

# Decision deadlines and uncertainty monitoring: the effect of time constraints on uncertainty and perceptual responses

Alexandria C. Zakrzewski · Mariana V. C. Coutinho · Joseph Boomer · Barbara A. Church · J. David Smith

Published online: 26 September 2013  
© Psychonomic Society, Inc. 2013

**Abstract** The behavioral uncertainty response has grounded the study of animal metacognition and influenced the study of human psychophysics. However, the interpretation of this response is debated—especially whether it is a behavioral index of metacognition. The authors advanced this interpretation using the dissociative technique of response deadlines. Uncertainty responding, if it is higher level or metacognitive, should depend on a slower, more controlled decisional process and be more vulnerable to time constraints. Humans performed sparse–uncertain–dense or sparse–middle–dense discriminations in which, respectively, they could decline difficult trials or positively identify middle stimuli. Uncertainty responses were sharply and selectively reduced under a decision deadline, as compared to primary perceptual responses (i.e., “sparse,” “middle,” and “dense” responses). This dissociation suggests that the uncertainty response does reflect a higher-level, decisional response. It grants the uncertainty response a distinctive psychological role in its task and encourages an interpretation of this response as an elemental behavioral index of uncertainty that deserves continuing research.

**Keywords** Metacognition · Uncertainty monitoring · Decision making · Response deadlines · Controlled processing

**Electronic supplementary material** The online version of this article (doi:10.3758/s13423-013-0521-1) contains supplementary material, which is available to authorized users.

A. C. Zakrzewski · M. V. C. Coutinho · J. Boomer · B. A. Church · J. D. Smith  
University at Buffalo, State University of New York, Buffalo, NY, USA

A. C. Zakrzewski (✉)  
Department of Psychology, University at Buffalo, Park Hall, Buffalo, NY 14260, USA  
e-mail: azakrzew@buffalo.edu

Humans’ metacognitive responses—uncertainty judgments, feelings of knowing, and so forth—ground the metacognition literature (e.g., Dunlosky & Bjork, 2008; Koriat & Goldsmith, 1996; Schwartz, 1994). Achieving a theoretical understanding of metacognition is important because metacognition is important to humans’ learning, thinking, and comprehension.

Toward achieving that understanding, researchers have begun to measure basic, behavioral forms of metacognition. For example, in the influential perceptual uncertainty-monitoring task, participants use categorization responses to place stimuli into perceptual categories. But they also have a response that lets them decline to complete any trials they choose, on the basis of the trial’s difficulty or of their uncertainty. Participants use this uncertainty response (UR) to decline difficult, potentially error-producing trials.

Illustrating this task, Balcomb and Gerken (2008) gave 3.5-year-old children paired-associate tests and the UR. These children would fail traditional metacognitive assessments, but they made URs to cope with uncertainty. They also performed poorly in later tests of items that they declined, indicating their responses to be a valid, internal cue of faint memory. Balcomb and Gerken concluded that young children have implicit access to internal knowledge states and that behavioral paradigms best reveal that access. Research like this could reveal the earliest developmental roots of human metacognition.

Similarly, comparative psychologists have asked whether other species have a functional analogue to human metacognition (reviewed in Smith, Beran, & Couchman, 2012). Macaques make URs to decline difficult memory trials (e.g., Hampton, 2001) and perceptual-classification trials (e.g., Smith, Coutinho, Church & Beran, 2013b). Thus, some primate species may share with humans a metacognitive capacity. Research like this could reveal the earliest evolutionary roots of human metacognition. It could also provide animal models for metacognition and suggest behavioral approaches

to foster metacognitive capacities in child populations that are developmentally or language delayed (e.g., Ruffman, 2000). Thus, these behavioral paradigms have applications for metacognition research and practice broadly, depending on whether the UR can be considered an elemental, behavioral index of metacognitive functioning.

This issue has been controversial, given comparative psychology's interpretative conservatism. Animals' URs, although seeming metacognitive, could nonetheless be stimulus- and reinforcement-based reactions to middling, indeterminate stimuli along a continuum. The possible low-level basis for URs has been the principal theoretical concern about cross-species metacognition research (e.g., Hampton, 2009; Jozefowicz, Staddon, & Cerutti, 2009; Smith, Beran, Redford, & Washburn, 2006; Smith et al., 2012). The psychological interpretation of URs remains sharply debated (Le Pelley, 2012; Smith, Couchman & Beran, 2013a). The present research clarifies that interpretation.

In early research on human psychophysics, the possible high-level basis for URs was an issue. Indeed, some believed that URs should be disallowed during psychophysical tasks because they were psychologically distinctive and possibly metacognitive reports of conscious uncertainty (Watson, Kellogg, Kawanishi, & Lucas, 1973, pp. 184–185). Boring (1921, p. 445) called the UR an “attitudinal seducer,” because he thought it would distract participants away from the necessary psychophysical attitude, in which they approached the task purely perceptually.

What is the correct interpretation? Are URs just another perceptual response to middling stimuli? Are they a higher-level cognitive assessment of discrimination failure? Are they an elemental index of metacognition, justifying their extension to other populations and species? It is scientifically casual just to interpret animals down, but humans up, as so often happens. It may also be unparsimonious. The present article models an empirical approach that allows for cross-species judgments about the cognitive level of URs in a way that could put different species on the same interpretative playing field.

We address these questions using a dissociative framework. If URs result from higher-level metacognitive processes, they should betray that information-processing character if used under stress. We chose the stressor of speeded responding. Participants performed a sparse–uncertain–dense (SUD) task in which they identified stimuli as being sparse or dense or responded “uncertain” for trials deemed too difficult, or a sparse–middle–dense (SMD) task in which they identified stimuli as sparse, middle, or dense. If URs—as compared to perceptual responses like “sparse,” “middle,” and “dense”—are more decisional and time constrained, then speeded responding should selectively undermine the UR's use. Evaluating this possibility was our primary empirical goal. SUD and SMD tasks have grounded several important animal metacognition articles (e.g., Beran, Smith, Coutinho, Couchman, & Boomer, 2009; Smith et al., 2006; Smith

et al., 2013b), so the psychological light thrown on them here is important in thinking about animals' performances, too.

This evaluation is also important because little is known about whether humans' online metacognitive processes—including URs—have a higher-level, reflective psychological character (though their *post-hoc* metacognitive justifications do). One needs information-processing benchmarks of reflective cognition in order to make this assessment. If URs are low-level, reactive responses, then we need not attribute metacognition to animals, young children, or human adults in uncertainty tasks. But if URs show distinctive benchmarks, the interpretative ground shifts. Then we have new ways to confirm metacognition in nonverbal human populations and animal populations, too, if they show these processing benchmarks. Thus we could illuminate the emergence of reflective mind during primate evolution and human development.

Our approach also offers to comparative psychology a transformative escape from a theoretical impasse. Some still label all animals' uncertainty performance “associative” (i.e., based in reinforcement and stimulus reactions—Jozefowicz et al., 2009; Le Pelley, 2012), which prevents researchers from distinguishing performances that may have different psychological characters (Smith et al., 2013a). Our approach using information-processing benchmarks offers principled ways to make meaningful distinctions among performances, fostering comparative psychology's ongoing theoretical development.

## Method

### Participants

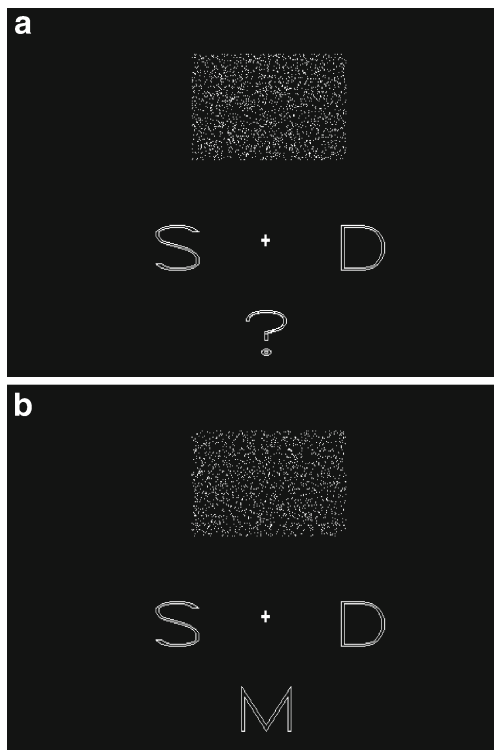
A group of 60 undergraduates—with apparently normal or corrected-to-normal vision—participated to fulfill a course requirement. To increase motivation, the top scorer received a \$10 cash prize.

### Density continuum

On each trial, a  $201 \times 101$  pixel unframed stimulus rectangle was presented at the computer screen's top center (Fig. 1). The rectangle contained varying numbers of randomly placed lit pixels. We used 42 stimulus levels, Levels 1–42 (1,085–2,255 pixels). Each level contained 1.8 % more pixels than the last, making the continuum logarithmic.

### Response modality

Three response icons were also presented on each trial. Responses were made by pressing one of three adjacent keyboard keys, arranged in the same left-to-right spatial order as the screen's icons. Adjacent keys allowed participants to respond rapidly during speeded trials.



**Fig. 1** The trial screens from the sparse–uncertain–dense task (a) and the sparse–middle–dense task (b) described in the text

### Stimulus distribution

The stimulus distributions for the SUD and SMD tasks were identical. In each task, one third of presented trials were at Levels 1–18 (1,085–1,470 pixels), Levels 19–24 (1,496–1,636 pixels), and Levels 25–42 (1,665–2,255 pixels).

### Sparse–uncertain–dense task

Levels 1–21 (1,085–1,550 pixels) and Levels 22–42 (1,578–2,255 pixels) were defined and reinforced to be the *sparse* and *dense* trials. Oversampling the difficult areas of the stimulus continuum (Levels 19–24) let us increase difficulty and uncertainty. Participants responded “S” and “D” in order to classify stimuli as sparse and dense, or made URs (“?” icon on screen) to decline the trial. For correct and incorrect responses, respectively, participants gained 1 point and heard a 1-s beep, or lost 3 points and heard an 8-s buzz as a penalty timeout. The UR produced no sound or outcome, but simply advanced the participant to the next trial. Following response, a white number representing total points appeared on the screen, with a green +1 for a correct response, a red –3 for an incorrect response, or a blue ? for a UR. The next trial followed.

Ideally, participants might just sharpen their perceptual sensitivity, so that they never erred and never made URs. But of course this would not happen: They would misperceive and they would err. They would need to optimize strategically,

since they could not perceive ideally, and therefore they would apply the UR most to the difficult trials near the sparse–dense discrimination breakpoint (Levels 21–22).

### Sparse–middle–dense task

Levels 1–18, 19–24, and 25–42 were defined and reinforced to be *sparse*, *middle*, and *dense* trials that deserved the “S,” “M,” or “D” response. The feedback was as we just described. No UR option was now available—it was replaced by the middle response (MR). All trials received correct/rewarding or incorrect/penalty kfeedback.

### Psychophysical control and matching

The SUD and SMD tasks were matched as follows. We could not control how many stimulus levels might make participants feel uncertain and prompt URs. The participants made this determination subjectively. However, from formal modeling in other studies, it is known that participants’ uncertainty regions span a narrow range of about six steps in an SUD task. Accordingly, we made the middle region in our SMD task span six steps (Levels 19–24). Thus, the intrinsic psychophysical prominence and availability of the UR and MR were equated a priori.

### Procedure

Participants were randomly assigned to the SUD or SMD tasks. During the 300-trial training phase, they had unlimited response time. Following training, a performance summary presented their total correct responses in green, their total incorrect responses and total points lost from errors in red, and (for SUD participants only) the total URs in blue with the total points potentially saved by those URs.

During testing, participants received—in counterbalanced order—300-trial speeded and unspeeded conditions. In the speeded condition only, trials received incorrect/penalty feedback if participants missed the imposed response deadline of 500 ms. The training, speeded, and unspeeded instructions are given in the [online supplement](#).

### Modeling performance and fitting data

We instantiated a formal model of both tasks. The model assumed that performance was organized along a continuum of psychological representations of increasing strength (from sparse to dense). It assumed that an objective stimulus level would create subjective impressions from trial to trial that vary in a Gaussian distribution around the objective level. This perceptual error would produce discrimination errors near the discrimination breakpoint. Finally, the model assumed a decision process in which criterion lines organized response regions. By the overlay of the sparse–uncertain or sparse–middle

criteria (SU, SM) and the uncertain–dense or middle–dense criteria (UD, MD), the stimulus continuum would be divided into sparse, uncertain (middle), and dense response regions.

We fit the observed performance by moving the model's parameters—perceptual error, the placement of the lower criterion (SU, SM), and the placement of the upper criterion (UD, MD)—through wide ranges. For each parameter configuration, we produced that simulated observer's predicted performance profile during a virtual session, finding its three response proportions—for “S,” “U” or “M,” and “D” responses—for 42 stimulus levels. We minimized standard fit measures to find the best-fitting parameter values. These procedures have been applied to human and animal uncertainty-monitoring data in other studies (e.g., Smith et al., 2006; Smith et al., 2013b).

## Results

### Overall analyses

The experiment was a 2 (task: SUD, SMD)  $\times$  2 (condition: speeded, unspeeded)  $\times$  42 (level) design. Task was a between-subjects factor, and Condition and Level were within-subjects factors. The dependent variable was the proportions of URs/MRs at different stimulus levels under different conditions. Accordingly, we analyzed the data with a 2  $\times$  2  $\times$  42 mixed factorial model (SAS 9.3, GLM procedure). Figure 2 shows the results from the unspeeded and speeded phases of both tasks. Late responses (i.e., beyond the 500-ms deadline) were not analyzed (480 and 846 trials of the 9,000 speeded SUD and SMD trials). Figure S1 (in the supplemental materials) shows late responses distributed across the stimulus continuum.

The effect of task,  $F(1, 58) = 108.78, p < .001, \eta_p^2 = .65$ , confirmed that the UR was used less than the MR. The effect of level,  $F(41, 2378) = 96.14, p < .001, \eta_p^2 = .62$ , confirmed that these responses were used more for intermediate stimulus levels (Fig. 2). The effect of condition was not significant,  $F(1, 58) = 2.83, p = .098, \eta_p^2 = .05$ .

The Task  $\times$  Level interaction,  $F(41, 2378) = 19.91, p < .001, \eta_p^2 = .26$ , confirmed that the MR region was broader and higher across the stimulus continuum than the UR region. The Condition  $\times$  Level interaction,  $F(41, 2358) = 3.51, p < .001, \eta_p^2 = .06$ , confirmed that the third response region (UR or MR) was broader and higher under unspeeded than under speeded conditions. These interaction patterns can be seen in Fig. 2. Most importantly, the Task  $\times$  Condition interaction,  $F(1, 58) = 11.58, p = .001, \eta_p^2 = .17$ , confirmed that the UR and MR reacted differently to the imposition of the deadline. To understand better this crucial interaction, we conducted separate analyses on the SUD and SMD tasks, as did Smith et al. (2013b) and Beran et al. (2009) in their studies of macaques' and capuchin monkeys' URs and MRs.

### Sparse–uncertain–dense task

The UR data were analyzed using an ANOVA with Condition and Stimulus Level as within-subjects factors. The effect of level,  $F(41, 1189) = 19.0, p < .001, \eta_p^2 = .40$ , confirmed that participants found most difficult the levels near the discrimination breakpoint and made URs selectively there (Fig. 2a and b). The effect of condition confirmed that URs were suppressed during the speeded phase,  $F(1, 29) = 10.4, p = .003, \eta_p^2 = .27$ . The Condition  $\times$  Level interaction,  $F(41, 1180) = 2.9, p < .001, \eta_p^2 = .09$ , confirmed that this suppression occurred mostly at the difficult trial levels at which most URs occurred, so this result is unsurprising.

The mean response latencies for the unspeeded “S,” “U,” and “D” responses were 0.48, 1.04, and 0.45 s, respectively. The UR was 124 % slower than the task's perceptual responses. However, we discounted these latency differences in interpreting our results. Beyond the large individual differences in latency (e.g., a sixfold variation in UR latency), we believe that the crucial question is whether the processing behind the UR is time compressible—that is, whether it can survive speeded responding. Evidently, it cannot.

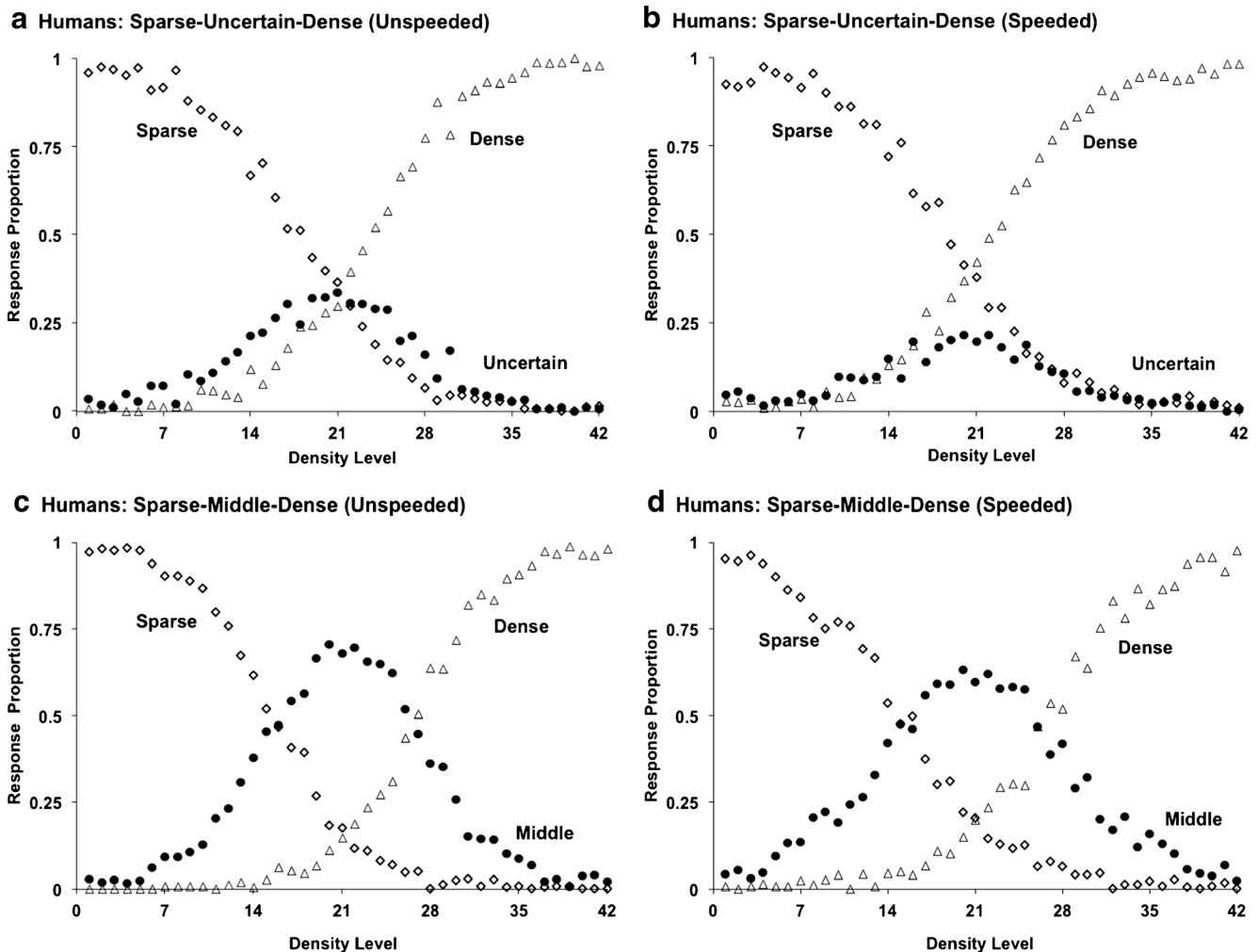
### Sparse–middle–dense task

The MR data were analyzed using the same ANOVA. The effect of level,  $F(41, 1189) = 85.0, p < .001, \eta_p^2 = .75$ , confirmed that participants made MRs most for veridically middle trials (Fig. 2c and d). We found no effect of condition,  $F < 2$ . The speeded phase did not reduce MRs. Rather, MRs actually increased slightly during speeded testing as the MR curve broadened across the stimulus continuum. A significant interaction occurred between condition and level,  $F(41, 1178) = 1.7, p = .005, \eta_p^2 = .05$ , a small effect reflecting this broadening.

The mean response latencies for the unspeeded “S,” “M,” and “D” responses were 0.71, 1.03, and 0.68 s, respectively. The MR was 48 % slower than the task's other responses. It was less differentiated in latency than was the UR [124 % slower;  $t(58) = 2.48, p = .016$ ]. Again, however, we stress that the crucial issue is whether the processing behind the MR is time compressible—that is, whether it can survive speeded responding. Evidently, it can.

Participants used the “S,” “M,” and “D” responses about equally often in both conditions. To confirm this, we found the average response proportions for “S,” “M,” and “D” responses across all 42 trial levels. This is a proxy for the area under each response curve. In the unspeeded and speeded conditions these averages were, respectively, .36, .27, .37 and .35, .28, .37.

Given the psychophysical matching of response regions that we arranged, it is a striking confirmation of the article's results and conclusions that MRs are used much more frequently than URs. This encourages the interpretation that



**Fig. 2** **a** Humans' performance in the sparse–uncertain–dense (SUD) task under unspeeded conditions. The horizontal axis indicates the density level of the box. The “sparse” and “dense” responses, respectively, were correct for boxes at Levels 1–21 and 22–42. The open diamonds and open triangles, respectively, show for each level the proportions of “sparse” and “dense” responses made. The closed circles show the proportions of trials receiving the uncertainty response at each level. **b** Humans' performance in the SUD task under speeded conditions, depicted in the same way. **c** Humans' performance in the sparse–

middle–dense (SMD) task under unspeeded conditions. The horizontal axis indicates the density level of the box. The “sparse,” “middle,” and “dense” responses, respectively, were correct for boxes at Levels 1–18, 19–24, and 25–42. The open diamonds and open triangles, respectively, show for each level the proportions of “sparse” and “dense” responses made. The closed circles show the proportions of trials receiving the “middle” response at each level. **d** Humans' performance in the SMD task under speeded conditions, depicted in the same way

MRs are psychologically more available than URs because they are primary perceptual responses.

The different response frequencies strengthen the results in another way. URs, used less often, nearer floor, had less room to fall. But they fell impressively under deadline conditions. MRs, used more often, had generous room to fall, but did not. These results were obtained despite scaling and regression forces, making the test of the hypothesis conservative and the pattern of results stronger.

An insightful reviewer asked whether one could arrange another kind of response region matching, wherein actual UR and MR response frequencies were equated. Perhaps if we

paid participants \$5 per UR, UR levels might reach MR levels. But, with this artificial inflation, URs would no longer be about uncertainty. Perhaps if we shrank the MR region to just one stimulus step, MR levels might drop to UR levels. But this would qualitatively change the nature of the SMD task. That methodological contortions would be required to equate MR and UR response levels confirms in another way the conclusion that URs and MRs are different psychologically.

We repeated the foregoing analyses, aggregating the data into three stimulus regions: Levels 1–18 (sparse), 19–24 (middle/difficult), and 25–42 (dense). These analyses (see the [supplementary materials](#)) produced essentially identical results.

## Model fits

Figure 3 shows the best-fitting performance profiles from modeling the data, with panels corresponding to those in Fig. 2. Table 1 summarizes the model fits. The model provided excellent fits. It predicted performance within about 0.02 per data point (Column AAD in the table).

The Width column in Table 1 confirms the study's main results: The UR region in the SUD—unspeeded task was 6.3 steps. This is the width that formal modeling in other studies had predicted, an important manipulation check of our paradigm. This justifies again our decision to make the MR region also six steps wide in the SMD task. Our matching process was successful. The UR region in the SUD task was halved during the speeded condition (3.1 steps), leaving a narrow region admitting few URs. The MR region in the SMD task was unaffected. These contrasts motivate the Discussion below.

## Discussion

We asked whether URs, more than primary perceptual responses, would show a distinctive information-processing profile—that is, vulnerability to a speeded-response requirement. Sixty humans completed 54,000 trials in SUD or SMD psychophysical tasks under speeded and unspeeded conditions. The speeded condition reduced URs but not MRs, though MRs and UR were optimally applied to the same stimulus levels.

These results inform the historical controversy surrounding the UR's psychological interpretation. The UR apparently is psychologically different and decisionally distinctive. It is not just a third, middle response applied to a third, middle region. The perceptual purists—who objected to URs in psychophysical tasks because they were psychologically distinctive—were correct to be concerned. The psychologists who embraced URs *because* they behaviorally index uncertain consciousness states were possibly correct, too.

These results may also explain why URs show special fragility and changeability (e.g., Smith et al., 2006). That is, the UR as used by humans and macaques—but not their sparse and dense responses—shows strong individual differences in use and underuse. This supports the idea that the UR is different from the primary perceptual responses and serves a distinctive role. Related observations—for example, that the UR is sharply affected by instructional set and temperamental tentativeness—originally led early psychophysicists to consider giving URs a higher-level psychological interpretation.

There are conceptual grounds for considering this higher-level interpretation. First, the perceptual responses are directly rewarded/penalized. They could be conditioned by primary reinforcement systems in brain. The UR, on the other hand, is never rewarded or penalized. It cannot be dependent on those conditioning systems in the same way.

Second, the perceptual responses—sparse, middle, and dense—are objectively associated with a concrete range of perceptual inputs. These ranges are entrained by reinforcements delivered. The UR, however, is not tied to any objective stimulus range. Its range is whatever the participant's uncertainty system says. Its range must be constructed through the participant's internal decisional processes.

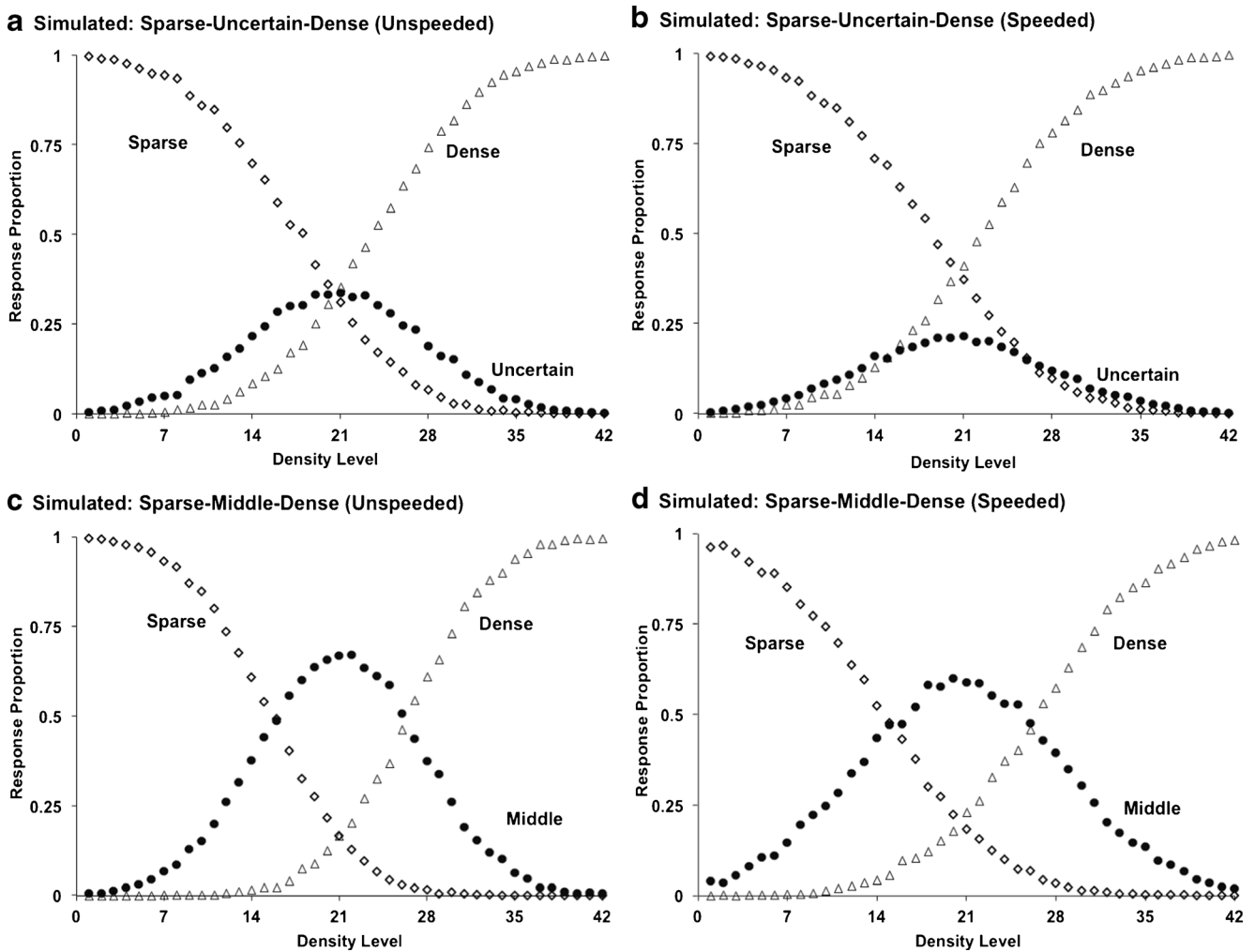
Third, tasks like the SUD task characteristically present to-be-classified perceptual stimuli that are inherently indeterminate—that is, possibly sparse or dense. Indeterminacy is the result when perceptual error scatters difficult stimuli near the task's discrimination breakpoint. To resolve the indeterminacy and choose adaptive behaviors, the participant has to engage higher levels of controlled, deliberate cognition—controlled processing, in Schneider and Shiffrin's (1977) sense. Thus, URs might be psychologically distinctive for being based on controlled processes recruited near a discrimination threshold when close perceptual calls require a referee.

Paul, Smith and Ashby (in preparation) offered support for this idea. Their fMRI study showed that URs, as compared to primary perceptual responses, activate distinctive neural networks that include anterior cingulate cortex, prefrontal cortical areas, and insula, suggesting that they are not just MRs.

The present results also inform the animal metacognition literature. They suggest that URs may have a more controlled, decisional basis that lifts them above the associative plane of processing—a goal that comparative-metacognition researchers have sought. The results may also help solve a cross-species metacognitive mystery. Beran et al. (2009) found that capuchin monkeys (*Cebus apella*, a New World primate) made almost no URs in an SUD task but made MRs generously in an SMD task. This dissociation suggests that capuchin monkeys have the perceptual processes that support MRs but possibly not the higher-level processes that support URs in humans and macaques (*Macaca mulatta*, an Old World primate).

More broadly, our research begins to trace the information-processing signature of basic metacognitive responses like the UR. Here we asked whether URs are dependent on temporally incompressible processes. In a companion study, Smith et al. (2013b) asked whether URs are working memory intensive. They found that concurrent working memory loads disrupted macaques' URs far more than their MRs.

Given this information-processing signature, comparative or developmental psychologists can then apply it to their species or age group and evaluate the metacognitive sophistication of their participants. Obviously, if macaques, other species, or young children fail to show the same information-processing signature, this down-interprets their metacognitive capacity. But if they do show the signature, it strengthens the isomorphism between human and animal metacognition, for example, with profound implications regarding the emergence of metacognition and reflective mind in the primate order.



**Fig. 3** **a** The best-fitting predicted performance profile when the signal-detection model described in the text was fit to humans’ unspeeded performance in the sparse–uncertain–dense (SUD) task. The horizontal axis indicates the density level of the box. The “sparse” and “dense” responses, respectively, were correct for boxes at Levels 1–21 and 22–42. The open diamonds and open triangles, respectively, show for each level the best-fitting proportions of “sparse” and “dense” responses. The closed circles show the best-fitting proportions of trials receiving the uncertainty response at each level. **b** The best-fitting predicted performance profile when the signal-detection model was fit to humans’ speeded performance in the SUD task, depicted in the same way. **c** The best-fitting predicted

performance profile when the signal-detection model was fit to humans’ unspeeded performance in the sparse–middle–dense (SMD) task. The horizontal axis indicates the density level of the box. The “sparse,” “middle,” and “dense” responses, respectively, were correct for boxes at Levels 1–18, 19–24, and 25–42. The open diamonds and open triangles, respectively, show for each level the best-fitting proportions of “sparse” and “dense” responses. The closed circles show the best-fitting proportions of trials receiving the “middle” response at each level. **d** Humans’ performance in the SMD task under speeded conditions, depicted in the same way

**Table 1** Details of model fits

Task	Condition	Figures	Perceptual error	Low criterion	High criterion	Width	SSD	AAD
SUD	Unspeeded	2A, 3A	6.8	17.3	23.6	6.3	0.0731	0.0184
SUD	Speeded	2B, 3B	7.5	19.0	22.1	3.1	0.0745	0.0187
SMD	Unspeeded	2C, 3C	5.6	15.2	26.9	11.7	0.0559	0.0162
SMD	Speeded	2D, 3D	7.3	14.9	26.8	11.9	0.1195	0.0239

SSD, sum of the squared deviations between the corresponding 126 observed and predicted data points produced by real and simulated observers. AAD, average of the absolute-value deviations between the corresponding 126 observed and predicted data points produced by the real and simulated observers, with the deviations always signed positively

Our research also grants comparative-metacognition research a constructive new perspective. Some theorists (e.g., Le Pelley, 2012) have mistakenly gathered all animals' uncertainty performances together under the rubric "associative." This approach neglects to analyze carefully these tasks' real information-processing requirements (Smith et al., 2013a). It blurs together animal performances that may be importantly different in psychological character. The present research is valuable because it models the practice of using information-processing benchmarks to distinguish tasks cognitively and psychologically. The benchmark approach used here is equally applicable to human minds or monkey minds, and it has the potential to contribute strongly to the continuing theoretical development of the animal metacognition literature. In fact, in a recent target-article-commentary cycle, this approach was a major topic of discussion (Smith et al., 2013a, and commentaries). In a sense, the present article points the way toward the next phase of animal metacognition research, by providing an empirical role model.

In the end, we conclude that the UR has the potential to support the development of animal models for metacognition, grounding the search for neurochemical blocks and enhancers. It could also support imaging research to map the distinctive brain organization of metacognition in humans. It extends the techniques available to child-development researchers, because young children can perform behavioral uncertainty tasks before typical verbal/introspective metacognition tasks. It supports the study of metacognition in language-delayed or autistic children and promotes the training of metacognition in educationally challenged populations. It might reveal which facets of metacognition are possible at a basic behavioral level, and which facets are denied nonverbal creatures. Thus, we believe that the behavioral uncertainty response will continue to play a constructive role in metacognition research.

**Author note** This research was supported by NICHD Grant No. 1R01HD061455 and NSF Grant No. BCS-0956993. We thank our undergraduate assistants for help with data collection.

## References

- Balcomb, F. K., & Gerken, L. (2008). Three-year-old children can access their own memory to guide responses on a visual matching task. *Developmental Science*, *11*, 750–760.
- Beran, M. J., Smith, J. D., Coutinho, M. V. C., Couchman, J. J., & Boomer, J. (2009). The psychological organization of "uncertainty" responses and "middle" responses: A dissociation in capuchin monkeys (*Cebus apella*). *Journal of Experimental Psychology: Animal Behavior Processes*, *35*, 371–381.
- Boring, E. G. (1921). The control of attitude in psycho-physical experiments. *Psychological Review*, *27*, 440–452. doi:10.1037/h0075451
- Dunlosky, J., & Bjork, R. A. (Eds.). (2008). *Handbook of metamemory and memory*. New York, NY: Psychology Press.
- Hampton, R. R. (2001). Rhesus monkeys know when they remember. *Proceedings of the National Academy of Sciences*, *98*, 5359–5362.
- Hampton, R. R. (2009). Multiple demonstrations of metacognition in nonhumans: Converging evidence or multiple mechanisms? *Comparative Cognition and Behavior Reviews*, *4*, 17–28.
- Jozefowicz, J., Staddon, J. E. R., & Cerutti, D. T. (2009). Metacognition in animals: How do we know that they know? *Comparative Cognition and Behavior Reviews*, *4*, 29–39.
- Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, *103*, 490–517.
- 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*, 686–708. doi:10.1037/a0026478
- Paul, E. J., Smith, J. D., & Ashby, F. G. (in preparation). *Neural correlates of uncertainty monitoring in human categorization*. Manuscript in preparation.
- Ruffman, T. (2000). Nonverbal theory of mind: Is it important, is it implicit, is it simulation, is it relevant to autism? In J. W. Astington (Ed.), *Minds in the making: Essays in honor of David R. Olson* (pp. 250–266). Oxford: Blackwell.
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*, 1–66. doi:10.1037/0033-295X.84.1.1
- Schwartz, B. L. (1994). Sources of information in metamemory: Judgments of learning and feelings of knowing. *Psychonomic Bulletin & Review*, *1*, 357–375.
- Smith, J. D., Beran, M. J., & Couchman, J. J. (2012). Animal metacognition. In T. Zentall & E. Wasserman (Eds.), *The Oxford handbook of comparative cognition* (pp. 282–304). Oxford, UK: Oxford University Press.
- Smith, J. D., Beran, M. J., Redford, J. S., & Washburn, D. A. (2006). Dissociating uncertainty responses and reinforcement signals in the comparative study of uncertainty monitoring. *Journal of Experimental Psychology: General*, *135*, 282–297. doi:10.1037/0096-3445.135.2.282
- Smith, J. D., Couchman, J. J., & Beran, M. J. (2013a). Animal metacognition: A tale of two comparative psychologies. *Journal of Comparative Psychology*. doi:10.1037/a0033105
- Smith, J. D., Coutinho, M. V. C., Church, B. A., & Beran, M. J. (2013b). Executive-attentional uncertainty responses by rhesus macaques (*Macaca mulatta*). *Journal of Experimental Psychology: General*, *142*, 458–475. doi:10.1037/a0029601
- Watson, C. S., Kellogg, S. C., Kawanishi, D. T., & Lucas, P. A. (1973). The uncertain response in detection-oriented psychophysics. *Journal of Experimental Psychology*, *99*, 180–185.