near the boundary value on which they would most probably err. The associative concern about the immediate appetitive attractiveness of the uncertainty response cannot explain this result.

(b) Reactions to primary stimulus qualities

Another associative concern was that difficulty level within uncertainty-monitoring tasks was often perfectly correlated with the objective stimulus level—for example, tone height (Hz) in the dolphin’s discrimination. Some stimuli would have caused animals frequent errors and ensured frequent penalty timeouts and lean rewards. These stimuli could have become aversive, and avoided through a default response that some mistook as a metacognitive response. This theoretical consideration recalls Hampton’s [29] concern that environmental cue associations could underlie seemingly metacognitive performances. In general, our typology of possible low-level descriptions is similar to Hampton’s typology.

To address this concern, researchers lifted uncertainty-monitoring tasks off the plane of concrete stimuli. In one case [15], macaques were allowed to make uncertainty responses in a same–different task. A same–different task—testing generalization over variable and novel stimulus contexts—requires some degree of abstraction beyond the absolute stimulus qualities that carry the relation. This abstractness explains why true same–different performances appear to be phylogenetically restricted [47] and why even non-human primates have distinctive weaknesses in same–different performance [48,49].

Accordingly, macaques made same or different responses to pairs of rectangles that had the same or different pixel densities. To cause them difficulty, the size of the density difference on different trials was adjusted in a threshold paradigm to constantly challenge subjects’ discrimination abilities. Moreover, same and different trials at several absolute pixel-density levels were intermixed to ensure a true relational performance. Yet, the macaques (figure 2) used the uncertainty response essentially identically to humans (a 0.97 cross-species correlation of the behavioural profiles), producing one of the closest correspondences between animals’ and humans’ performance. Shields et al. even reserved some regions of absolute density for use in immediate generalization tests to confirm the macaques’ generalizable same–different performance. This illustrates the constructive use of transfer tests to show the representational generality of macaques’ uncertainty processes, which might also indicate their similarity to humans’ uncertainty processes.

Uncertainty responses cannot have been triggered by low-level stimulus cues, because the performance survived immediate transfer tests and because in any case the relevant cue was abstract. Rather, uncertainty responses had to be prompted by the indeterminacy of the same–different relation instantiated by difficult and highly variable stimulus pairs.

Hampton [44] explored macaques’ metamemory using a delayed matching-to-sample task. With longer delays between sample presentation and match-choice selection, matching performance decreased because monkeys remembered the sample less well. Monkeys selectively declined memory tests at long retention intervals when they had mostly forgotten the sample. In addition, one monkey performed better at each delay level on the trials he chose to complete than on the trials he was forced to complete. This undermines the explanation that the monkey was just reacting to long delays with an escape response. Instead, it suggests that he was monitoring some psychological signal of (not) remembering. Both monkeys also responded uncertain more, no matter the length of the retention interval, when blank trials occurred with no sample shown, guaranteeing that macaques could not know what to match.

These monkeys cannot have been conditioned to avoid particular concrete stimuli. The memory targets only applied for a single trial, so longer term avoidance learning was useless. Moreover, the uncertainty response was made with no visible sample stimulus to trigger an avoidance response. These macaques showed a kind of metamemory. They appeared to monitor memory’s
contents to decline tests of weaker memories. This memory-strength signal—abstract, cognitive and non-associative—is profoundly different from the stimulus signal available in traditional operant situations. There are converging metamemory results [18,30]. In one case [18], macaques proved able to adaptively decline memory tests of the most difficult serial positions in lists of to-be-remembered items (figure 3a).

In this experiment, they showed an additional kind of cognitive self-regulation. That is, as the metamemory task was made more difficult, macaques held their error rate near 10 per cent (figure 3b) by responding uncertain more in difficult task conditions (figure 3c). Thus, macaques accepted memory tests if they were 90 per cent certain of remembering. Finally, researchers have also assessed macaques’ metamemory by using trans-cranial magnetic stimulation (TMS) to interfere with visual working memory [39]. TMS interfered with a macaque’s matching-to-sample performance and also increased his uncertainty responding. It was not just some global TMS effect that caused this increase. The effect of TMS was hemisphere specific. That is, TMS caused an increase in uncertainty responding only when it occurred contralaterally to the presentation of the sample in the visual field (when it

Figure 2. (a) Performance by monkeys in a same–different discrimination [15]. The monkeys manipulated joysticks to make different or same responses, respectively, when two pixel boxes had the same or different internal density of lit pixels. In addition, they could make an uncertainty response to decline the trial. The pitch of the density difference on different trials was adjusted dynamically to titrate the monkeys’ perceptual limit for distinguishing sameness from difference and to examine their response patterns in detail within this region of maximum difficulty. The horizontal axis gives the ratio between the densities of the two pixel boxes seen on each trial. The same response was correct for ratio 1—these trials are plotted as the rightmost data point for each curve. The different response was correct for all other trials. The solid line represents the percentage of trials receiving the uncertainty response at each density ratio. The percentages of trials ending with the different response (dotted line) or same response (dashed line) are also shown. (b) Performance by humans in the same–different discrimination, depicted in the same way. Reprinted with permission from Shields et al. [15, p. 158]. Copyright © 1997 by the American Psychological Association.

Figure 3. (a) Performance by a macaque in a metamemory task [18]. NT denotes ‘not there’ trials in which the probe picture was not in the memory list of pictures. The serial position (1–4) of the probe picture in the list of pictures on ‘there’ trials is also given along the x-axis. The percentage of each type of trial that received the uncertainty response is shown (solid line with squares). The percentage correct (of trials on which the memory test was attempted) is also shown (dashed line with circles). Macaques responded ‘uncertain’ most for the trials on which their memories were most indeterminate. (b) Percentage error rates by two monkeys (black and grey bars) when the difficulty of the memory test was increased by increasing the memory list from two to six pictures. (c) Percentage uncertainty responses (URs) by two monkeys when the difficulty of the memory test was increased in the same way. Reprinted with permission from Smith et al. [18, p. 236, p. 238]. Copyright © 1998 by the American Psychological Association.
also probably maximally interfered with registering and remembering the to-be-remembered sample).

By dissociating difficulty and uncertainty from concrete stimuli, and thereby protecting against stimulus-based interpretations, all of these metamemory studies required animals to monitor uncertainty on a more abstract and cognitive level, producing results that the stimulus-based associative concern cannot explain.

(c) Entrainment to reinforcement contingencies

A third concern centred on trial-by-trial feedback as always provided by the early paradigms. Every outcome could be associated with the stimulus–response pair that produced it. As also suggested by Hampton [29], animals might have been conditioned to make uncertainty responses when facing the trials that were associated with the worst stored reinforcement history.

To address this concern about entrained reinforcement gradients, researchers replaced trial-by-trial feedback with deferred feedback whereby animals worked for blocks of trials before receiving a performance evaluation [19,26]. Moreover, in that evaluation, rewards and timeouts were bundled separately, so that the temporal sequence of trials completed and outcomes obtained bore no relation. This defeated the normal processes of association and conditioning. Animals could not know which trials had gone unreinforced through a clear and immediate feedback signal, and so, based on the objective feedback of the task itself, they could not know which trials they had missed.

By this technique, researchers uncoupled objective performance from subjective difficulty. Animals had to set decision criteria and define response regions on their own—cognitively and decisionally. Yet, animals still made uncertainty responses proactively and adaptively under these circumstances. Emphasizing this uncoupling, the study by Smith et al. [19, p. 292, fig. 8a,b] showed that there was no relationship between the proportion of uncertainty responses and the proportion of (in)correct responses across trial levels, as there would have to be if uncertainty responses were conditioned avoidance responses. Instead, there was a strong relationship between the proportion of uncertainty responses and the distance of the trial level from the animal’s decisional breakpoint in the discrimination [19, p. 292, fig. 8c,d], as there would be if the animal were monitoring difficulty or uncertainty.

These experiments dissociated for the first time strategies based in reinforcement history from strategies based in decisional difficulty. Animals can monitor uncertainty adaptively using the latter strategy. Therefore, the reinforcement-based concern about the uncertainty response is not fully justified. More generally, the deferred-feedback technique has broad potential applicability within comparative psychology. It forces animals to self-construe the task and to self-construct a task approach, and thus it provides a more cognitive read on their behaviour.

(d) Representational specificity

Other research has tested the representational rigidity or narrowness of animals’ uncertainty responses, expected if these responses are low-level behavioural reactions. For example, researchers [35] asked macaques to monitor uncertainty while multi-tasking. Four different difficult discriminations were randomly intermixed trial by trial. Despite this multi-tasking requirement, macaques were able to decline the difficult trials across domains. This shows that uncertainty responses are not reactive to just one well-trained trial type at a time. This kind of simultaneous transfer test suggests that uncertainty responses result from a general psychological signal that transcends a single task and perhaps is similar to the general psychological state of uncertainty that humans would bring to the same collection of discriminations.

Washburn et al. [38] tested another form of generalization in macaques’ metamcognitive performances, by asking whether they would respond uncertain adaptively on a novel task’s first trial. The researchers adapted the learning-set paradigm [50], in which a new two-choice discrimination began every six trials. Macaques responded uncertain far more often on trial 1 of each problem than on trials 2–6, consistent with the fact that they could not know the answer on trial 1 but could know the answer on trials 2–6 (in this experiment, the uncertainty response revealed each discrimination’s answer but gave no appetitive reward). This rapid, flexible application of the uncertainty response to new discrimination problems also strongly discourages associative interpretations and encourages metacognitive interpretations of the uncertainty response. It illustrates again the utility of using transfer to show the generality and flexibility with which some species use uncertainty-monitoring processes.

(e) Phylogenetic restrictions in uncertainty monitoring

Cross-species research also undermines lower-level, associative interpretations of the uncertainty response, sometimes through the failure of animals to respond adaptively. Capuchin monkeys (Cebus apella) represent another major primate lineage (the New World primates). Researchers [23] tested capuchins’ uncertainty monitoring along a sparse–to-dense perceptual continuum with the difficult and uncertain trials surrounding the discrimination’s breakpoint. Strikingly, capuchins did not respond uncertain, though this sharply reduced their reward efficiency. A similar result was obtained when the error timeout was increased to 90 s, so that with each error capuchins potentially forfeited 30 trials and 30 food rewards (figure 4a).

In other sessions, capuchins performed a sparse–middle–dense task in which they could earn rewards or timeouts for (in)correctly making middle responses. Capuchins responded middle easily from the beginning of testing, in sharp contrast to their negligible uncertainty responding (figure 4b).

These two tasks were structured similarly—indeed, the same intermediate stimuli should have recruited middle and uncertain responses. Thus, the two tasks—strong mutual controls—produced a striking dissociation. The capuchins easily brought middle responses, but not uncertain responses—under the control of those intermediate stimuli.

Phil. Trans. R. Soc. B (2012)
Capuchins are such apt learners that they are known as the poor person’s chimpanzee. If uncertainty responses were triggered by conflict, aversion, avoidance, fear, competing response strengths, reward maximization, hesitation-wavering behaviours, hesitation-wavering latencies or any other first-order cue, capuchins would have used that cue to prompt adaptive uncertainty responses. Clearly, the mechanisms that underlie middle responding and uncertainty responding are different psychologically. And clearly, the uncertainty-monitoring capacities of capuchins and macaques are different as well, a conclusion that has been reached independently [21, 28, 31].

3. INTERIM CONCLUSION
The results considered in §2 have produced a growing consensus that some species have shown metacognition. ‘Metamemory, the ability to report on memory strength, is clearly established in rhesus macaques (Macaca mulatta) by converging evidence from several paradigms’ [37, p. 266]. ‘Evidence for metacognition by nonhuman primates has been obtained in great apes and old world monkeys’ [28, p. 575]. ‘Substantial evidence from several laboratories converges on the conclusion that rhesus monkeys show metacognition in experiments that require behavioral responses to cues that act as feeling of knowing and memory confidence judgments’ [32, p. 130].

This debate has shown some of comparative psychology’s best practices, including interpretative conservatism, incisive criticism, testable low-level interpretations, disconfirmed low-level interpretations, and rapid empirical and theoretical progress towards a consensual conclusion.

4. POOR INTERPRETATIVE PRACTICES IN ANIMAL-METACOGNITION RESEARCH
However, not all the assertions of low-level processes have been disciplined and principled. There have been misunderstandings, shallow descriptive accounts and misapplications of Morgan’s canon. In this section, we discuss these poorer practices within this area of comparative psychology.

(a) A misconception about formal models
Researchers commonly use signal-detection models to describe animals’ metacognitive performances [20, 34, 43, 52]. A misconception surrounds these models to which comparative psychologists should attend. The misconception is that if a formal model fits behavioural data, then one can and should interpret the data in a low-level, associative manner [53].

This supposition lacks a scientific basis. In signal-detection models, the parameters, decision criteria and response regions are defined purely mathematically. These models do not specify cognitive representations, cognitive processes, levels of awareness or brain regions. The elements of the models are psychologically empty because they are purely mathematical. They cannot imply a low-level information-processing description: they imply no information-processing description.

It is a problem that one can be led by a model’s simple mathematics to assume that it reflects simple psychological processes. There is no correlation of this kind. Signal-detection models would fit humans’ metacognitive data perfectly well, even though humans often complete uncertainty-monitoring tasks using fully conscious cognition.

The broader implication is that it is not principled behaviourism to say that a model explains animals’ behaviour. Instead, we must reckon with the processes and representations that underlie the behaviour, including their level in the animal’s cognitive system and in its awareness.

(b) A misconception about reinforcement’s benefits
Another misconception is that one can explain animals’ uncertainty responses by saying that animals make them to reduce the time to the next reward. On this view, uncertainty responses are inherently low level and associative because they are about reward maximization.

However, the reinforcement-maximization hypothesis is also psychologically empty. Though animals (and humans) may try to maximize rewards, the psychological

Figure 4. (a) The performance of capuchin monkeys in a sparse–uncertain–dense task [23]. The horizontal axis indicates the density level of the box. The sparse and dense responses, respectively, were correct for boxes at density levels 1–21 and 22–42. The solid line represents the percentage of trials receiving uncertainty responses at each trial level. The percentages of trials ending with the sparse response (dotted line) or dense response (dashed line) are also shown. (b) The performance of the same capuchin monkeys in the sparse–middle–dense task [23], depicted in a similar way. From Smith et al. [51, p. 48].
question is how they do so. Reward-maximization processes are not necessarily low level. They could sometimes be linked to meta-level processes and representations [14]. Even a human’s conscious, declarative metacognitive behaviours are compatible with reward maximization. Therefore, reward maximization cannot point to a low-level, information-processing description: it points to no information-processing description.

It is also a problem that one can be led by the simple premise of reward maximization to mistakenly assume that it reflects low-level processes. There is no correlation of this kind. In fact, we saw in §2 that low-level reward maximization—using traditional stimulus and reinforcement cues—does not fit the data.

High-level reward maximization might explain the data, if animals choose to complete easy trials that bring immediate reward and decline difficult trials they believe could produce error and reinforcement delay. Only in this way will they avoid difficult trials (speeding reinforcement) without avoiding easy trials (slowing reinforcement). But difficulty monitoring is a higher level, metacognitive process. Thus, reward maximization using the uncertainty response is not evidence against metacognition. To the contrary, it shows the animal using its metacognitive understanding productively.

The broader implication is that it is not principled behaviourism to say that animals make responses because they have a benefit. Instead, one must describe how the animal gets to the benefit psychologically—how its mind produces the benefit.

(c) A misconception about present stimuli

A third misconception is that uncertainty tasks can only reflect metacognition when humans or animals respond with the relevant stimuli absent. The idea is that stimulus absence prevents organisms from responding directly to stimulus properties, and ensures that they must (if they can) represent their mental states in some way and make a judgement based on those.

The animal-metacognition literature transcends this idea factually. Macaques have performed both stimulus-present and stimulus-absent metamemory tasks [18,44]. The results converged strongly, and modelling showed that macaques in those studies had the same memory-strength criterion for choosing to complete memory trials [20]. Metacognitive judgements unfold some way and make a judgement based on those.

The animal-metacognition literature transcends this idea factually. Macaques have performed both stimulus-present and stimulus-absent metamemory tasks [18,44]. The results converged strongly, and modelling showed that macaques in those studies had the same memory-strength criterion for choosing to complete memory trials [20]. Metacognitive judgements unfold some way and make a judgement based on those.

On reflection, this convergence will be intuitive. As a student considers a multiple-choice test question, the question and response alternatives are fully visible. But he or she will still make metacognitive assessments (Where in the book was that material? Is [b] a lure? Do I know this or should I skip on to use time better? Should I change major?). Present stimuli do not dampen metacognition [54]. Indeed, the mind could be freed towards more efficient metacognition when it is not occupied with stimulus maintenance.

It is another problem that one can be led by present stimuli to assume that the presence of those stimuli implies low-level psychological processes. There is no correlation of this kind. The broader implication is that it is not principled behaviourism to suppose that a stimulus-present task is low-level and associative. By doing so, one decides the issue using an associative bias that lacks a scientific basis. Instead, one must find out how the animal thinks about the present stimuli; and the level that thinking occupies in its cognitive system. One’s preferences for associative explanations cannot provide those answers, but science may provide them.

Low-level/high-level disputes especially occur, in the animal-metacognition literature and elsewhere, when the facts are mixed and there is a temptation to tie-break using a theoretical preference. However, mixed data patterns especially deserve no strong interpretation. In these situations, there is a constructive place for agnostic silence while the facts accumulate and the empirical reality asserts itself. We can wait and see, while we document animals’ capacities: here is what they do and fail to do.

Reading the animal-metacognition literature in this way is illuminating. Adaptive uncertainty responses can be independent of stimuli and reinforcement. They are used flexibly, during metacognitive multitasking, and even selectively on the first trial of novel tasks while animals discover what to do. They are cognitive and decisional, available alike for difficult same—different judgements and memory reports. It is an extraordinary set of performances, no matter the interpretative lens one views it through.

Even Morgan [42, p. 59] may have expressed agnosticism. He said: ‘In no case is an animal activity to be interpreted as the outcome of the exercise of a higher psychical faculty, if it can be fairly interpreted as the outcome of the exercise of one which stands lower in the psychological scale’. He also said: ‘it is clear that any animal may be at a stage where certain higher faculties have not yet been evolved from their lower precursors; and hence we are logically bound not to assume the existence of these higher faculties until good reasons shall have been shown for such existence’. He does urge caution in making the high-level attribution. But he neither strongly denies the high-level capacity nor strongly asserts the low-level explanation. What he says is consistent with applying a razor of silence—wait and see—rather than a razor of denial [55].

(d) Selective criticism

An additional problem is that some pursue low-level interpretations of metacognitive performances by being selective—they only discuss the paradigms that allow the criticism. This practice disregards a growing empirical literature.

Disciplined theoretical interpretations must treat the totality of the research findings. Research findings bootstrap off of one another. Some early findings may have had a potential low-level interpretation, as we have discussed. But once later findings address the issue, the original study can regain some lustre, because now the parsimonious explanation might be that the older finding also deserved a high-level interpretation. If the species is metacognitive in one task, it is not parsimonious to suppose that it became qualitatively non-metacognitive in a closely related task.

In effect, the later task fulfils Morgan’s addendum to the canon, which is extremely important though
often overlooked. He said [42, p. 59]: ‘To this, however, it should be added, lest the range of the principle be misunderstood, that the canon by no means excludes the interpretation of a particular activity in terms of the higher processes, if we already have independent evidence of the occurrence of these higher processes in the animal under observation’. The later study provides the independent evidence, and it should affect one’s theoretical interpretation of the earlier study. A philosophical analysis of Morgan’s canon [56] reached a similar conclusion.

(e) Shopping associative mechanisms
An additional concern is that there has been a kind of associative musical chairs in the field of animal metacognition. By turns, stimulus aversion/avoidance, reinforcement history, reward maximization and other associative explanations have been asserted. It betrays a bias to give associative theory many bites of the apple like this, because it reveals an insistence to find a low-level interpretation. This problem becomes worse, as we have discussed, when the associative hypotheses become non-psychological and less principled. It also becomes worse when different associative mechanisms are used to explain different individual findings (i.e. task A and B, respectively, need associative interpretations X and Y). It is not parsimonious to depend on multiple low-level descriptions of animals’ performance across tasks, when a basic metacognitive process explains everything simply and naturally, using an adaptive capacity with which cognitive evolution would likely have endowed some species.

(f) Implausible dualism
Finally, it is an implausible scientific dualism that humans and animals are qualitatively metacognitive and associative, respectively [57]. In no other instance, be it younger versus older human children, or younger versus older human adults, in which the groups’ performance profiles correlated at 0.97 (figure 2) [15], would one think to offer qualitatively different low- and high-level interpretations of the behavioural data. Instead, one would naturally interpret similar performances similarly. Thus, our literature provides a case wherein the weight of parsimony has shifted towards the metacognitive interpretation, and the burden of proof has shifted towards associative theorists to demonstrate the necessity for, and the sufficiency of, low-level cues in supporting animals’ uncertainty performances. That burden has not been met.

Remember, too, that macaques and humans share evolutionary histories, homologous brain structures and so forth. This also makes it implausible that humans would produce their highly similar graph in a qualitatively different way. As De Waal [58, p. 316] said: ‘the most parsimonious assumption concerning nonhuman primates is that if their behavior resembles human behavior the psychological and mental processes are similar’. Though surely there are limits on this application of evolutionary parsimony, it is at least likely that, if humans perform metacognitively, this provides some evidence that non-human primates use similar information-processing mechanisms [59].

One could argue that metacognition had no evolutionary depth, no phases in its development, and no antecedents. To the contrary, we suppose that metacognition had some evolutionary course of development. This predicts psychological continuities between macaques and humans in this capacity, just as we know there are biological continuities. However, this does not mean that macaques must have every conscious and self-aware facet of humans’ metacognition. Those facets could have been the add-ons of human evolution as the metacognitive capacity matured and flowered. These are important remaining theoretical questions for the field to explore.

5. DETECTING A METACOGNITIVE SIGNAL WITHIN ANIMAL MINDS
Figure 5 summarizes the theoretical situation in the animal-metacognition literature from the perspective of signal-detection theory. The signal-detection framework is apt because we are evaluating whether research has yet detected the signal from animal minds of a higher-level cognitive capacity called metacognition. Animals’ true metacognition (figure 5a)—which we cannot see into their minds to directly confirm—will produce a distribution of cognitive performances across paradigms that appear sometimes more or less cognitively sophisticated. So will their true associative capacity (figure 5b). But the appearances presented by these capacities may overlap, creating a difficult interpretative problem. Therefore, through this decision space, behavioural analysts place a decision criterion or theoretical dividing line. Performances above the line meet the theoretical grade and are interpreted as metacognitive. Performances below the line do not and are deemed to reflect lower-level, associative processes.

Within this decision space, there are four possible interpretative outcomes. Hits occur to the right in figure 5a when the scientist correctly concludes for metacognition. Correct rejections occur to the left in figure 5b when the scientist correctly concludes for an associative mechanism. Hits and correct rejections are salutary scientific events. Misses occur to the left in figure 5a when the scientist wrongly concludes against metacognition. In the illustration, about 75 per cent of true metacognitive performances by animals would be interpreted away. False alarms occur to the right in figure 5b when the scientist wrongly concludes for metacognition. In the illustration, about 3 per cent of all associative performances by animals would falsely be called metacognitive. Misses and false alarms are infelicitous scientific events.

The decisional stance within the animal-metacognition literature, as with all domains of comparative psychology, followed Morgan’s interpretative lead. His canon was designed to counter the anecdotal reports of animal intelligence and the introspective methods of mental attribution that had produced an anthropomorphic bias. This interpretative lead yielded a distinctive and familiar theoretical culture. Our literature set the high decision criterion shown in figure 5 for accepting the presence of metacognition, so that few animal performances exceeded it.
There was skepticism about animal metacognition. There was a lack of agnosticism. The preference for associative interpretations was used to break interpretative ties.

One sees this culture operating when mathematical parameters are deemed to reflect low-level processes, and when casual reinforcement-maximization hypotheses—not rooted in a concrete information-processing description—are mistaken for low-level processes. One sees it operating when stimulus-based tasks are automatically assumed to elicit reactive processes, and when behavioural analysts persistently try out different low-level mechanisms, taking multiple bites from the associative apple.

In §§2 and 3 of this article, we praised some elements of this decisional stance within our field—it produced strong theoretical progress. In §4, we pointed out some flawed aspects of this scientific culture. Now we consider some structural weaknesses within this scientific culture that produce a complementary interpretative bias to the old, anthropomorphic bias.

First, the high-threshold stance of the animal-metacognition literature guarantees that we will detect less accurately the true psychological signals issuing from animal minds. The highest level of overall correct responding in a detection experiment comes from a criterion level that is midway between the extremes, not at an extreme position. The criterion chosen in the animal-metacognition literature could possibly be expected to double the incorrect scientific conclusions reached in our area. This is a serious problem that has rarely been noted in discussions of scientific inference within the animal-metacognition literature or within comparative psychology more broadly.

Second, the high-criterion stance copes poorly with the fact that the two types of interpretative error inexorably trade off with one another. The more avoind we are of false alarms, creating a high bar for a metacognitive theoretical interpretation, the more misses we will experience because many actual metacognitive performances will not clear the high-bar inferential hurdle.

One can roughly quantify this tradeoff using figure 5. In figure 5b, as the decisional threshold moves rightward near its position in the figure, one
continues to avoid a tiny set of additional false alarms, as one correctly puts more of the vanishing tail of the associative distribution to the left of the criterion. But, in exchange, in figure 5a, one moves the line through the probability-dense centre of that Gaussian distribution. As a result, one dispaces large numbers of true metacognitive events to the left of the decision criterion, and thus we would interpret away these metacognitive performances and increase the number of misses. There would be a many-to-one exchange of misses incurred to false alarms avoided. This is a lousy tradeoff. This would only be acceptable scientific practice if we were for some reason many times more accepting of misses than of false alarms. Of course, comparative psychologists have historically been asymmetrically avoidant of false alarms.

The third problem is that the decisional stance in our literature has caused misses to receive less attention. Indeed, rarely does anyone warn of the dangers of misses in the scientific interpretation of animals’ uncertainty performances (or their performances in other cognitive domains). We issue that warning here. It is easy to specify in detail the dangers of misses, and we believe these dangers are serious.

Misses in animal-metacognition research are as wrong as false alarm interpretations. Misses create artificial discontinuities between human and animal minds. Misses may cause us to underestimate the experience of pain and suffering by animals and threaten the ethical conduct of animal studies. These artificial discontinuities can blind us to the origins of human capacities and to their emergence during phylogeny. Misses make it seem that animal models have no place in studying human capacities, because animal minds are qualitatively different and low-level. As a result, misses downgrade the relevance of animal research. They downgrade its fundability, too. The more that animals are qualitatively different, the less that animal studies have to contribute to issues of humans’ mental health and psychological functioning. Misses also make animal research less accessible and interesting to the wider academic community. They isolate comparative science and reduce its societal impact and footprint. These isolating effects have become increasingly clear and problematic over the last 20 years. Indeed, one can see that some elegant fields of comparative study have retreated—like beautiful glaciers—to the higher elevations of behavioural analysis where few seek access.

6. A MIDDLE GROUND FOR INTERPRETATION IN ANIMAL-METACOGNITION RESEARCH

In contrast, the animal-metacognition literature is struggling to achieve a middle ground of theoretical interpretation. We will end by describing this middle ground as it presents itself in our field.

First, research encourages the conclusion that animals’ uncertainty responses are qualitatively different psychologically from their primary perceptual responses (e.g. the dolphin’s low/high responses). Acknowledging this difference seems necessary to explain the dissociation between uncertain and middle responses [23] and the differences in uncertainty responding between macaques and capuchins [24]. Notice, though: qualitatively different does not mean qualitatively conscious, self-aware and so forth.

Second, research encourages the conclusion that animals’ uncertainty responses are not appropriately considered associative responses in the traditional sense. They can be independent from stimuli, reinforcement and so forth. Uncertainty responses should probably be elevated interpretatively to the level of cognitive-decisional processes in animals’ minds.

Third, the research also shows that uncertainty responses serve animals on the first trial of novel tasks, in abstract tasks, in metamemory tasks and even during multi-tasking. They are sometimes used with so much flexibility and agility that they appear to have some continuities with humans’ explicit and declarative cognitive processes—that is, those processes that let humans’ reflective minds turn on a dime. Notice: the presence of some continuities does not imply the presence of all continuities.

In conceptualizing these continuities psychologically, some have followed Shiffrin & Schneider [60] by noting that uncertainty-monitoring tasks are inconsistently mapped. That is, they feature indeterminate mental representations that map unreliably onto responses so that those representations become inadequate guides to behaviour. The animal must engage higher levels of controlled cognitive processes to adjudicate the indeterminacy and choose adaptive behaviours. One might say that uncertainty responses represent controlled decisions, at the limit of perception or memory, to decline difficult trials. This is a careful statement of our field’s middle ground. It grants animals’ uncertainty responses some deserved cognitive sophistication, without attributing to them all the sophisticated features that human metacognition can show.

Thus, the animal-metacognition literature is seeking a moderate decisional stance by which it avoids extreme interpretations depending on either low-level associations or florid, conscious metacognition. Instead, it has characterized in more specific and more accurate information-processing terms the mental representations and cognitive processes that underlie animals’ metacognitive performances.

This stance confers benefits on the research area, by increasing the theoretical and empirical scope granted to researchers. For one example, researchers are now freed to try to pinpoint the nature and level of the controlled uncertainty processes that animals show. To this end, our laboratory is asking whether uncertainty responses reflect an executive cognitive utility that especially requires attentional resources or working-memory capacity. We are also asking whether macaques can experience sudden realizations of knowing. Through empirical approaches of this kind, one can continue to cautiously elevate one’s interpretation of the uncertainty response, showing that it is higher-level, attentional, executive and even perhaps conscious.

These approaches were held in abeyance in the early years of animal-metacognition research, when the focus was on the associative content of uncertainty responses. A high decisional threshold is not just a
quantitative threshold by which a field expresses its general conservatism. To the contrary, the decisional threshold qualitatively changes the nature of the theoretical discussion and affects the kind of research questions that are naturally asked. A field has more scope to ask diverse empirical questions given a more temperate theoretical climate.

Likewise, our field has gained more scope to consider human and animal metacognitive capacities in relation to one another. What are the benefits and affordances of language-based metacognition that is propositionally encoded and in which animals cannot share? What is the essential nature of metacognition that can occur without language and propositions? Do humans feel like uncertain selves in metacognition tasks in ways that monkeys do not? Why, when, and how did conscious cognitive regulation come to play a substantial role within humans’ cognitive system? These questions open up as one honours the homologies in the uncertainty-monitoring performances of humans and animals. In these homologies, one also clearly sees the value of animal models for human metacognition, and the possibility of searching for biochemical blocks that might be removed and biochemical enhancers that might be applied to improve metacognitive regulation.

The animal-metacognition literature has also begun to instantiate the distinctive theoretical premise that metacognition is not all-or-none. A great deal of energy has been spent debating the qualitative choice: do animals have metacognition or are they being associative? However, there is a constructive theoretical middle ground wherein one grants organisms a basic uncertainty-monitoring capacity without over-interpreting what they do. In this middle ground may lie the phylogenetic emergence of human metacognition and the ontogenetic emergence of metacognition in human development.

This perspective grants animal-metacognition research strong links to and implications for metacognition research in human development. The behavioural animal paradigms expand the range of metacognition paradigms available for testing young human children. Using them, researchers may uncover the earliest developmental roots of human metacognition [40]. The animal paradigms can also be used to explore the metacognitive capacities of language-delayed and autistic children, or children with mental retardation. It is an important possibility that there might be more basic forms of cognitive regulation (more implicit; less language-based) that could be preserved or fostered in children who are challenged in the highest-level aspects of metacognition.

Thus, one sees that a middle ground of theoretical interpretation is not the compromise of weakness. It balances our field better between the two types of inferential errors. It grants the field more theoretical scope. It opens new lines of research, concerning psychological content, consciousness, human origins, language affordances and so forth. It broadens our field, granting it outreach to issues of human development and psychological well-being. It makes the research in animal metacognition accessible to a wider range of interested but non-expert consumers of science in the public domain. And, remarkably, the empirical picture in our field makes plain that, in addition, we are now reading more accurately than ever the uncertain signals emanating from animal minds.

The preparation of this article was supported by grant 1R01HD061455 from NICHD and grant BCS-0956993 from NSF.

REFERENCES
3 Koriat, A. 1993 How do we know that we know? The accessibility model of the feeling of knowing. Psychol. Rev. 100, 609–639. (doi:10.1037/0033-295X.100.4.609)


