COGNITIVE NEUROSCIENCE

Perceptual meta-uncertainty reveals enhanced calibration of sensory confidence in autism

Laurina Fazioli[1], Bat-Sheva Hadad[1,2], Rachel N. Denison[3], Amit Yashar[1,2]*

Atypical metacognition has been suggested to underlie autistic phenotypes, given its role in social cognition and behavioral flexibility. However, no study has quantitatively assessed metacognitive abilities in autism. Here, we measured meta-uncertainty—the noise corrupting estimates of one's own decision uncertainty—in autism. In three experiments, autistic and non-autistic participants ($N = 145$) performed orientation categorization tasks while simultaneously reporting their choice confidence. By independently manipulating each Bayesian component—sensory uncertainty, prior, and reward—and fitting a recently established process model, we assessed metacognitive abilities and their contingency on the Bayesian components while controlling for first-order decisions. Unlike non-autistic participants, autistic participants' meta-uncertainty depended on which decision component was manipulated and was lower than that of non-autistic participants, specifically when decisions were adjusted for sensory uncertainty. These findings reveal that metacognition in autism is not generally reduced but rather enhanced for inferences that rely primarily on sensory information.

INTRODUCTION

Metacognition—the ability to monitor and evaluate one's own mental states—plays a fundamental role in human cognition. It allows individuals to assess the reliability of their perceptions, guide learning, and adjust behavior in uncertain environments (*1*). Failures of metacognition have been implicated across a range of psychiatric and neurodevelopmental conditions (*2*, *3*). In autism spectrum disorder—marked by atypical social cognition, restricted and repetitive behavior, and altered sensory processing (*4*)—impaired metacognition could contribute to alterations in both social reasoning and perceptual processing. However, despite intense interest, the nature of metacognition in autism remains unclear. Some accounts suggest a broad reduction in metacognitive ability (*5*), while others point to more selective alterations—such as difficulties in metacognitive control (*6*) or in tasks involving social cognition (*7*). Previous research has primarily relied on the accuracy of confidence reports in memory, cognitive, or perceptual tasks [reviewed in (*5*)]. However, confidence reports reflect not only self-monitoring abilities but also general task performance and biases in confidence decision boundaries (*8*–*12*), making it difficult to isolate genuine metacognitive differences.

Here, we adopt a computational approach that directly formalizes the processes underlying confidence (*13*). We draw on Bayesian theories of perception and decision-making (*14*, *15*), according to which perceptual decision-making is formalized as an inference process that combines sensory uncertainty (i.e., likelihood), prior expectations (i.e., internal models), and reward (i.e., cost function) to compute a decision criterion that minimizes expected cost (*14*, *15*). Recent perceptual categorization studies have shown that, contrary to longstanding views (*16*, *17*), autistic individuals integrate all Bayesian components in perceptual decision-making in a manner similar to non-autistic individuals (*18*–*20*). Nevertheless, it remains possible that atypicalities in perceptual decision-making arise at the level of metacognitive monitoring in autism.

Across three experiments, autistic and non-autistic participants performed a perceptual categorization of orientation task (first-order decision-making) while reporting their confidence (second-order decision-making). By independently manipulating prior probabilities, reward structures, and sensory uncertainty, we were able to ask whether—and how—distinct Bayesian components shape metacognitive ability in autism. The present study builds on previously collected datasets in which perceptual decisions were measured (*18*, *20*); here, we analyze the associated confidence reports, which have not been previously reported, to characterize metacognitive processes.

We independently manipulated Bayesian components by varying category base rate (experiment 1) and reward points attributed per correct answer (experiment 2) in task 1. We manipulated sensory uncertainty (experiment 3; task 2) by using an embedded task, designed specifically to induce a decision criterion shift by varying sensory evidence only—here, stimulus contrast (*21*–*23*). We quantified metacognitive abilities by using a recently developed process model of metacognition on the confidence data (the "CASANDRE" or "confidence as a noisy decision reliability estimate" model) (*13*), which quantifies the noise corrupting internal estimates of uncertainty (hence, meta-uncertainty). More specifically, the meta-uncertainty parameter estimates the consistency with which internal decision uncertainty—arising from evaluating sensory noise—is mapped onto confidence reports across trials, providing a measure of metacognitive ability that is independent of task difficulty and confidence bias. Our quantitative approach to metacognition reveals a qualitative divergence between autistic and non-autistic individuals in how Bayesian factors contribute to metacognitive ability. While for non-autistic participants, metacognitive abilities remained stable across experiments, autistics' meta-uncertainty varied depending on which Bayesian component biased first-order decisions. Specifically, they exhibited enhanced abilities, relative to non-autistic participants, when perceptual decisions relied on sensory information alone.

RESULTS

Participants (52 autistic and 93 non-autistic) categorized the orientation of grating stimuli (first-order task) and reported their

[1]The Department of Special Education, Faculty of Education, University of Haifa, Haifa, Israel. [2]The Edmond J. Safra Brain Research Center for the Study of Learning Disabilities, University of Haifa, Haifa, Israel. [3]Department of Psychological and Brain Sciences, Boston University, Boston, USA.
*Corresponding author. Email: amit.yashar@edu.haifa.ac.il

confidence (second-order task) on every trial (Fig. 1A). In all experiments, to manipulate sensory uncertainty, we varied stimulus contrast across seven values. We manipulated information level regarding orientation category by randomly varying the stimulus orientation on each trial. In task 1, for each trial, stimulus orientation was drawn from one of the two partially overlapping Gaussian distributions (*21–23*), with means $m_A = -4°$ (category A) and $m_B = 4°$ (category B) and SDs $s_A = s_B = 5°$ (Fig. 1B, task 1). Participants reported simultaneously the stimulus category (category A versus B) and confidence rating (four-point scale) by pressing one of eight keys, ranging from category A highly confident to category B highly confident (Fig. 1A). This simultaneous report prevented postdecision bias (*24*).

In task 1, we tested how participants made second-order confidence decisions when first-order perceptual decisions integrated prior (experiment 1) or reward (experiment 2) information with sensory uncertainty. In experiment 1, we manipulated priors by explicitly varying category base rate probability across blocks. On a given block of trials, categories could appear with balanced (category A = 50% and category B = 50%) or unbalanced base rate probabilities (category A = 25% and category B = 75% or category A = 75% and category B = 25%). In experiment 2, we varied the points awarded for correct answers across three blocks of trials. In the unbiased reward block, each correct response was awarded 2 points. In the two biased reward blocks, a correct response was awarded 3 points for

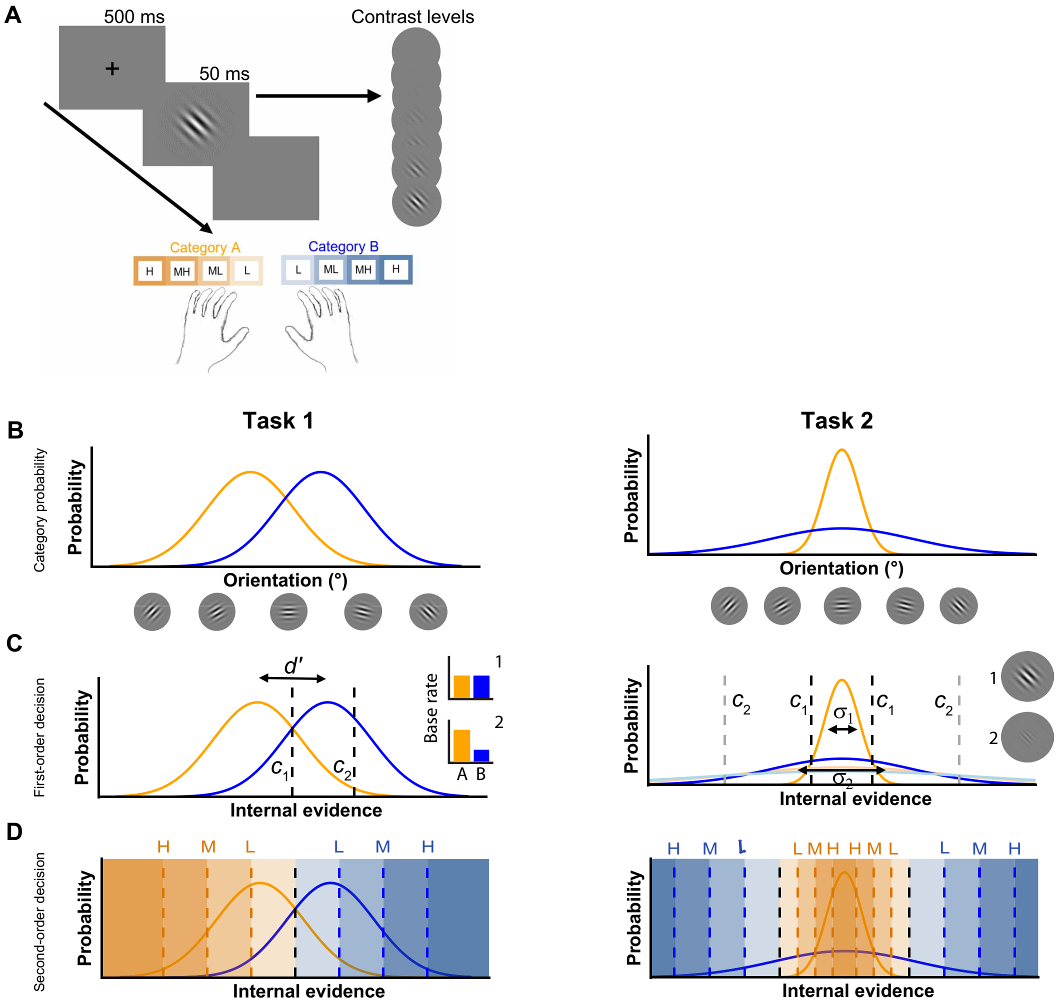

Fig. 1. Task descriptions. (**A**) Sequence of events within a trial in experiments 1, 2, and 3, respectively, manipulating prior, reward, and sensory uncertainty. Participants simultaneously reported the Gabor's category (category A or category B) based on orientation and their confidence level (high, medium-high, medium-low, and low), using eight response keys, ranging from high-confidence category A to high-confidence category B. Stimulus contrast randomly varied between trials from seven fixed values. (**B**) Stimulus orientation distributions for each category in task 1 (experiments 1 and 2) and task 2 (experiment 3). (**C**) Internal representation of the category distributions. In task 1, the sensitivity d' represents the ability to separate the categories, and c represents the decision criterion, adjusting from equal prior/reward (c_1) to prior/reward that favors category A (c_2). In task 2, vivid colors represent the internal representations of the categories when the sensory noise is low (high contrast) and faded colors when the sensory noise is high (low contrast). σ represents the SD of the internal representation of category A (i.e., combination of internal—inverse of d'—and external noises), leading to a narrow distribution when sensory noise is low (σ_1) and wider when sensory noise is high (σ_2). c_1 represents the decision boundaries, shifting outward when sensory noise is increasing (c_2). (**D**) Second-order decisions for each task. The black vertical lines represent c. The shaded areas represent the confidence reported as a function of orientation, from high (darker area) to low confidence (lighter area), for categories A (orange) and B (blue). The colored vertical lines represent the confidence criterion c_c, indicating the stimulus information (i.e., orientation) required to report a specific level of confidence. Three confidence criteria are displayed, representing the confidence boundary between high and medium-high (H), medium-high and medium-low (M), and medium-low and low confidence (L).

one category and 1 point for the other. Specifically, in one biased block, category A was awarded 3 points, while in another biased reward block, category B was awarded 3 points.

We implemented the CASANDRE model (*13*) to compute metacognitive abilities from participants' categorization and confidence responses. From the model fits, we extracted signal-detection-theory–like parameters, which provided estimates of first-order decisions—sensitivity (i.e., d') and decision criterion (i.e., c) (*9*, *11*). Here, we expected the decision criterion to shift toward the more likely or rewarded category (Fig. 1C, task 1) and confidence to increase as orientations deviated from points of maximum overlap between categories (Fig. 1D, task 1).

In experiment 3, we aimed to assess second-order decisions when first-order decisions are driven solely by sensory uncertainty. In task 1, when prior probabilities and rewards are balanced, the optimal decision criterion lies at the intersection between the category distributions and does not depend on stimulus uncertainty (*22*, *23*). To isolate the effect of sensory uncertainty, we therefore used an embedded category task (task 2), a well-established paradigm in which decision criteria adjust as a function of the reliability of sensory evidence (*21*–*23*). As in task 1, participants categorized stimuli based on orientation. Stimuli were drawn from two Gaussian distributions with identical means ($\mu_A = \mu_B = 0°$) but different SDs ($\sigma_A = 3°$ and $\sigma_B = 12°$), corresponding to narrow and broad categories (Fig. 1B). Thus, orientations closer to 0° were more likely to come from category A, whereas more tilted orientations were more likely to come from category B. These category distributions—and thus the prior—are fixed across trials. Sensory uncertainty was manipulated via stimulus contrast, which affects the reliability of the sensory evidence and broadens the internal likelihood function on each trial. As uncertainty increases, the relative evidence for the two categories changes across stimulus orientations, leading to outward shifts in the optimal decision boundary (Fig. 1C). Thus, whereas in task 1 decision shifts are driven by prior or reward information, in task 2, they are driven by trial-by-trial changes in sensory uncertainty alone. We further expected confidence to reflect the relative evidence for each category (Fig. 1D). In task 1, orientations around the category overlap (0°) were expected to be reported with low confidence for both categories and orientations that deviated enough from the mean to be reported with high confidence. In task 2, orientations around the means of the two categories (0°) were expected to be associated with high confidence for category A, and the confidence for category A was expected to diminish as orientations deviated, until stimulus evidence favored category B. Extreme deviation of orientation was expected to lead to greater confidence when reporting B.

After removing participants with accuracy at chance level (see the "Outlier removal" section of Materials and Methods and Table 1), the final sample included 46 non-autistic and 31 autistic participants in experiment 1, 44 non-autistic and 30 autistic participants in experiment 2, and 42 non-autistic and 30 participants in experiment 3. A small number of additional participants were excluded from specific analyses (e.g., raw data and criterion) when individual parameter values exceeded 2.5 SD from the mean, as detailed in Materials and Methods. All findings remained consistent when analyses were performed on the final sample, including these specific outliers (see the "Robustness check—Analyses of all participants including outliers" section of Supplementary Results).

Confidence ratings reflect sensory uncertainty for both groups

First, we examined how decision and confidence varied with stimulus orientation (11 orientations) and strength (7 contrast levels). Figure 2 illustrates the proportion of reporting category B (top row) and the mean of confidence report (bottom row), as a function of orientation, contrast level, and group. Values reported are across base rate and reward blocks for task 1 (Fig. 2, A to D). For experiments 1 and 2, manipulating category base rate and reward (Fig. 2, A to D, top row), category report was characterized by a sigmoid shape, with a proportion of reports for category B increasing as the orientation became more clockwise. The sigmoid was steeper with high contrasts, reflecting that category report was more sensitive to orientation as contrast increased. See fig. S6 for visualization per prior/reward blocks. In experiment 3, which manipulated sensory uncertainty, we observed that the probability of reporting category B (wider distribution) increased as orientations deviated from 0° (i.e., mean of the narrow distribution), and the categorization became more sensitive to orientation as contrast increased (Fig. 2, E and F, top row; see the "Additional results" section of Supplementary Results, fig. S6, and table S1 for the statistical analyses). Fazioli *et al.* (*18*–*20*) conducted optimal observer analyses on the data and showed similar first-order decisions for autistic and non-autistic groups when comparing them to optimal observers.

The variability in category reports is reflected in the confidence choices. Figure 2 (A to F, bottom row) illustrates how confidence increased as the probability of selecting a given category rose, with this relationship becoming sharper when stimulus strength increases (i.e., higher contrast). This consistent relationship between category

Table 1. Description of the sample size in the three experiments for the overall sample, and statistical analyses performed on each estimate: raw data (i.e., category and confidence reports), sensitivity, decision criterion, confidence criterion, meta-uncertainty, and guess rate.

	Overall *n*	Raw data	Sensitivity	Decision criterion	Confidence criterion	Meta-uncertainty	Guess rate
Prior experiment	$n_{autistic} = 34$	$n_{autistic} = 31$	$n_{autistic} = 29$	$n_{autistic} = 30$	$n_{autistic} = 31$	$n_{autistic} = 31$	$n_{autistic} = 31$
	$n_{non\text{-}autistic} = 49$	$n_{non\text{-}autistic} = 46$	$n_{non\text{-}autistic} = 46$	$n_{non\text{-}autistic} = 45$	$n_{non\text{-}autistic} = 46$	$n_{non\text{-}autistic} = 46$	$n_{non\text{-}autistic} = 46$
Reward experiment	$n_{autistic} = 32$	$n_{autistic} = 30$	$n_{autistic} = 29$	$n_{autistic} = 28$	$n_{autistic} = 30$	$n_{autistic} = 30$	$n_{autistic} = 30$
	$n_{non\text{-}autistic} = 48$	$n_{non\text{-}autistic} = 44$	$n_{non\text{-}autistic} = 43$	$n_{non\text{-}autistic} = 42$	$n_{non\text{-}autistic} = 44$	$n_{non\text{-}autistic} = 44$	$n_{non\text{-}autistic} = 44$
Sensory uncertainty experiment	$n_{autistic} = 34$	$n_{autistic} = 30$	$n_{autistic} = 28$	$n_{autistic} = 29$	$n_{autistic} = 29$	$n_{autistic} = 30$	$n_{autistic} = 30$
	$n_{non\text{-}autistic} = 44$	$n_{non\text{-}autistic} = 42$	$n_{non\text{-}autistic} = 42$	$n_{non\text{-}autistic} = 40$	$n_{non\text{-}autistic} = 41$	$n_{non\text{-}autistic} = 42$	$n_{non\text{-}autistic} = 42$

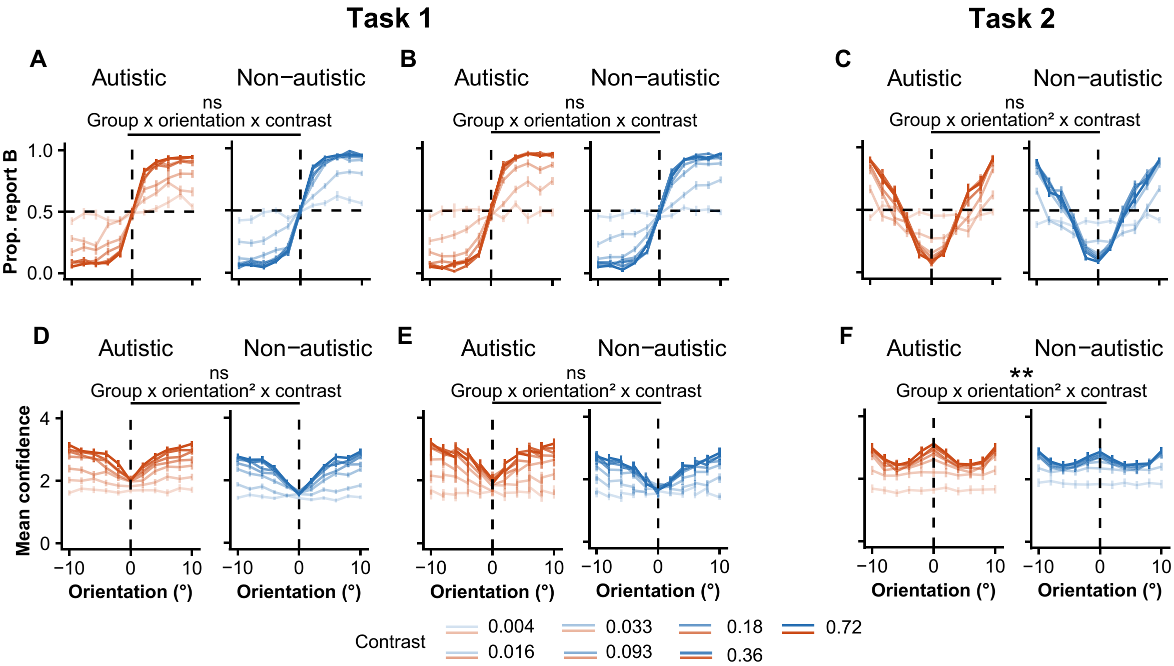

Fig. 2. Behavioral data. Category and confidence report data for (**A** and **D**) experiment 1, prior manipulation, (**B** and **E**) experiment 2, reward manipulation, and (**C** and **F**) experiment 3, sensory uncertainty manipulation. The top row represents the proportion of reporting category B (*y* axis) as a function of orientation (*x* axis) and contrast (line transparency). The bottom row illustrates the mean of confidence (*y* axis) as a function of orientation (*x* axis) and contrast level (line transparency). In (A) and (B), data show means across participants and base rate/reward blocks. In (C), data show means across participants. Error bars represent ±SE. The interactions between group, orientation, and contrast, on the proportion of reporting B, and confidence reported, were evaluated using linear mixed-effect models. not significant (ns): $P \geq 0.05$ and ** $0.01 > P \geq 0.001$. The sample size consisted of 31 autistic and 46 non-autistic participants in (A) and (D), 30 autistic and 44 non-autistic participants in (B) and (E), and 30 autistic and 42 non-autistic participants in (C) and (F).

and confidence choices indicates that participants can assess the reliability of their decision. To quantify how confidence choice was associated with stimulus information, we performed linear and quadratic mixed-effect models with binned orientation and contrast as within-subject factors and group as a between-subject factor, on confidence ratings.

Confidence depends on the stimulus value and strength in both groups

The mixed-effect models investigating the effects of stimulus information (i.e., orientation and contrast), block conditions (i.e., base rate for experiment 2, reward for experiment 2, and no block condition for experiment 3), and group on confidence report included both linear and quadratic factors of orientation. Here, we focused on the overall V-shaped pattern and did not interpret linear terms separately.

Prior manipulation (experiment 1). In experiment 1, as expected, confidence ratings were higher as contrast increased [$t(31.17) = 5.58$, $P < 0.001$] and as orientations deviated from 0° [$t(57.79) = 4.27$, $P < 0.001$]. The effect of contrast was more pronounced when orientations deviated from 0°, as indicated by the interaction between squared orientation and contrast [$t(61140) = 6.95$, $P < 0.001$]. Participants reported overall higher confidence when base rate was unbalanced, compared to balanced [$t(61140) = 2.84$, $P = 0.005$] (see fig. S7A). Overall confidence ratings did not significantly vary across groups [$t(5312) = 1.20$, $P = 0.229$]. All other main effects and interactions were not significant (see the "Additional results" section of Supplementary Results and table S2).

Reward manipulation (experiment 2). Similarly to experiment 1, confidence ratings were higher as contrast increased [$t(26.19) = 4.42$, $P < 0.001$] and when orientations deviated from 0° [$t(64.66) = 5.86$, $P < 0.001$], and the effect of contrast was more pronounced when orientations deviated from 0°, as indicated by the interaction between squared orientation and contrast [$t(61890) = 5.30$, $P < 0.001$]. Participants were more confident in the unbalanced reward block than the balanced reward block [$t(61860) = 3.00$, $P = 0.003$] (see fig. S7B). A significant three-way interaction between contrast, reward, and squared orientation [$t(61890) = 2.38$, $P = 0.017$] revealed that confidence ratings were more sensitive to a stimulus value and strength when reward was unbalanced. However, these effects did not vary across groups, as all remaining effects, including the main effect of group [$t(12.26) = -0.86$, $P = 0.410$] and interactions, were not significant (see the "Additional results" section of Supplementary Results and table S2).

Sensory uncertainty manipulation (experiment 3). As in the categorization task in experiments 1 and 2, confidence ratings in task 2 increased with contrast [$t(72.89) = 8.37$, $P < 0.001$] and changed as a function of orientation [$t(67.34) = 2.56$, $P = 0.012$]. Although overall confidence ratings did not significantly vary across groups [$t(43.80) = 0.32$, $P = 0.753$], autistic participants' confidence was more sensitive to a stimulus value and strength (Fig. 2C), as indicated by a significant interaction between contrast and group [$t(72.48) = -2.42$, $P = 0.018$] and a three-way interaction between group, squared orientation, and contrast [$t(113{,}900) = 3.05$, $P = 0.003$]. All remaining effects and interactions were not significant (see the "Additional results" section of Supplementary Results and table S2).

These results indicate that in both groups in task 1 (experiments 1 and 2), confidence ratings were driven by stimulus strength (i.e., contrast), with this effect modulated by a stimulus value (i.e., orientation). The relation between a stimulus value and strength varied between autistic and non-autistic individuals when first-order decision boundaries were adjusted for sensory uncertainty alone, as in task 2 (experiment 3). However, these relations rely on average adjustment of confidence based on participants' assessment of the strength of their first-order decision, but these relations do not reflect trial-to-trial variability in this assessment. This variability, termed meta-uncertainty, constitutes metacognitive abilities. To estimate each participant's meta-uncertainty, we next fitted the CASANDRE models of confidence (*13*) that quantify their estimation of their own decision reliability, separately for each experiment.

Meta-uncertainty explains confidence reports in autism

According to the CASANDRE model, in perceptual decision-making, metacognitive ability in confidence report is determined by the observer's reliability of their uncertainty estimate on their perceptual choice. This well-established model explains previous works using both tasks 1 and 2 (*13*, *23*). The model has two-stage processes (a first-order decision and a second-order confidence), and it is well rooted in traditional signal detection theory. The model separates discrimination abilities and response bias from metacognitive abilities. By fitting the model to the individual data, we estimated for each participant the following parameters: sensitivity (s), decision criterion (c_d), guess rate (GR), confidence criterion (c_c), and meta-uncertainty (σ_m). Hence, σ_m provides a measure of estimation of internal noise that is independent of sensitivity and first-order decision. These parameters were optimized to minimize the negative log-likelihood and best capture participants' behavioral data (see the "Computational model" section of Materials and Methods).

Fit of meta-uncertainty model on category and confidence reports

See Supplementary Methods for details about comparisons between different variants of the meta-uncertainty model and between models with meta-uncertainty as a free parameter (meta-uncertainty mode) and identical models with meta-uncertainty fixed at 0 (restricted model). Results demonstrate that meta-uncertainty plays a role in confidence reports in all experiments (see Supplementary Results).

Figure 3 illustrates how category and confidence reports were predicted by the meta-uncertainty model. The figure displays proportion of reporting category B (top row) and mean of confidence report (middle row) as a function of stimulus orientation and the two extreme contrast values. Bottom row displays the mean confidence as a function of the proportion of reporting B and contrast level. All subplots show observed (points) and predicted (solid lines) data

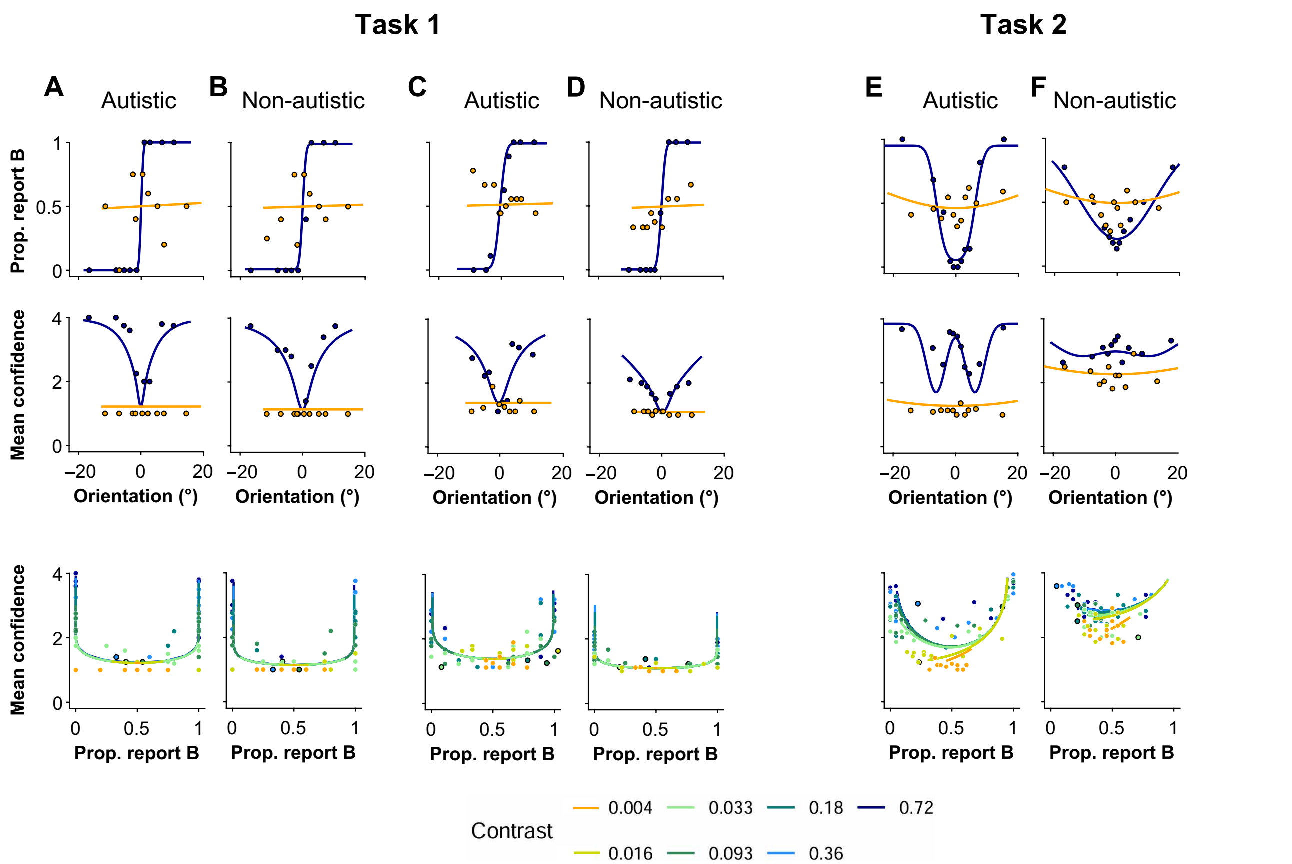

Fig. 3. Model fitting for a sampled participant from each group in each experiment. Proportion of reporting category B (top row, *y* axis) and mean of confidence report (middle row, *y* axis) as a function of stimulus orientation (*x* axis) and the two extreme contrast values—0.004 and 0.72, for the prior (**A** and **B**), reward (**C** and **D**), and sensory evidence experiments (**E** and **F**). The bottom row displays the mean of confidence (*y* axis) as a function of proportion of reporting category B (*x* axis), contrast (color), and group, for each experiment. Each subplot displays the observed and predicted behavior for an individual participant. Data points illustrate choice behavior, with size proportional to the number of trials. The solid lines represent the fit of the meta-uncertainty model using the maximum likelihood estimation method. The model was fit to all data simultaneously for each experiment and group.

for individual subjects, for the prior (Fig. 3, A and B), reward (Fig. 3, C and D), and sensory uncertainty (Fig. 3, E and F) experiments. Figures showing the individual model fittings are displayed in figs. S9 to 37.

The sigmoid, "V," and "W" shapes observed in Fig. 2 for high contrasts—characteristic of category and confidence reports that are sensitive to stimulus information—were reproduced by the model predictions and fitted properly with the observed data (Fig. 3, top and middle rows). For low contrasts, the fitted lines flattened, reproducing well the reduced association of category and confidence reports with stimulus information. These observations demonstrate that the pattern of category and confidence reports was well captured by the meta-uncertainty model. In the sensory uncertainty experiment, we noticed that autistic individuals exhibited a more detailed association between reports and stimulus information, as illustrated by the steeper curve in task 2 for the autistic participant, suggesting a greater association between confidence and choice consistency. Therefore, participants' behavior was well captured by the meta-uncertainty in all experiments. Figure 3 (bottom row) illustrates how the association between choice consistency and confidence was predicted by the meta-uncertainty model. For the prior and reward experiments, we observed a single association ("U" shape) between confidence and category report across contrast conditions. This behavior—captured by the meta-uncertainty model, as the predicted behavior closely fitted the data—demonstrates participants' ability to assess the reliability of their decision. In the sensory uncertainty experiment, the more complex pattern of association between confidence and choice consistency was also captured by the meta-uncertainty model. Here, we noticed a greater variance in this association for autistic participants, as illustrated in Fig. 3 (E and F), suggesting more nuanced mapping of confidence on choice consistency.

Comparable first-order sensitivity

We next analyzed model-derived parameters reflecting first-order processes (i.e., sensitivity and decision criterion) to confirm that any differences in perceptual decisions between groups did not stem from atypical first-order decisions in autistic participants. Sensitivity was computed during the model fitting using a Naka-Rushton function, leading to three parameters estimating the gain of sensory response as sensory evidence increases. We then used an inverted function to derive sensitivity values for each level of contrast for each participant. Figure 4 illustrates perceptual sensitivity (top row) and decision criterion (bottom row) as a function of contrast level and group. In Fig. 4 (A and B), values are reported across base rate and reward blocks. During the model fitting for experiments 1 and 2, the decision criterion in the balanced block was fixed at 0 to reduce the number of free parameters. Therefore, Fig. 4 (bottom row) shows the criterion for the unbalanced block, which reflects the deviation from the balanced-block criterion.

Prior manipulation (experiment 1). The analysis of variance (ANOVA) performed on sensitivity (s) revealed that sensitivity decreased with decreasing contrasts [$F(1.36, 99.47) = 115.13$, $P < 0.001$, $\eta_p^2 = 0.61$], indicating that the contrast variation was an effective manipulation of stimulus reliability (Fig. 4A, top row). None of the other effects were significant (see the "Additional results" section of Supplementary Results), indicating that both groups exhibited a comparable perceptual sensitivity to the contrast levels.

Reward manipulation (experiment 2). The ANOVA performed on s revealed that sensitivity decreased as contrast decreased [$F(1.33, 92.76) = 81.97$, $P < 0.001$, $\eta_p^2 = 0.54$] (Fig. 4, top row). Furthermore, sensitivity was higher in the unbalanced, compared to the balanced reward blocks [$F(1, 70) = 7.06$, $P = 0.010$, $\eta_p^2 = 0.09$] (see fig. S8), more

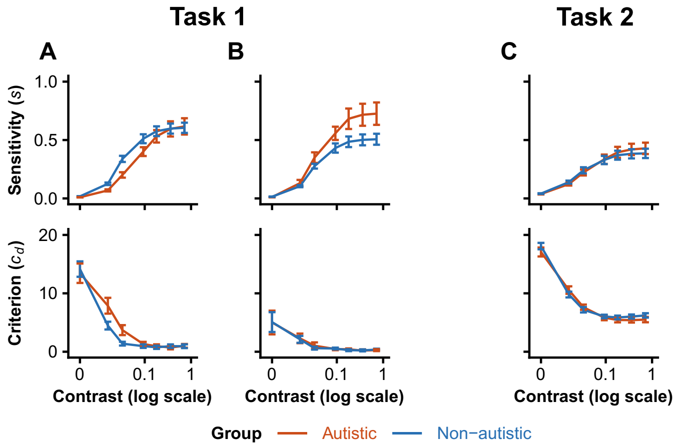

Fig. 4. Signal detection parameters for each experiment. Parameters are presented for (**A**) experiment 1 manipulating prior, (**B**) experiment 2 manipulating reward, and (**C**) experiment 3 manipulating sensory uncertainty. (A to C, top row) Perceptual sensitivity (s)—estimated during the model fitting using a Naka-Rushton function—as a function of contrast (x axis) and group (line color). The data are averaged across base rate in (A, top) and reward in (B, top). (A to C, bottom row) Decision criterion (c_d) as a function of contrast (x axis) and group (line color). The data displayed in (A, top row) and (B, top row) are for the unbalanced base rate/reward blocks. All data points and bars show means across participants, and error bars represent ±SE. The sample size consisted of 29 autistic and 46 non-autistic participants in (A, top row), 30 autistic and 45 non-autistic participants in (A, bottom row), 29 autistic and 43 non-autistic participants in (B, top row), 28 autistic and 42 non-autistic participants in (B, bottom row), 28 autistic and 42 non-autistic participants in (C, top row), and 29 autistic and 40 non-autistic participants in (C, bottom row).

specifically in the contrasts 0.004 and 0.016, as indicated by a significant interaction between reward and contrast [$F(1.12, 78.11) = 4.44$, $P = 0.034$, $\eta_p^2 = 0.06$]. Sensitivity did not significantly differ between groups [$F(1, 70) = 3.10$, $P = 0.083$, $\eta_p^2 = 0.04$], although autistic participants showed a tendency toward greater sensitivity. All remaining effects were not significant (see the Supplementary Materials). Therefore, the two groups exhibited similar sensitivity to the contrast manipulation.

Sensory uncertainty manipulation (experiment 3). Sensitivity declined with decreasing contrast [$F(1.16, 79.02) = 119.50$, $P < 0.001$, $\eta_p^2 = 0.63$] (Fig. 4C, top row), and the two groups did not differ in sensitivity [$F(1, 68) = 0.06$, $P = 0.806$, $\eta_p^2 < 0.01$], across levels of contrast [$F(1.16, 79.02) = 0.90$, $P = 0.360$, $\eta_p^2 = 0.01$].

Comparable first-order decision criterion

Prior manipulation (experiment 1). Figure 4A (bottom row) shows the criterion shift as a function of contrast and group in the unbalanced base-rate block. The ANOVA performed on the decision criterion (c_d) showed a greater criterion shift for unbalanced, compared to balanced base rate [$F(1, 73) = 134.13$, $P < 0.001$, $\eta_p^2 = 0.65$]. Furthermore, criterion shift increased with decreasing contrast [$F(1.87, 136.82) = 114.79$, $P < 0.001$, $\eta_p^2 = 0.61$], and this only when base rate was unbalanced, as indicated by the interaction between base rate and contrast [$F(1.87, 136.82) = 114.79$, $P < 0.001$, $\eta_p^2 = 0.61$]. Therefore, participants shifted their decision criterion toward the category with higher base rate probability as contrast decreased, and this in a comparable manner between groups, as the main effect of group was not significant [$F(1, 73) = 1.41$, $P = 0.239$, $\eta_p^2 = 0.02$], as well as all other effects (see the "Additional results" section of Supplementary Results). Autistic individuals performed the categorization task similarly to non-autistics, by exhibiting similar sensitivity and integration of prior information during first-order decisions.

Reward manipulation (experiment 2). Figure 4B (bottom row) shows the criterion shift as a function of contrast and group in the unbalanced reward block. Similar to the base-rate manipulation, participants shifted their decision criterion more in the unbalanced than in the balanced reward block [$F(1, 68) = 21.48$, $P < 0.001$, $\eta_p^2 = 0.24$]. Furthermore, criterion shift increased as contrast decreased [$F(1.29, 87.55) = 12.05$, $P < 0.001$, $\eta_p^2 = 0.15$], and this in a greater manner when reward was unbalanced [$F(1.29, 87.55) = 12.05$, $P < 0.001$, $\eta_p^2 = 0.15$], with no difference between groups [$F(1, 68) = 0.01$, $P = 0.911$, $\eta_p^2 < 0.001$]. All other effects were not significant (see the "Additional results" section of Supplementary Results).

Sensory uncertainty manipulation (experiment 3). Figure 4C (bottom row) shows the criterion shift as a function of contrast and group in task 2. The ANOVA performed on c_d revealed that c_d increased as contrast decreased [$F(2.61, 175.17) = 332.01$, $P < 0.001$, $\eta_p^2 = 0.83$]. The two groups did not differ in the overall shift of decision criterion [$F(1, 67) = 0.27$, $P = 0.608$, $\eta_p^2 < 0.01$], across contrast levels [$F(2.61, 175.17) = 1.78$, $P = 0.160$, $\eta_p^2 = 0.03$].

Overall, these results confirm that in each experiment, the first-order decision reflects the specific contribution of one Bayesian component—prior, reward, or sensory uncertainty—to perceptual inference. Moreover, autistic individuals performed the first-order task similarly to non-autistics. They exhibited comparable sensitivity to the orientation distributions and integrated all Bayesian components to the same extent. Although the autistic group showed a trend toward enhanced sensitivity in the reward experiment, meta-uncertainty, the estimate of perceptual metacognitive abilities, is independent of first-order performance. Therefore, marginal differences in the first-order tasks cannot account for the observed meta-uncertainty results.

Comparable confidence criterion across groups

In each experiment, the model estimated three confidence criteria, c_c, per participant, representing the amount of internal evidence a participant requires to report increasing levels of confidence. Specifically, c_c-low marks the boundary between the low and medium-low confidence keys, c_c-medium between medium-low and medium-high, and c_c-high between medium-high and high. A higher c_c value indicates a more conservative threshold, meaning that the participant requires stronger evidence to shift the confidence level. Figure 5 illustrates the change in confidence criterion as a function of confidence level and group. In Fig. 5 (A and B), values are reported across base rate and reward blocks.

Prior manipulation (experiment 1). The ANOVA performed on c_c revealed that the confidence criterion increased as the confidence level increased [$F(1.51, 113.24) = 89.85$, $P < 0.001$, $\eta_p^2 = 0.55$]—where c_c-high was significantly higher than c_c-medium [$t(76) = 10.3$, $P < 0.001$], and c_c-medium was higher than c_c-low [$t(76) = 4.25$, $P < 0.001$] (Fig. 5A)—indicating that participants adopted more conservative thresholds when reporting higher confidence levels, demonstrating an appropriate use of the confidence rating scale. All remaining effects, including the effects of group [$F(1, 75) = 3.45$, $P = 0.068$, $\eta_p^2 = 0.04$] and base rate [$F(1, 75) = 0.02$, $P = 0.898$, $\eta_p^2 < 0.01$], were not significant (see the "Additional results" section of Supplementary Results).

Reward manipulation (experiment 2). The ANOVA performed on the c_c showed that the confidence criterion was more conservative as the confidence level increased [$F(1.57, 112.95) = 51.11$, $P < 0.001$, $\eta_p^2 = 0.42$] (Fig. 5B), with a greater confidence criterion for c_c-high compared to c_c-medium [$t(73) = 7.55$, $P < 0.001$] and c_c-medium compared to c_c-low [$t(73) = 3.48$, $P < 0.001$]. The confidence criterion did not vary between groups [$F(1, 72) = 2.38$, $P = 0.128$, $\eta_p^2 = 0.03$] and reward [$F(1, 72) = 2.79$, $P = 0.105$, $\eta_p^2 = 0.04$]. All other effects were not significant (see the "Additional results" section of Supplementary Results).

Sensory uncertainty manipulation (experiment 3). The ANOVA performed on c_c showed that the confidence criterion increased with the confidence level [$F(1.30, 90.83) = 11.05$, $P < 0.001$, $\eta_p^2 = 0.14$] (Fig. 5C), with a greater confidence criterion for c_c-high compared to c_c-medium [$t(71) = 3.48$, $P = 0.003$] and c_c-low [$t(71) = 3.48$, $P = 0.003$], while c_c-low and c_c-medium did not differ [$t(71) = 1.23$, $P = 0.672$]. The confidence criterion did not vary between groups [$F(1, 70) = 0.01$, $P = 0.965$, $\eta_p^2 < 0.01$] or between confidence levels and groups [$F(1.30, 90.83) = 0.45$, $P = 0.544$, $\eta_p^2 < 0.01$]. Therefore, both groups similarly adjusted their confidence criterion by adopting a more conservative c_c when reporting higher confidence in their categorization in all experiments, demonstrating a similar use of the confidence rating scale.

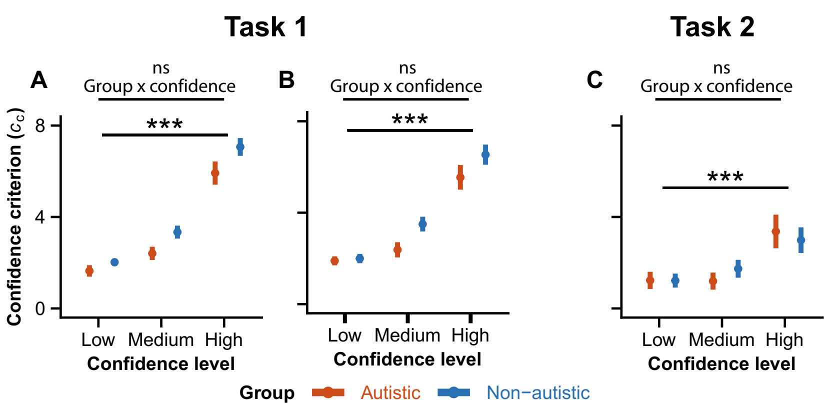

Fig. 5. Confidence criterion for each experiment. The parameter is presented for (**A**) experiment 1 manipulating prior, (**B**) experiment 2 manipulating reward, and (**C**) experiment 3 manipulating sensory uncertainty. Confidence criterion (c_c) as a function of confidence level (x axis) and group (bar color). Confidence levels reflect the position of the criterion between two adjacent confidence keys: low/medium-low, medium-low/medium-high, and medium-high/high, respectively. The data are averaged across base rates in (A) and reward in (B). Horizontal bars show means across participants, and vertical bars represent ±SE. The main effect of confidence level and the interaction between group and confidence level were evaluated using mixed-design ANOVAs. ns: $P \geq 0.05$ and $^{***}P < 0.001$. The sample size consisted of 31 autistic and 46 non-autistic participants in (A), 30 autistic and 44 non-autistic participants in (B), and 29 autistic and 41 non-autistic participants in (C).

Meta-uncertainty in autism depends on first-order Bayesian source of uncertainty

Critically, we analyzed the model estimation of meta-uncertainty (σ_m)—referring to the variability (uncertainty) in estimating the internal noise of the first-order decision variable—the estimate of perceptual metacognitive abilities (see the "Computational model" and the "Data analyses" sections of Materials and Methods). A high σ_m value is associated with higher meta-uncertainty and hence lower metacognitive ability. Normality analyses of meta-uncertainty revealed that the distributions significantly deviated from normality across all experiments. We used $\log(\sigma_m)$ in our analysis as meta-uncertainty expresses a ratio of two SDs, and ratios are inherently logarithmic (see fig. S2). Consistent with Boundy-Singer *et al.* (*13*), σ_m values distributed in two clusters: A small boundary cluster ($0.1 < \sigma_m < 0.2$) concentrated near the lower boundary (0.1) imposed by the model, hence, may reflect less reliable estimation of meta-uncertainty. The main cluster ($0.2 < \sigma_m$) had an approximately normal distribution. In all experiments, frequency analyses show no difference across groups in the probability of these clusters, indicating that unreliable estimation of meta-uncertainty did not vary between autistic and non-autistic participants. See Supplementary Methods for details about the clustering procedure and Supplementary Results and fig. S2 for the distribution analyses. All following analyses were performed on the main cluster, on the log-transformed meta-uncertainty values. One participant from the autistic group was removed from the analysis here due to an extreme value (>2.5 SD). Notably, including this participant does not change the pattern of results and significant effects (see the "Robustness check—Analyses of all participants including outliers" section of Supplementary Results and fig. S3).

Meta-uncertainty was lower for autistic compared to non-autistic participants, reflecting enhanced metacognitive sensitivity, specifically when decision behavior depended on sensory uncertainty alone. Figure 6 illustrates meta-uncertainty as a function of experiment and group, and means are reported across base rate and reward blocks. A 3 × 2 factor ANOVA with Bayesian component manipulation experiment (prior, reward, and likelihood) and group revealed that group differences between autistic and non-autistic in meta-uncertainty were specific to Bayesian component, as indicated by the significant interaction between experiment and group [$F(2, 190) = 4.23$, $P = 0.016$, $\eta_p^2 = 0.04$] (Fig. 6). Planned two-sample *t* test comparisons showed that the autistic group had lower meta-uncertainty compared to the non-autistic group in the sensory uncertainty experiment [$t(49) = 3.27$, $P = 0.002$, $d = 0.90$], but not in the prior nor in the reward experiments [$t(73) = 0.74$, $P = 0.461$] and [$t(68) = 0.12$, $P = 0.897$], respectively. We performed a contrast analysis comparing the meta-uncertainty in the sensory uncertainty experiment to the average meta-uncertainty in the prior and reward experiments (0.5 × prior + 0.5 × reward − sensory uncertainty). The analysis revealed that, in the autistic group, meta-uncertainty was significantly lower in the sensory uncertainty experiment, compared to the prior and reward experiments [$t(190) = 2.61$, $P = 0.009$]. The difference was not significant in the non-autistic group [$t(190) = 1.34$, $P = 0.179$]. See Supplementary Results and fig. S5 for investigation of the effect of prior and reward within individual experiments.

Because many participants participated in experiment 3 after completing experiment 1 and/or 2, we aimed to control for a possible training effect in experiment 3. Therefore, to directly investigate if participation in previous experiments could improve metacognitive abilities, we tested whether meta-uncertainty varied between and within groups as a function of familiarity with the confidence report

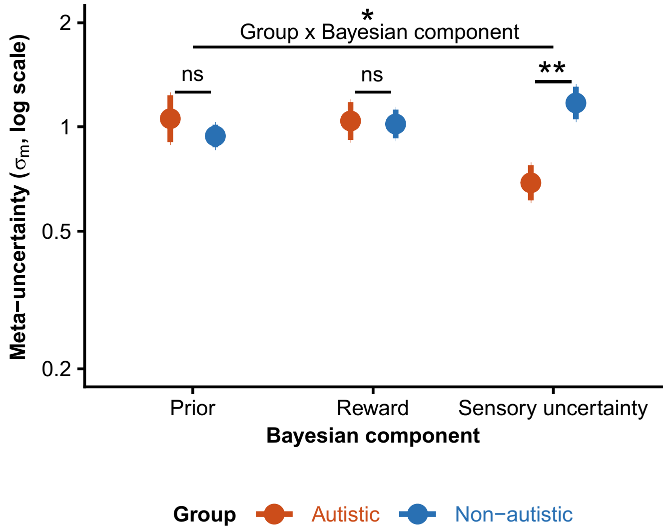

Fig. 6. Meta-uncertainty results per experiment (Bayesian component). The log-transformed meta-uncertainty σ_m (*y* axis) as a function of Bayesian component (*x* axis) and group (color). Data points show the group means per experiment, and vertical bars represent ±SE. The interaction between group and Bayesian component was evaluated using a between-subject ANOVA. The effects of group per experiment were evaluated using unpaired *t* tests. ns: $P \geq 0.05$, *$0.05 > P \geq 0.01$, and **$0.01 > P \geq 0.001$. The sample size consisted of 29 autistic and 46 non-autistic participants in the prior experiment, 29 autistic and 41 non-autistic participants in the reward experiment, and 21 autistic and 30 non-autistic participants in the sensory uncertainty experiment.

(i.e., whether participants performed experiment 1 or 2 before completing experiment 3). Results showed no effect of experience on meta-uncertainty in either group (see Supplementary Results and fig. S4), suggesting that group differences in meta-uncertainty cannot be explained by experimental training.

DISCUSSION

Alterations in self-monitoring and evaluation have been proposed to play a key role in several neurodevelopmental conditions (*2*, *3*), including autism (*25*). However, it remains unclear whether atypical self-monitoring in autism arises from differences in metacognitive ability and confidence bias or is confounded by differences in first-order decisions. Using a computational modeling approach, we independently quantified these processes and identified a fundamental divergence in the factors that determine metacognitive ability in autistic versus non-autistic individuals.

Both autistic and non-autistic participants associated confidence with sensory uncertainty and adjusted their confidence criteria similarly. However, a key group difference emerged in meta-uncertainty—the computational estimate of uncertainty about internal noise. While meta-uncertainty remained stable in non-autistic participants, it varied in autistic participants depending on task manipulations of Bayesian components, as metacognitive ability was enhanced (i.e., lower meta-uncertainty) when decisions relied solely on sensory evidence. Notably, group differences were specific to confidence reports. Perceptual sensitivity and decision criteria were comparable across groups, and both groups demonstrated similar integration of Bayesian component during first-order perceptual decision-making, as confirmed by comparisons with an ideal observer model in previous studies (*18*–*20*). Together with the computational modeling results, these findings suggest atypical metacognitive abilities, even when sensory processing and first-order decision-making are indistinguishable.

In non-autistic individuals, metacognitive ability is typically considered a domain-general capacity—stable across tasks and sensory modalities (*26*–*29*). This consistency is also reflected in meta-uncertainty, which varies between individuals but is strongly correlated within individuals across sessions (*13*). Our findings replicate this pattern: the average meta-uncertainty in the non-autistic group remained consistent across tasks. By contrast, metacognitive monitoring abilities in autism are context dependent. Rather than exhibiting a global reduction, autistic participants appear to monitor their own decisions more accurately when those decisions depend directly on sensory evidence than when prior knowledge or reward information influences the decision process. This suggests a reduced weighting of non-sensory information in self-evaluation—an effect that corresponds with claims of attenuated contextual integration in autistic perception (*17*, *30*). However, the present study, together with Fazioli *et al.* (*18*), reveals that first-order perceptual decisions are intact in autism, and atypical inference processes may rather emerge at the metacognitive level.

Metacognition and decision-making are dynamically coupled: Confidence shapes learning, exploration, and behavioral adaptation (*31*, *32*) and affects subsequent choices and behavior (*33*, *34*). Thus, differences in metacognitive ability may partly explain downstream behavioral differences. This is particularly relevant to sensory-related reactions in autism—an area that has received increasing attention in recent years—as atypical sensory reactivity, such as hypo- or hyperresponsiveness to sensory stimuli, has become recognized as a core feature of the autistic phenotype. While previous research has predominantly focused on sensory sensitivity and first-order decision-making (*16*, *17*, *35*, *36*), and some studies have proposed differences in higher-level expectations and priors (*30*, *37*, *38*), less attention was given to second-order perceptual processes, and the mechanisms underlying sensory reactivity in autism remain poorly understood. Our findings suggest that metacognitive monitoring capabilities may play an important and previously underappreciated role in shaping how autistic individuals engage with sensory input. Specifically, the reduced noise in estimating knowledge based on sensory evidence demonstrates a more accurate monitoring of sensory information in autistic individuals, suggesting a stronger bias toward sensory information at the metacognitive level. These findings correspond to the claim of enhanced perception in autism (*36*). However, we suggest that enhanced abilities emerge at the second-order level rather than first-order sensitivity.

This bias toward sensory information may affect high- and low-order processes in decision-making, such as monitoring decisions that integrate priors and therefore adjusting decision-making strategies based on prior information. Notably, results showing atypical use of prior information, such as reduced prior updating in autism during first-order decisions (*39*, *40*), may not necessarily originate in atypical first-order inference, but rather reflect reduced influence of contextual information during high-level perceptual processes due to a bias toward sensory information at a metacognitive level. In particular, if autistic individuals rely more on sensory information when monitoring perceptual decision, contextual changes may not elicit the change in decision confidence that signals the need for adapting decision strategies through prior updating. Furthermore, a reduced ability to update priors based on environmental changes can impair the capacity to form predictive models of the environment, leading to an overestimation of volatility. This interpretation aligns with recent theoretical accounts proposing overestimation of volatility as a key feature of perceptual processing in autism (*37*, *38*).

Moreover, the present findings highlight the promise of this mechanistic framework to independently quantify first- and second-order perceptual processes—a notable contribution in developmental condition research. This computational approach of metacognitive ability can capture individual differences overlooked by traditional accuracy-based measures, offering a framework to link self-monitoring with neural, developmental, and behavioral outcomes.

Last, these results point to metacognition as a bridge for integrating perceptual processing and social accounts of autism. Metacognitive monitoring underpins one's capacity to interpret not only one's own mental states but also those of others; impairments in this domain may therefore contribute to challenges in communication and perspective-taking (*41*–*43*). Understanding how metacognition operates in autism may thus illuminate the developmental relationship between introspection, self-awareness, and theory of mind. This approach is relevant to the study of a broader range of mental disorders, as alterations in metacognitive computations are considered to play a critical role in many forms of psychopathology (*44*).

MATERIALS AND METHODS

This study is based on data from a three-experiment project investigating perceptual decision-making in autism using orientation categorization tasks. Analyses of first-order decisions (i.e., category choice) were reported in Fazioli *et al.* (*18*–*20*). The present article focuses on confidence data that have not been previously reported. The present study uses the same experimental procedure and task design as in those articles.

Participants

This study included 52 adults diagnosed with autism (41 males and 11 females) and 93 non-autistic individuals (18 males and 75 females). Participants chose between receiving monetary compensation (40 shekels/hour) or university credits (3 credits/hour). Autistic participants were recruited from a pool of participants regularly involved in research at the Department of Special Education. The two groups did not differ in age [$t(105) = 0.55$, $P = 0.59$], with a mean of $m = 26.70$ years old (SE = 0.86) for the autistic group and $m = 27.30$ (SE = 0.64) for the non-autistic group. Intellectual Quotient (IQ) was assessed using the Test of Non-Verbal Intelligence 4, which measures cognitive functioning independent of language skills (*45*). The groups did not differ in IQ [$t(60.3) = 0.90$, $P = 0.37$], with a mean of $m = 99.3$ (SE = 11.40) for the autistic group and $m = 101.0$ (SE = 9.72) for the non-autistic group. Autistic traits were measured using the Autistic Quotient (AQ) questionnaire, and a t test [$t(64.9) = 6.97$, $P < 0.001$] revealed a significantly higher AQ score for the autistic group, $m = 27.0$ (SE = 8.11), compared to the non-autistic group, $m = 16.7$ (SE = 6.69). For each participant, we maintained a minimum 24-hour interval between experiments or experimental sessions.

The autism diagnosis was confirmed using standardized clinical assessments, including the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) (*4*), the Autism Diagnostic Interview (i.e., ADI-R52), and the Autism Diagnostic Observation Schedule (i.e., ADOS-2). All participants completed the Community

Assessment of Psychic Experiences (i.e., CAPE) and AQ questionnaires, in their preferred language (Hebrew or English), either during the clinical assessment or after the experimental sessions. Non-autistic individuals with a history of epilepsy or learning disorders were excluded from the study, as well as individuals diagnosed with autism who have known genetic disorders (e.g., Down syndrome).

Ethics statements

The three experiments received ethical clearance from the Institutional Review Board at the University of Haifa under the reference number 046/20, and participants provided written informed consent before every experimental session.

Apparatus and stimuli

Participants were set in a dimly lit room in front of a computer. A chinrest was used to set viewing distance at 57 cm, and participants responded via a keyboard. See the study of Fazioli *et al.* (*18*) for information about the monitor and display background. Experimental design, tasks, and stimuli (Fig. 1A) were based on Qamar *et al.* (*21*), Denison *et al.* (*22*), and Adler and Ma (*23*). All stimuli were presented at the center of the screen. Each trial began with a 500-ms fixation (black circle, 0.2° visual angle), followed by a 50-ms stimulus—a sinusoidal grating (Gabor patch) with a two-dimensional Gaussian envelope (SD = 0.325°, 85% contrast, 3 cycles per degree). In each trial, the grating's orientation was randomly drawn from one of two Gaussian distributions, corresponding to the two stimulus categories (Fig. 1A). Observers were asked to report from which category they thought the stimulus belonged to, based on its orientation, and how confident they were about their answer. Following stimulus onset, they reported both their category choice (category A or B) and their level of confidence using a four-point scale using a single key. The confidence rating scale ranged from high-confidence category A to high-confidence category B (see Fig. 1A). To manipulate sensory uncertainty, we randomly varied stimulus contrast across trials, using seven fixed values (0.004, 0.016, 0.033, 0.093, 0.18, 0.36, and 0.72). The sensory uncertainty manipulation was used to (i) modulate the integration of prior and reward information into the perceptual decision in experiments 1 and 2 and directly investigate the effect of sensory uncertainty on the decision boundaries in experiment 3 and (ii) assess participants' ability to evaluate the reliability of their decision across different sensitivity levels—an estimate of metacognitive ability.

Categories

The stimulus categories were defined by continuous Gaussian orientation distributions. In task 1 (experiments 1 and 2), distributions were centered at $m_A = -4°$ and $m_B = 4°$ (relative to the horizontal line), with SDs of $SD_A = SD_B = 5°$ (Fig. 1B, task 1). In task 2 (experiment 3), we used embedded categories, a design allowing to test how changes in sensory uncertainty only influence perceptual decisions (*21–23*). Here, distributions had identical means, $m_A = m_B = 0°$ (horizontal), but differing SDs, $SD_A = 3°$ and $SD_B = 12°$ (Fig. 1B, task 2). These parameters were selected to yield an optimal accuracy rate of ~80%.

Blocks

Each experiment consisted of three blocks. In experiment 1, we manipulated prior information by explicitly varying category base rate probabilities across blocks, with either balanced (B = 50% and A = 50%) or unbalanced (B = 25% and A = 75% or B = 75% and A = 25%) prior base rate between the two categories. In experiment 2, we manipulated reward information by explicitly varying the number of points awarded for correct answers in each category across blocks, with balanced (B = 2 points and A = 2 points) or unbalanced (B = 1 point and A = 3 points or B = 3 points and A = 1 point) reward value between categories. In both experiments, the balanced block was always performed second. The order of the unbalanced blocks was counterbalanced between participants. In experiment 3, as the sensory uncertainty was the main manipulation, there was no difference between the three experimental blocks.

Procedure and design

Trainings

Each experiment started with extensive category (40 trials) and confidence (40 trials) training, with stimulus displayed for 300 ms at 100% contrast; see the study of Fazioli *et al.* (*18*) for more information.

Main experiment

Participants were explicitly introduced to the variation between each experimental block (e.g., base rate for experiment 1 and reward points for experiment 2) with a verbal explanation. At the beginning of each block, they were informed of the new base rate/reward condition and performed a 40-trial practice session in which they reported both category and confidence. After each response, a text displayed the chosen category, along with correctness auditory feedback. We ensured that participants reached around 70% accuracy during this practice session, reflecting that they were familiar enough with the categories, the response keys, and the block conditions. Then, they completed the block of 280 test trials.

During the test blocks, no trial-to-trial feedback was given to prevent for feedback-based learning and ensure that decision boundaries were generated internally. However, participants received a summary of their categorization accuracy every 50 trials to maintain engagement. In experiments 2 and 3, participants also received information about the number of points earned in the previous 50 trials and the total points accumulated throughout the experiment.

To ensure participants understood the main manipulations (i.e., base rate or reward), a "check question" was randomly introduced. In experiment 1, participants were asked to gamble an amount (0 to 99 cents) on the chances for the next stimulus to belong to a specific category, with the remaining money assigned to the other category. They were informed that their prediction accuracy would influence a bonus added to their original compensation. In experiment 2, participants were asked about how many points they would earn for correctly categorizing a stimulus from a given category. In addition, they were informed that the accumulated number of points earned during the experiment would determine a bonus added to their original compensation. No comprehension checks were needed in experiment 3, so the same gambling question from experiment 1 was used to maintain consistency. A reward system was also implemented to keep the same level of implication as in experiments 1 and 2. Participants were informed that every correct answer was worth two points, and the total accumulated points would determine a bonus added to their compensation. In experiments 1 and 2, participants completed 960 experimental trials over ~50 min. Preliminary data indicated that experiment 3 was more susceptible to noise. Therefore, participants performed two separate sessions of 960 trials, with a minimum 24-hour gap between them.

Data analyses

We used MATLAB (R2024b) to fit the computational model to our data. Statistical analyses were conducted in R (4.4.1).

On the basis of previous analyses, we assumed a symmetry in participants' criterion shift between opposite base rate/reward blocks

(*18*). Therefore, before implementing the model, we converted the responses from blocks where category A had low base rate/reward, to combine the trials with the other unbalanced block for a single model fit. We reversed stimulus category and stimulus responses and multiplied the stimulus orientation by −1.

Outlier removal

In all experiments, participants with an accuracy below 0.6 at the three highest contrast levels and across blocks were excluded from all analyses. In the ANOVA analyzing the effects of Bayesian component and group on meta-uncertainty, we excluded participants whose meta-uncertainty was inaccurately estimated due to the lower boundary imposed by the model (cutoff value: $\sigma_m < 0.02/\log \sigma_m < 0.02 < 1.61$) (see Supplementary Results and fig. S2). Before analyzing each model parameter, we also excluded participants whose value fell below or above 2.5 SD from the group mean.

Computational model

To estimate metacognitive abilities, we fitted a recent computational model (the CASANDRE or confidence as a noisy decision reliability estimate model) (*13*) that was shown to explain well behavioral confidence reports in previous studies using the same basic stimuli and task (*13*). The model estimates meta-uncertainty, a metacognitive parameter reflecting how precisely an individual can assess their decision reliability. The model assumes that on each trial, an observer estimates the reliability of their decision (makes a confidence decision, V_c) by comparing the absolute distance between the decision variable V_d (i.e., strength of sensory evidence) and the contrast-specific decision criterion c_d and normalizing it by $\hat{\sigma}_d$, the estimate of the dispersion of V_d (Eq. 1). Here, V_d is derived from a normal distribution centered on the true stimulus value, with a variability given by sensory noise (i.e., inverse of sensitivity). The absolute difference between V_d and V_c reflects the strength of the decision, with a higher value indicating stronger evidence. Therefore, V_c can be explained as the strength of evidence for the choice, scaled by how noisy the internal system is perceived to be. This framework assumes that observers do not have access to their actual sensory uncertainty and estimate for every decision. This estimate $\hat{\sigma}_d$ is modeled as a random variable drawn from a lognormal distribution with a mean of s_d (i.e., the true sensory noise) and a trial-to-trial variability of s_m. Last, confidence rating was obtained by comparing V_d to a fixed confidence criterion c_c. Therefore, the noise when estimating decision reliability mainly comes from variability in assessing sensory uncertainty, also called meta-uncertainty s_m. A larger σ_m indicates greater variability in estimating internal noise and, therefore, lower metacognitive abilities. In addition, a guess rate (GR) parameter is included to account for random response.

$$V_c = \frac{|V_d - c_d|}{\hat{\sigma}_d} \tag{1}$$

For each participant in each contrast level, the model estimates sensitivity (s) and decision criterion (c_d). In addition, for each participant across all contrast levels, the model estimates guess rate (GR), confidence criterion (c_c), and meta-uncertainty (σ_m).

To apply the model, we modeled sensitivity using a Naka-Rushton function, defined by a maximum sensitivity R_{max}, a semisaturation constant C_{50}, and a slope n (Eq. 2). Therefore, the model contained 15 free parameters: GR, meta-uncertainty (σ_m), three sensory parameters (R_{max}, C_{50}, and n), seven decision criteria (c_d), and three confidence criteria (c_c, one less than the number of confidence levels). Each parameter was optimized to minimize negative log-likelihood, using the MATLAB fmincon function. The fitting procedure was structured in three nested loops, and the best fit was selected on the basis of the lowest log-likelihood

$$s = \frac{R_{max} * C^n}{C^n + {C_{50}}^n} \tag{2}$$

The starting values of GR, s_m, R_{max}, C_{50}, n, and c_c were respectively 0.01, 0.5, 1.5, 0.3, 2, and 1. We defined strict lower and upper bounds for each parameter to ensure valid estimates: $0 < \text{GR} < 0.1$; $0.1 < \sigma_m < 5$; $0.005 < R_{max} < 5$, $0.5 < n < 5$; $0.005 < C_{50} < 1$; $0 < c_c < 10$.

In task 1, the starting values of the decision criteria c_d were set at 0. For the trials from unbalanced blocks (i.e., unequal reward or prior between categories), the boundaries were $-20 < c_d < 20$. For the trials from the balanced block (i.e., equal reward or prior between categories), the boundaries were $-0.002 < c_d < 0.002$, as the criterion was not expected to vary when sensory evidence decreased, to reduce the number of free parameters to 12.

In task 2, the observer sets two decision boundaries to distinguish between the narrow category A and the broad category B (Fig. 1C, task 2). To reduce the number of free parameters, we assumed these boundaries to be symmetrical around zero degrees. Therefore, we estimated the lower value of c_d for each contrast and multiplied it by minus one to estimate the upper value. The boundaries were set at $-20 < c_d < 0$, and the starting values were set from −5 (high contrast) to −11 (low contrast). Therefore, similar to the unbalanced trials of task 1, we used 15 free parameters for modeling data from task 2.

Model comparison

We started by fitting the original model to the data. In that version, sensitivity was estimated separately for each contrast level (i.e., 7 values), using a signal-detection-theory–like model. For task 1, the model did not conditionally constrain c_d, and for task 2, both boundaries for c_d were estimated. This resulted in a total of 19 free parameters for task 1 and 26 for task 2. To reduce the number of free parameters and minimize the risk of overfitting, we tested three alternative model variants, where estimates for sensitivity and decision criterion were reduced. The model described above, using the Naka-Rushton function to estimate sensitivity, demonstrated the best fit (see Supplementary Results). The results reported here in the main text are based on the parameters extracted from this model. The description and comparison between models are provided in Supplementary Methods, Supplementary Results, and fig. S1.

Statistical analyses

Categorization task. For each experiment, we investigated how the category choice varied across contrast and orientations. For experiments 1 and 2, we conducted 2 × 7 × 11 × 2 linear mixed-effect models with group (non-autistic, autistic) as a between-subject factor and contrast (0.004, 0.016, 0.033, 0.093, 0.18, 0.36, and 0.72), orientation (−10, −8, −6, −4, −2, 0, 2, 4, 6, 8, and 10), and base rate/reward (balanced and unbalanced) as within-subject factors, on the proportion of reporting category B. For experiment 3, we conducted a 2 × 7 × 11 linear and quadratic mixed-effect model to account for the V-shaped pattern of response. The model included group (non-autistic and autistic) as a between-subject factor and contrast (0.004, 0.016, 0.033, 0.093, 0.18, 0.36, and 0.72) and orientation (−10, −8, −6, −4, −2, 0, 2, 4, 6, 8, and 10) as within-subject factors, as well as the squared orientation factor, and was performed on the proportion of reporting category B (see Supplementary Results and table S1).

Confidence task. We investigated how the confidence report was influenced by the different manipulations for each experiment. In experiments 1 and 2, we conducted 2 × 7 × 11 × 2 linear and quadratic mixed-effect models with group (non-autistic and autistic) as a between-subject factor and contrast (0.004, 0.016, 0.033, 0.093, 0.18, 0.36, and 0.72), orientation (−10, −8, −6, −4, −2, 0, 2, 4, 6, 8, and 10), and base rate/reward block (balanced and unbalanced) as within-subject factors, on the confidence report. A squared orientation factor was added in each model to account for the V-shaped behavior. In experiment 3, we conducted a 2 × 7 × 11 linear mixed-effect model with group (non-autistic and autistic) as a between-subject factor and contrast (0.004, 0.016, 0.033, 0.093, 0.18, 0.36, and 0.72) and orientation (−10, −8, −6, −4, −2, 0, 2, 4, 6, 8, and 10) as within-subject factors, on the confidence report.

Sensitivity and decision criterion. We used the parameters from the Naka-Rushton fitting to estimate the sensitivity s_i for each level of contrast C_i for each participant (Eq. 3). In experiments 1 and 2, we performed 2 × 2 × 7 mixed-design ANOVAs with group (non-autistic and autistic) as a between-subject factor and prior/reward (balanced and unbalanced) and contrast (0.004, 0.016, 0.033, 0.093, 0.18, 0.36, and 0.72) as within-subject factors on the s and c_d. In experiment 3, we performed 2 × 7 mixed-design ANOVAs with group (non-autistic and autistic) as a between-subject factor and contrast (0.004, 0.016, 0.033, 0.093, 0.18, 0.36, and 0.72) as a within-subject factor on the s and c_d

$$s_i = \frac{R_{\max} * C_i^{\,n}}{C_i^{\,n} + C_{50}^{\,n}} \tag{3}$$

Confidence criterion. Because there were four confidence levels, there were three confidence-level boundaries: 1 = between low and mid-low, 2 = between mid-low to mid-high, 3 = between mid-high and high. To investigate the shift in confidence criterion c_c, we performed a 2 × 3 × 2 mixed-design ANOVA with group (non-autistic and autistic) as a between-subject factor and confidence-level boundaries (1, 2, and 3) and base-rate block (balanced and unbalanced) as within-subject factors on the c_c. In experiment 2, we performed a similar 2 × 3 × 2 mixed-design ANOVA with group, confidence-level boundary, and reward block (balanced and unbalanced). In experiment 3, we performed a 2 × 3 mixed-design ANOVA with group (non-autistic and autistic) as a between-subject factor and confidence level (1, 2, and 3) as a within-subject factor on the c_c.

Meta-uncertainty. To investigate whether meta-uncertainty differed between groups, and whether the type of Bayesian component involved in perceptual decisions affected the metacognitive abilities, we performed a 2 × 3 between-subjects ANOVA with group (non-autistic and autistic) and experiment (prior, reward, and sensory uncertainty) as between-subject factors, on the meta-uncertainty.

To control for potential improvement in metacognitive abilities over time among participants who completed multiple experiments, we directly tested whether meta-uncertainty in experiment 3 varied as a function of familiarity with the task, by conducting a 2 × 2 between-subject ANOVA on the meta-uncertainty in experiment 3, with group (non-autistic and autistic) and previous participation (with and without) as between-subject factors. See Supplementary Results and fig. S4.

To investigate the difference between groups and block conditions (prior and reward) in metacognitive abilities for experiments 1 and 2, we conducted for each experiment a 2 × 2 mixed-design ANOVA with group (non-autistic and autistic) as a between-subject factor and prior/reward block (balanced and unbalanced) as a within-subject factor on the σ_m. The results are described in Supplementary Results and fig. S5.

Guess rate. In experiments 1 and 2, we performed a 2 × 2 mixed-design ANOVA with group (non-autistic and autistic) as a between-subject factor and prior/reward block (balanced and unbalanced) as a within-subject factor on GR. In experiment 3, we conducted an unpaired t test on GR with group (non-autistic and autistic) as the between-subject factor. The results are displayed in the Supplementary Materials.

Significant effects from the ANOVAs were further investigated using paired and unpaired t tests as appropriate to elucidate the nature of the observed differences. Bonferroni corrections were applied to control for multiple comparisons. Effect sizes were calculated using partial eta square (η_p^2) for ANOVAs and Cohen's standardized mean difference (d) for t tests.

In the linear mixed-effects models, t-statistics were computed as the ratio of the estimated coefficient to its SE. Degrees of freedom and P values were obtained using the Satterthwaite approximation as implemented in the R package "lmerTest."

Supplementary Materials

This PDF file includes:
Supplementary Methods
Supplementary Results
Figs. S1 to S37
Tables S1 and S2

REFERENCES

1. J. H. Flavell, Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *Am. Psychol.* **34**, 906–911 (1979).
2. T. X. F. Seow, M. Rouault, C. M. Gillan, S. M. Fleming, How local and global metacognition shape mental health. *Biol. Psychiatry* **90**, 436–446 (2021).
3. T. Wise, O. J. Robinson, C. M. Gillan, Identifying transdiagnostic mechanisms in mental health using computational factor modeling. *Biol. Psychiatry* **93**, 690–703 (2023).
4. American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders: DSM-5* (American Psychiatric Association, ed. 5, 2022).
5. K. L. Carpenter, D. M. Williams, A meta-analysis and critical review of metacognitive accuracy in autism. *Autism* **27**, 512–525 (2023).
6. K. Maras, J. E. Norris, N. Brewer, Metacognitive monitoring and control of eyewitness memory reports in autism. *Autism Res.* **13**, 2017–2029 (2020).
7. K. M. DeBrabander, A. E. Pinkham, R. A. Ackerman, D. R. Jones, N. J. Sasson, Cognitive and social cognitive self-assessment in autistic adults. *J. Autism Dev. Disord.* **51**, 2354–2368 (2021).
8. T. O. Nelson, A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychol. Bull.* **95**, 109–133 (1984).
9. S. M. Fleming, H. C. Lau, How to measure metacognition. *Front. Hum. Neurosci.* **8**, 82285 (2014).
10. P. Mamassian, Visual confidence. *Annu. Rev. Vis. Sci.* **2**, 459–481 (2016).
11. B. Maniscalco, H. Lau, A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Conscious. Cogn.* **21**, 422–430 (2012).
12. S. M. Fleming, HMeta-d: Hierarchical Bayesian estimation of metacognitive efficiency from confidence ratings. *Neurosci. Conscious.* **2017**, nix007 (2017).
13. Z. M. Boundy-Singer, C. M. Ziemba, R. L. T. Goris, Confidence reflects a noisy decision reliability estimate. *Nat. Hum. Behav.* **7**, 142–154 (2023).
14. D. C. Knill, W. Richards, *Perception as Bayesian Inference* (Cambridge Univ. Press, 1996).
15. P. Mamassian, M. Landy, L. T. Maloney, "Bayesian modelling of visual perception," in *Probabilistic Models of the Brain: Perception and Neural Function* (MIT Press, 2002), pp. 13–36.
16. R. P. Lawson, G. Rees, K. J. Friston, An aberrant precision account of autism. *Front. Hum. Neurosci.* **8**, 302 (2014).
17. E. Pellicano, D. Burr, When the world becomes 'too real': A Bayesian explanation of autistic perception. *Trends Cogn. Sci.* **16**, 504–510 (2012).
18. L. Fazioli, B.-S. Hadad, R. N. Denison, A. Yashar, Suboptimal but intact integration of Bayesian components during perceptual decision-making in autism. *Mol. Autism.* **16**, 2 (2025).
19. L. Fazioli, B.-S. Hadad, R. Denison, A. Yashar, Intact Bayesian perceptual decision making and metacognition in autism. *J. Vis.* **23**, 5283 (2023).
20. L. Fazioli, B. S. Hadad, R. N. Denison, A. Yashar, Intact integration of reward information during perceptual decision-making in autism. PsyArXiv 10.31234/osf.io/j3tmx_v1 (2025).

21. A. T. Qamar, R. J. Cotton, R. G. George, J. M. Beck, E. Prezhdo, A. Laudano, A. S. Tolias, W. J. Ma, Trial-to-trial, uncertainty-based adjustment of decision boundaries in visual categorization. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 20332–20337 (2013).
22. R. N. Denison, W. T. Adler, M. Carrasco, W. J. Ma, Humans incorporate attention-dependent uncertainty into perceptual decisions and confidence. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 11090–11095 (2018).
23. W. T. Adler, W. J. Ma, Comparing Bayesian and non-Bayesian accounts of human confidence reports. *PLoS Comput. Biol.* **14**, e1006572 (2018).
24. J. Navajas, B. Bahrami, P. E. Latham, Post-decisional accounts of biases in confidence. *Curr. Opin. Behav. Sci.* **11**, 55–60 (2016).
25. H. M. Bednarz, J. A. Trapani, R. K. Kana, Metacognition and behavioral regulation predict distinct aspects of social functioning in autism spectrum disorder. *Child Neuropsychol.* **26**, 953–981 (2020).
26. N. Faivre, E. Filevich, G. Solovey, S. Kühn, O. Blanke, Behavioural, modeling, and electrophysiological evidence for domain-generality in human metacognition. bioRxiv 095950 [Preprint] (2016). https://doi.org/10.1101/095950.
27. J. Morales, H. Lau, S. M. Fleming, Domain-general and domain-specific patterns of activity supporting metacognition in human prefrontal cortex. *J. Neurosci.* **38**, 3534–3546 (2018).
28. J. Carpenter, M. T. Sherman, R. A. Kievit, A. K. Seth, H. Lau, S. M. Fleming, Domain-general enhancements of metacognitive ability through adaptive training. *J. Exp. Psychol. Gen.* **148**, 51–64 (2019).
29. M. Rouault, A. McWilliams, M. G. Allen, S. M. Fleming, Human metacognition across domains: Insights from individual differences and neuroimaging. *Pers. Neurosci.* **1**, e17 (2018).
30. K. J. Friston, R. Lawson, C. D. Frith, On hyperpriors and hypopriors: Comment on Pellicano and Burr. *Trends Cogn. Sci.* **17**, 1 (2013).
31. D. G. Lee, T. A. Hare, Value certainty and choice confidence are multidimensional constructs that guide decision-making. *Cogn. Affect. Behav. Neurosci.* **23**, 503–521 (2023).
32. N. Yeung, C. Summerfield, Metacognition in human decision-making: Confidence and error monitoring. *Philos. Trans. R. Soc. B* **367**, 1310–1321 (2012).
33. D. Rahnev, A. Koizumi, L. Y. McCurdy, M. D'Esposito, H. Lau, Confidence leak in perceptual decision making. *Psychol. Sci.* **26**, 1664–1680 (2015).
34. B. A. Purcell, R. Kiani, Hierarchical decision processes that operate over distinct timescales underlie choice and changes in strategy. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E4531-40 (2016).
35. B.-S. Hadad, A. Yashar, Sensory perception in autism: What can we learn? *Annu. Rev. Vis. Sci.* **8**, 239–264 (2022).
36. L. Mottron, M. Dawson, I. Soulières, B. Hubert, J. Burack, Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception. *J. Autism Dev. Disord.* **36**, 27–43 (2006).
37. R. P. Lawson, C. Mathys, G. Rees, Adults with autism overestimate the volatility of the sensory environment. *Nat. Neurosci.* **20**, 1293–1299 (2017).
38. S. Van De Cruys, K. Evers, R. Van Der Hallen, L. Van Eylen, B. Boets, L. de-Wit, J. Wagemans, Precise minds in uncertain worlds: Predictive coding in autism. *Psychol. Rev.* **121**, 649–675 (2014).
39. L. Sapey-Triomphe, L. Timmermans, J. Wagemans, Priors bias perceptual decisions in autism, but are less flexibly adjusted to the context. *Autism Res.* **14**, 1134–1146 (2021).
40. R. Twito, B. Hadad, S. Szpiro, Is she still angry? Intact learning but no updating of facial expressions priors in autism. *Autism Res.* **17**, 934–946 (2024).
41. C. Grainger, D. M. Williams, S. E. Lind, Metacognitive monitoring and control processes in children with autism spectrum disorder: Diminished judgement of confidence accuracy. *Conscious. Cogn.* **42**, 65–74 (2016).
42. T. Nicholson, D. M. Williams, S. E. Lind, C. Grainger, P. Carruthers, Linking metacognition and mindreading: Evidence from autism and dual-task investigations. *J. Exp. Psychol. Gen.* **150**, 206–220 (2021).
43. S. M. Fleming, Metacognition and confidence: A review and synthesis. *Annu. Rev. Psychol.* **75**, 241–268 (2024).
44. M. Rouault, T. Seow, C. M. Gillan, S. M. Fleming, Psychiatric symptom dimensions are associated with dissociable shifts in metacognition but not task performance. *Biol. Psychiatry* **84**, 443–451 (2018).
45. M. Goldberg Edelson, S. M. Edelson, S. Jung, Assessing the intelligence of individuals with autism: A cross-cultural replication of the usefulness of the TONI. *Focus Autism Other Dev. Disabil.* **13**, 221–227 (1998).

Acknowledgments: We would like to thank all the participants who took part in this project, as well as M. Hartston for recruiting the autistic participants. **Funding:** We also would like to thank our funding sources: Israel Science Foundation (grant no. 1459/24 to A.Y. and 2025/25 to B.-S.H.) and US-Israeli Binational Science Foundation (grant no. 234/2020 to B.-S.H.). **Author contributions:** Resources, funding acquisition, and supervision: A.Y. and B.-S.H. Conceptualization and methodology: A.Y., B.-S.H., and R.N.D. Investigation: L.F. Data curation: L.F. and A.Y. Formal analyses: L.F. and A.Y. Visualization: L.F. and A.Y. Original draft: L.F. and A.Y. Review and editing: All authors. **Competing interests:** The authors declare that they have no competing interests. **Data, code, and materials availability:** All data and code needed to evaluate and reproduce the results in the paper are present in the paper and/or the Supplementary Materials and are also available in the Open Science Framework (OSF) repository as a registered project: https://doi.org/10.17605/OSF.IO/KHB2R. This study did not generate new materials.

Submitted 27 November 2025
Accepted 9 July 2026
Published 19 August 2026
10.1126/sciadv.aee2332

Perceptual meta-uncertainty reveals enhanced calibration of sensory confidence in autism

Laurina Fazioli, Bat-Sheva Hadad, Rachel N. Denison, and Amit Yashar

Sci. Adv. **12** (34), eaee2332. DOI: 10.1126/sciadv.aee2332

Science Advances (ISSN 2375-2548) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.