

Neural correlates of hierarchical predictive processes in autistic adults

Received: 18 November 2021

Accepted: 8 May 2023

Published online: 19 June 2023

 Check for updates

Laurie-Anne Sapey-Triomphe^{1,2}  , Lauren Pattyn¹, Veith Weilhhammer^{3,4}, Philipp Sterzer^{3,4}  & Johan Wagemans^{1,2} 

Bayesian theories of autism spectrum disorders (ASD) suggest that atypical predictive mechanisms could underlie the autistic symptomatology, but little is known about their neural correlates. Twenty-six neurotypical (NT) and 26 autistic adults participated in an fMRI study where they performed an associative learning task in a volatile environment. By inverting a model of perceptual inference, we characterized the neural correlates of hierarchically structured predictions and prediction errors in ASD. Behaviorally, the predictive abilities of autistic adults were intact. Neurally, predictions were encoded hierarchically in both NT and ASD participants and biased their percepts. High-level predictions were following activity levels in a set of regions more closely in ASD than NT. Prediction errors yielded activation in shared regions in NT and ASD, but group differences were found in the anterior cingulate cortex and putamen. This study sheds light on the neural specificities of ASD that might underlie atypical predictive processing.

Even though we live in a complex and changing environment, most of us manage to minimize uncertainty and therefore avoid being overwhelmed by surrounding stimuli. This is achieved using our predictive abilities which rely on prior beliefs. Priors are accumulated over previous experiences and capture the hierarchical statistical regularities of sensory stimuli. Using a generative model, the brain would act like a predictive engine that constantly tries to anticipate or predict the incoming inputs, and infer their causes^{1–3}. This process has been formalized in the Bayesian brain framework, where the brain implements and updates a hierarchical model of the world^{3–5}. Priors are combined with sensory inputs to generate percepts, and their relative weights depend on their precision (i.e., inverse variance). The prior/sensory precision balance needs to be flexibly adjusted to the context to yield optimal perception. Discrepancies between priors and sensory inputs are signaled through prediction errors, which can be used to update priors or may be ignored when signaling irrelevant noise⁶. Predictive mechanisms are hypothesized to rely on a bidirectional cascade of top-down predictions and bottom-up prediction errors, across hierarchically organized cortical layers and brain regions⁷.

Impaired predictive skills would lead to an atypical perception and a sensation of an unpredictable world. In particular, atypical predictive mechanisms might underlie the symptoms encountered in Autism Spectrum Disorders (ASD)^{8–14}. ASD is characterized by deficits in social interactions and communication, and by restrictive and repetitive behaviors and interests (Diagnostic and Statistical Manual of Mental Disorders, 5th edition, DSM-5¹⁵). Even though this neurodevelopmental condition affects >1% of the population¹⁶, the core mechanisms of ASD remain to be identified. The Bayesian theories of ASD offer a promising framework to account for the heterogeneous symptoms of ASD. Indeed, suboptimal predictive abilities might, for instance, explain the higher uncertainty intolerance in autistic individuals¹⁷ or their difficulties to rapidly build up an internal representation to categorize stimuli¹⁸. More broadly, if ASD is a disorder of perceptual inference¹⁹, it would prevent autistic individuals from accurately inferring the meaning of stimuli or making predictions. The social domain would be particularly impacted as it is a complex, dynamic, and noisy domain, where predictions need to be frequently updated and adjusted to the context. Non-social symptoms, such as

¹Department of Brain and Cognition, Leuven Brain Institute, KU Leuven 3000 Leuven, Belgium. ²Leuven Autism Research (LAuRes), KU Leuven 3000 Leuven, Belgium. ³Department of Psychiatry, Charité-Universitätsmedizin Berlin, 10117 Berlin, Germany. ⁴Berlin Institute of Health, Charité-Universitätsmedizin Berlin, 10178 Berlin, Germany. ✉e-mail: laurie-anne.sapey-triomphe@inserm.fr

repetitive behaviors or insistence on sameness, could be a way to restore some predictability. More specifically, Bayesian theories have suggested a reduced impact of priors in ASD, either because of low prior precision¹¹ or because of high sensory precision⁸, both resulting on a very realistic perception of the world according to these theories. One of the hypotheses focusing on the ratio of prior and sensory precisions^{9,10,13} is that there might be a high and inflexible precision of prediction errors in autism (HIPPEA)¹³.

As detailed in a recent review²⁰, the predictive abilities of autistic individuals are intact or impaired depending on the context. While the perceptual bias induced by structural priors seems typical in ASD (e.g., ref. 21,22), the learning of predictive associations is sometimes impaired²⁰. Associative learning would be particularly difficult in ASD when predictive features have low salience or consistency²⁰. In terms of learning dynamic, autistic individuals were slower at updating predictions²³, which could explain a more inflexible weighting of priors in ASD²⁴. The ability to learn probabilistic associations in volatile environments was intact in autistic children²⁵, but atypical in autistic adults²⁶ who showed less surprise to unexpected outcomes in ASD²⁶, consistently with the hypothesis of more inflexible prediction errors in ASD¹³. This is also in line with a mismatch negativity study showing a less flexible modulation of prediction errors in autistic adults²⁷.

Despite the thought-provoking debate about the Bayesian hypotheses of ASD, very little is known about their neural correlates in ASD. Three fMRI studies recently investigated reward and social prediction errors in ASD. In a false belief paradigm, social prediction errors were atypically encoded in the gyral surface of the anterior cingulate cortex (ACC) of autistic adults²⁸. In two studies investigating the neural correlates of reward prediction errors, group differences between neurotypicals (NT) and autistic individuals were found in the paracingulate gyrus, insula, and frontal pole²⁹ and in the ACC and frontal regions³⁰. Yet, note that these two last studies did not identify neural correlates of prediction errors within groups. The neural mechanisms underlying predictions and prediction errors in ASD should be characterized in order to shed light on the predictive processes in ASD.

Unlike the literature on ASD, the neural correlates of predictive mechanisms have been quite extensively investigated in NT^{31–36}. Predictions and prediction errors are thought to pass between superficial and deeper cortical layers within each brain region (e.g., ref. 37), and some key regions have been identified. For instance, when processing contextual associations, the parahippocampal and retrosplenial cortices represent familiar associations at different levels of abstraction, while the orbitofrontal cortex (OFC) updates the internal

representation of the current context^{38,39}. More broadly, a recent meta-analytic approach³⁶ highlighted the role of the inferior frontal gyrus (IFG) and insula in signaling both prediction errors and predictions. A few model-based fMRI studies investigated hierarchically structured predictions and prediction errors^{31–33}. In two associative learning studies^{31,32}, lower-level prediction errors involved the dopaminergic midbrain, and visual, frontal, parietal, and cingulate regions, while higher-level prediction errors correlated with activity in the cholinergic basal forebrain. In another associative learning task³³, higher-level predictions involved regions such as the OFC or hippocampus, whereas lower-level predictions only triggered activity in the visual cortex. The advantage of these model-based fMRI studies is that they allow to identify predictions or prediction errors at different levels of the hierarchy, on a trial-by-trial basis and with individualized trajectories of the model parameters.

The aim of the present study was to characterize the brain regions involved in signaling hierarchical predictions and prediction errors in NT and autistic adults, and to assess whether there were between-group differences. For this purpose, we used model-based fMRI and the same experimental paradigm as in the study by Weilhhammer et al.³³. NT and autistic adults performed a crossmodal associative learning task, where a high- or a low-pitch tone was predictive of a clockwise or counterclockwise rotation of two dots, respectively (Fig. 1). In a similar behavioral task, but with a low contingency (i.e., 62.5% of expected outcomes), autistic adults did not update their predictions when the association reversed and were less biased by priors²⁴. Here, we used a higher probabilistic association (i.e., 75% as in³³) to increase the chances that all participants would learn predictions, so that we could identify their neural correlates and pinpoint genuine group differences in neural activation.

The paradigm included three forms of uncertainty. First, sensory uncertainty was introduced with a subset of ambiguous trials where the two dots did not rotate, but simply jumped from their vertical to their horizontal position. In these trials, participants tend to report a rotation direction that is consistent with the main contingency, suggesting that they are biased by their priors^{24, 33}. Second, expected uncertainty was introduced by the probabilistic association between the tone and the rotation direction (i.e., only 75% of the trials showed the expected rotation). Third, the task included unexpected uncertainty, as the tone-rotation association could suddenly reverse. We modeled the behavioral data using a three-level Hierarchical Gaussian Filter⁴⁰, which offers a way to model perception in uncertain contexts using a hierarchical generative model of the environment. We correlated the trial-wise estimates of the model parameters signaling

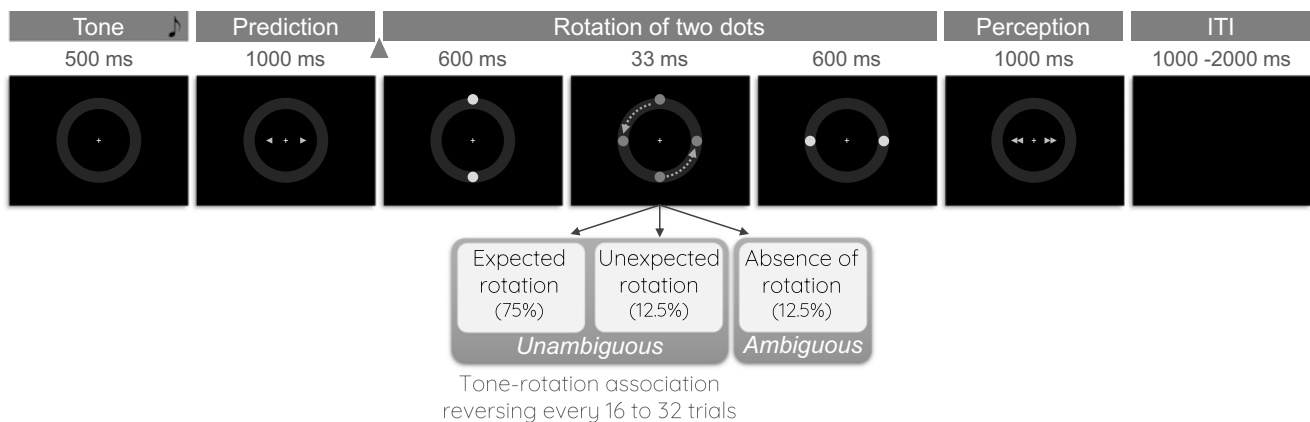


Fig. 1 | Experimental paradigm. After hearing a low- or a high-pitch tone, participants had to predict the rotation direction of a pair of dots, and to report their percept. There was a probabilistic association between the tone and the rotation direction (main contingency: 75%), reversing every 16, 24, or 32 trials. In a subset of

ambiguous trials, the dots did not rotate but simply appeared in their vertical and then horizontal position. CW: Clockwise, CCW: Counterclockwise, ITI: Inter-trial interval. The triangle indicates a jitter of 100 to 300 ms.

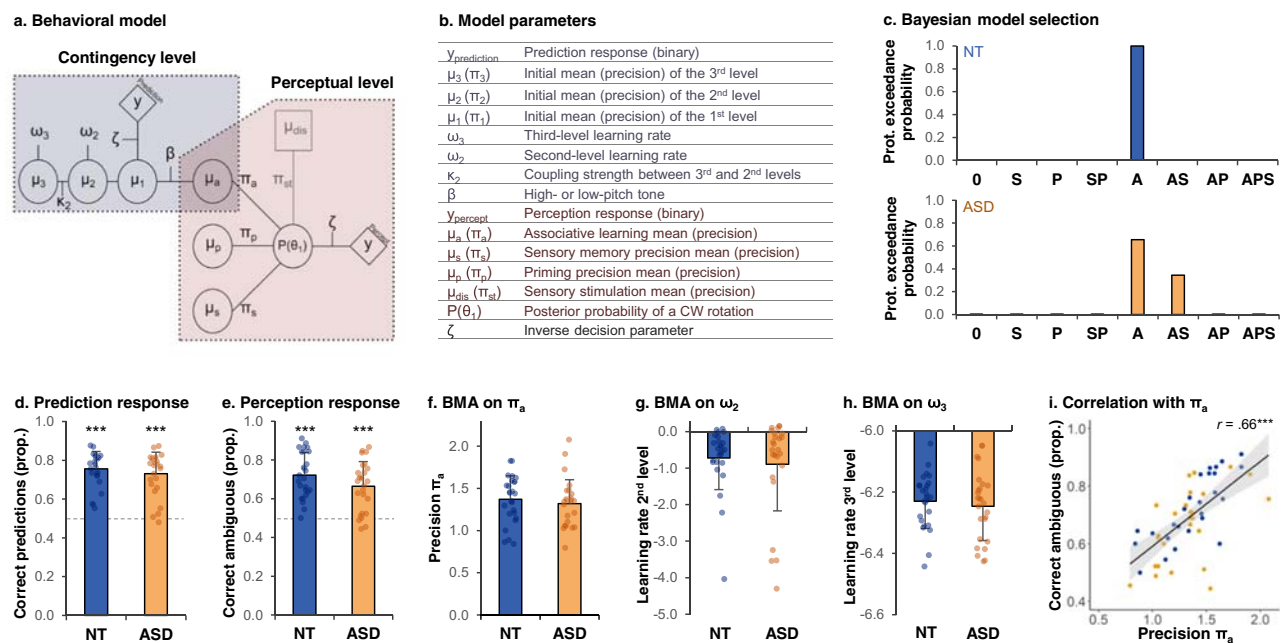


Fig. 2 | Behavioral and modeling results. **a** Schematic representation of the model, adapted from Weinhhammer, V. A., Stuke, H., Sterzer, P. & Schmack, K. The Neural Correlates of Hierarchical Predictions for Perceptual Decisions. *J. Neurosci. Off. J. Soc. Neurosci.* 38, 5008–5021 (2018)³³. **b** Model parameters. **c** Bayesian Model Selection comparing 8 models that included an influence of sensory memory (S) and/or priming (P) and/or associative learning (A) or none (O). Top/Bottom graphs: protected exceedance probability in the NT/ASD groups. **d** Proportion (prop.) of correct prediction responses in the NT group (mean = 0.76%) and ASD group (mean = 0.73%). **e** Proportion of ambiguous trials perceived according to the current contingency in the NT group (mean = 0.72%) and ASD group (mean = 0.66%). **f** Associative learning precision π_a , estimated using BMA, in the NT group

(mean = 1.37) and ASD group (mean = 1.32). **g** 2nd level learning rate ω_2 , estimated using BMA, in the NT group (mean = -0.72) and ASD group (mean = -0.89). **h** 3rd level learning rate ω_3 , estimated using BMA, in the NT group (mean = -6.23) and ASD group (mean = -6.25). **i** Correlation between π_a and the percentage of ambiguous trials perceived according to the current contingency. The error band represents the confidence interval. Correlations were assessed using Pearson's correlation test. In the histograms **d–h**, error bars indicate standard deviations, and groups were compared using Student's *t* tests (two-sided, not adjusted for multiple comparisons). Blue: NT participants ($n = 26$). Orange: ASD participants ($n = 26$). *** $p < 0.001$.

predictions and prediction errors with the fMRI time-courses. We also used Dynamic Causal Modeling to investigate how top-down and bottom-up connections were modulated by predictions and prediction errors, respectively. Finally, autistic traits and symptoms were assessed using questionnaires to better characterize our sample and to investigate whether these symptoms could contribute to the group differences.

In NT, we hypothesized that the neural correlates would be in line with the literature^{31,33,41}. We predicted a rather large overlap between NT and ASD regarding the set of regions encoding predictions, as several behavioral studies found an intact influence of priors in ASD^{21,22,42–45}. Given that in a similar paradigm, autistic individuals updated their predictions less than NT²⁰, differences might be found in regions involved in learning and adjusting predictions, such as the OFC, IFG, retrosplenial cortex, or (para)hippocampus. In general, group differences are more likely to be found at the higher levels of the hierarchy than at lower levels, as suggested by behavioral results interpreted as an overestimation of the volatility in ASD²⁶ and by fMRI results showing a similar effect of visual illusions on the primary visual cortex in NT and ASD⁴⁶. The weak prior hypothesis¹¹ would predict a decreased activation of regions encoding priors, and a less proactive brain in ASD. Specifically, proactive processes at the neural level were evidenced in NT³³ with this paradigm, as activity in the retinotopic areas of the dot trajectories was elicited when hearing a tone. Regarding the neural correlates of prediction errors, the HIPPEA hypothesis would predict a stronger reaction to precision-weighted prediction errors in ASD in regions such as the ACC, caudate nucleus, putamen, basal forebrain, or anterior insula.

Results

Behavioral results

Both groups got percentages of correct prediction responses above chance level: 76% (± 9) in NT ($t(25) = 14.6$, $p < 0.0001$, $d = 2.86$) and 73% (± 11) in ASD ($t(25) = 10.7$, $p < 0.0001$, $d = 2.10$) (Fig. 2d). These percentages did not significantly differ between groups ($t(48) = 0.88$, $p = 0.38$, $d = 0.24$).

Perception responses were analyzed in unambiguous and ambiguous trials, separately. In unambiguous trials, both groups got 99% (± 1) of correct perception responses. In ambiguous trials, the percentage of trials perceived as following the current contingency was 72% (± 12) in the NT group and 66% (± 13) in the ASD group (no significant group difference despite a trend, $t(50) = 1.7$, $p = 0.095$, $d = 0.47$) (Fig. 2e). These percentages were above chance level in the NT group ($t(25) = 9.7$, $p < 0.0001$, $d = 1.90$) and ASD group ($t(25) = 6.6$, $p < 0.0001$, $d = 1.29$), suggesting that they were both biased by their expectations.

Finally, in the confidence rating task, a repeated-measure ANOVA on confidence rating scores showed a main effect of ambiguity ($F(1,50) = 40.0$, $p < 0.0001$), with ambiguous trials being rated as more uncertain than unambiguous trials (60% \pm 30 certain in ambiguous trials vs. 90% \pm 14 certain in unambiguous trials, $t(51) = 6.4$, $p < 0.0001$, $d = 0.88$), but revealed no group effect nor interaction.

Behavioral modeling

After inverting the eight models and performing BMS, group-level inference showed that the *Associative learning* model (A) best explained the data in both groups (Fig. 2c). Indeed, the protected

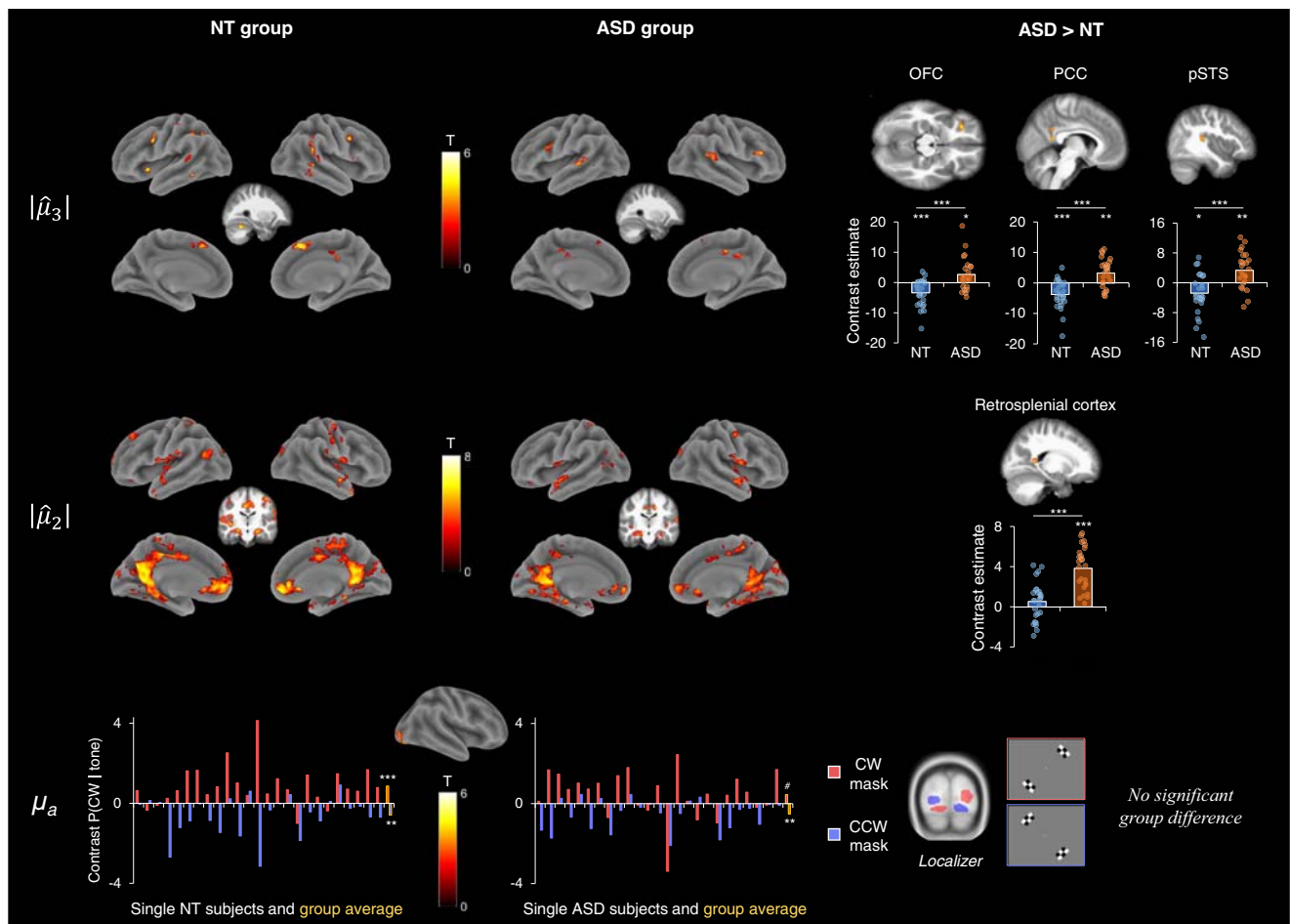


Fig. 3 | Neural correlates of the prior mean. Top row: positive effect of the 3rd level prior mean $|\hat{\mu}_3|$ in the NT and ASD groups, and group differences in the left orbitofrontal cortex (OFC), bilateral posterior cingulate cortex (PCC, note that the two PCC clusters were merged into one region for plotting the contrast estimate) and right posterior superior temporal sulcus (pSTS). OFC histogram: one-sample t test in NT: $p < 0.001$, in ASD: $p = .023$; two-sample t test between groups: $p < 0.001$. PCC histogram: one-sample t test in NT: $p < 0.001$, in ASD: $p = 0.002$; two-sample t test: $p < 0.001$. pSTS histogram: one-sample t test in NT: $p = 0.014$, in ASD: $p = 0.002$; two-sample t test: $p < 0.001$. NT group: $n = 26$, ASD group: $n = 25$. Middle row: Positive effect of the 2nd level prior mean $|\hat{\mu}_2|$ in the NT and ASD groups, and group differences in the right retrosplenial complex (overlapping the precuneus and retrosplenial cortex). Retrosplenial histogram: one-sample t test in NT:

$p = 0.182$, in ASD: $p < 0.001$; two-sample t test: $p < 0.001$. NT group: $n = 26$, ASD group: $n = 25$. Bottom row: positive effect of the low-level prior mean μ_a across groups (central figure) and participants (plots) in the individual masks defined by the localizer. Histograms show the contrast estimates of μ_a (P(CW|tone)) within the individually-defined CW (red) and CCW (purple) masks. The yellow bars on the right of the graphs indicate the group means. NT histogram: one-sample t tests: $p < 0.001$ (CW mask) and $p = 0.006$ (CCW mask). ASD histogram: one-sample t tests: $p = 0.0843$ (CW mask) and $p = 0.002$ (CCW mask). NT group: $n = 25$, ASD group: $n = 25$. In the fMRI analyses, significance level was set at $p < 0.001$ at voxel level and $p < 0.05$ at cluster level. In the histograms, the significance level of the one- or two-sample t tests (two-sided) are indicated as follows: # $p = 0.08$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

exceedance probability was 1.00 in the NT group and 0.65 in the ASD group. In the ASD group, the second model that best explained the data was the model *Associative learning & Sensory memory* (AS) (protected exceedance probability of 0.35).

Following BMA, the precision of associative learning π_a was 1.37 (± 0.28) in the NT group and 1.32 (± 0.28) in the ASD group (no significant group difference: $t(50) = 0.6$, $p = 0.52$, $d = 0.18$) (Fig. 2f). The second-level learning rates ω_2 were -0.72 (± 0.86) in NT and -0.89 (± 1.28) in ASD (no significant group difference: $t(44) = 0.5$, $p = 0.59$, $d = 0.15$) (Fig. 2g). The third-level learning rates ω_3 were -6.23 (± 0.09) in NT and -6.25 (± 0.11) in ASD (no significant group difference: $t(48) = 0.6$, $p = 0.56$, $d = 0.16$) (Fig. 2h). The estimates of the other posterior parameters are given in the Supplementary Table S1.

As expected for a successful inversion of our model, the strength of associative learning π_a was positively correlated with the percentage of correct prediction responses ($r = 0.50$, $p < 0.001$) and of ambiguous trials perceived according to the current contingency ($r = 0.66$, $p < 0.001$, Fig. 2i).

We investigated whether the strength of associative learning was correlated with autistic traits (AQ), intolerance of uncertainty (IU), and atypical sensory sensitivity (GSQ). After correcting for multiple comparisons using FDR correction, none of these correlations were significant on the entire sample or within the ASD group, but two correlations remained significant in NT. Within the NT group, the strength of associative learning π_a was negatively correlated with the number of autistic traits (AQ, $r = -0.54$, $p_{\text{FDR-corr}} < 0.01$) and with the atypical sensory sensitivity (GSQ, $r = -0.54$, $p_{\text{FDR-corr}} < 0.01$).

fMRI preamble

After having identified that the associative learning model best explained the behavioral data in both groups, we aimed at characterizing the neural correlates of prior mean and precision as well as prediction errors at several levels of the hierarchy in NT and ASD participants. In all the analyses reported below, the statistical threshold was set at $p < 0.001$ at voxel level and $p < .05$ at cluster level, unless

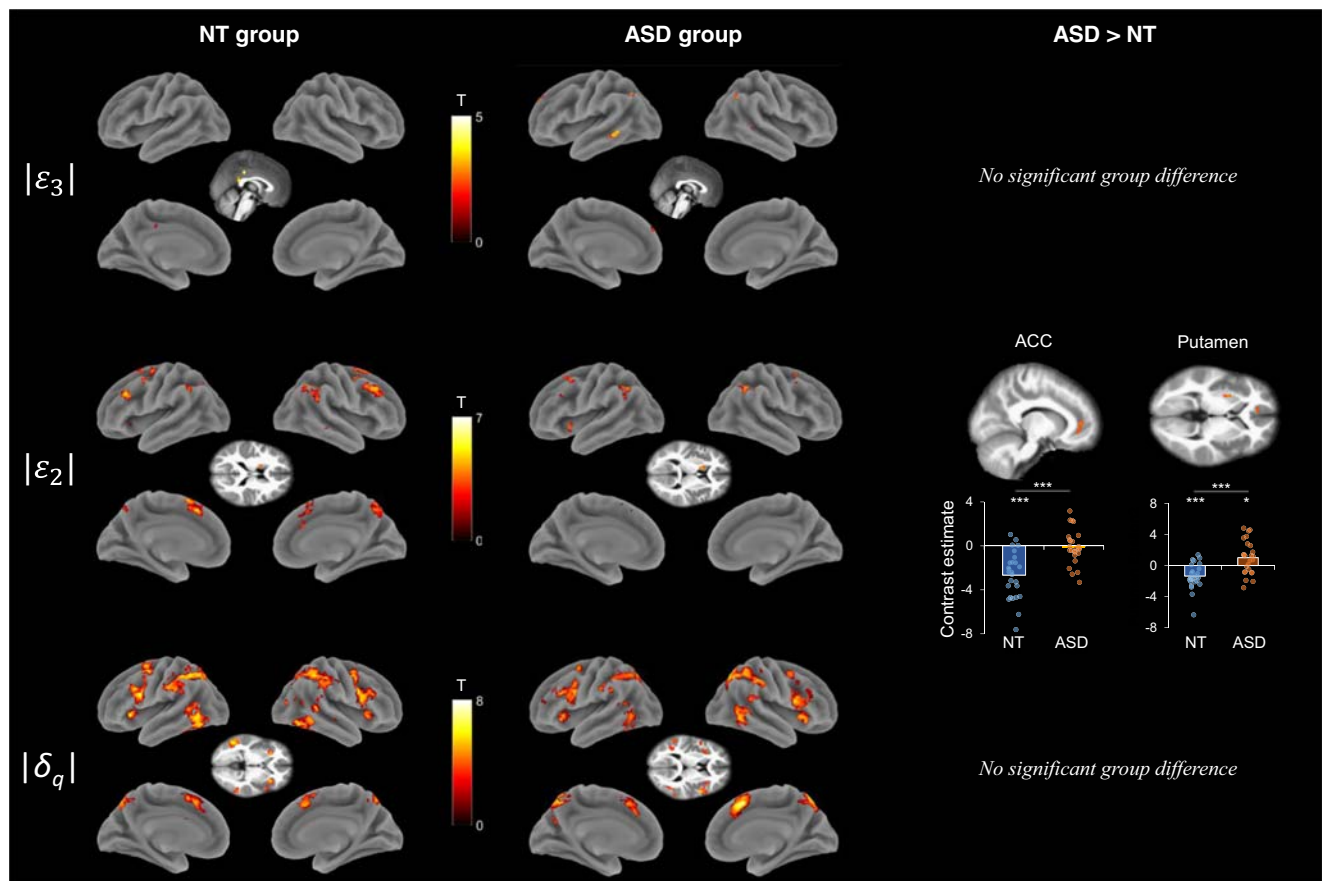


Fig. 4 | Neural correlates of the prediction errors. Top row: positive effect of the 3rd level precision-weighted prediction error $|\varepsilon_3|$ in NT and ASD. Middle row: positive effect of the 2nd level precision-weighted prediction error $|\varepsilon_2|$ in NT and ASD, and group differences in the left anterior cingulate cortex (ACC) and left putamen. ACC histogram: one-sample t test in NT: $p < 0.001$, in ASD: $p = 0.709$; two-sample t test: $p < 0.001$. Putamen histogram: one-sample t test in NT: $p < 0.001$, in

ASD: $p = 0.029$; two-sample t test: $p < 0.001$. Bottom row: Positive effect of the absolute perceptual prediction error $|\delta_q|$ in the NT and ASD groups. Significance level was set at $p < 0.001$ at voxel level and $p < 0.05$ at cluster level. In the histograms, the significance level of one- and two-sample t tests (two-sided) are indicated as follows: * $p < 0.05$, *** $p < 0.001$. NT group: $n = 26$, ASD group: $n = 25$.

it is specified “ $p_{FWE-corr} < 0.05$ ”, in which case FWE correction was applied at cluster level.

As expected, the contrast *tone vs. baseline* showed significant activation in the auditory cortex (superior temporal cortex, Heschl gyrus), precentral cortex, superior frontal, and superior parietal gyri in both groups ($p_{FWE-corr} < 0.05$). The contrast *rotation vs. baseline* involved the bilateral occipital cortex, precentral gyri, cerebellum and putamen in both groups, as well as the superior frontal gyrus in NT and superior parietal cortex in ASD ($p_{FWE-corr} < 0.05$). There were no significant group differences for these two contrasts. The main focus of the analyses was on the correlates of the estimated model parameters that were added as parametric modulators of the *tone* and *rotation* regressors, as detailed below.

Neural correlates of predictions

The correlates of predictions (high-level: $|\widehat{\mu}_3|$, mid-level: $|\widehat{\mu}_2|$ and low-level: μ_a) are illustrated in Fig. 3 and detailed in Supplementary Table S2.

In NT, high-level predictions $|\widehat{\mu}_3|$ (i.e., strength of the belief about the volatility) involved the cerebellum, supplementary motor area (SMA), precentral gyrus, middle frontal gyrus, insula, posterior superior (pSTS) and middle temporal gyri and inferior parietal lobe ($p_{FWE-corr} < 0.05$). In ASD, high-level predictions involved the SMA, pSTS, IFG, and middle cingulate cortex ($p_{FWE-corr} < 0.05$). In comparison with NT, the ASD group showed a stronger correlation between

$|\widehat{\mu}_3|$ and the activity in the left OFC, bilateral posterior cingulate cortex and right pSTS.

In both groups, mid-level predictions $|\widehat{\mu}_2|$ (i.e., strength of the belief about the probabilistic association) involved a broad cluster covering several medial and central regions (including the cingulate cortex, SMA, medial frontal, OFC, precuneus, lingual gyrus, cerebellum, striatum, hippocampus, parahippocampus), as well as the pre- and postcentral gyri, auditory cortex, pSTS and temporal pole ($p_{FWE-corr} < 0.05$). The correlation between $|\widehat{\mu}_2|$ and activity in the right retrosplenial complex was stronger in ASD than NT.

Low-level predictions μ_a were associated with activity in the occipital cortex and left postcentral gyrus (whole-brain analysis) and were further explored through a ROI-analysis. Using the individual masks mapping the CW and CCW rotations, we extracted the contrast estimate μ_a corresponding to the probability of having a CW rotation given the tone. As displayed in Fig. 3, for most of the participants, hearing a tone that was predictive of a CW or CCW rotation yielded activity in the CW or CCW mask, respectively. In the CW mask, it was the case for 84% of NT and 68% of ASD participants, and in the CCW mask, for 72% of the participants in each group (no significant proportion differences). Across NT participants, the mean activation level was significantly different from zero for both the CW ($t(24) = 4.2$, $p < 0.001$, $d = 0.84$) and CCW ($t(24) = 3.0$, $p < 0.01$, $d = 0.61$) masks. Across ASD participants, the mean activation level was significantly different from zero for the CCW mask ($t(24) = 3.4$, $p < 0.01$, $d = 0.68$) but not for the CW mask ($t(24) = 1.8$, $p = 0.08$, $d = 0.36$).

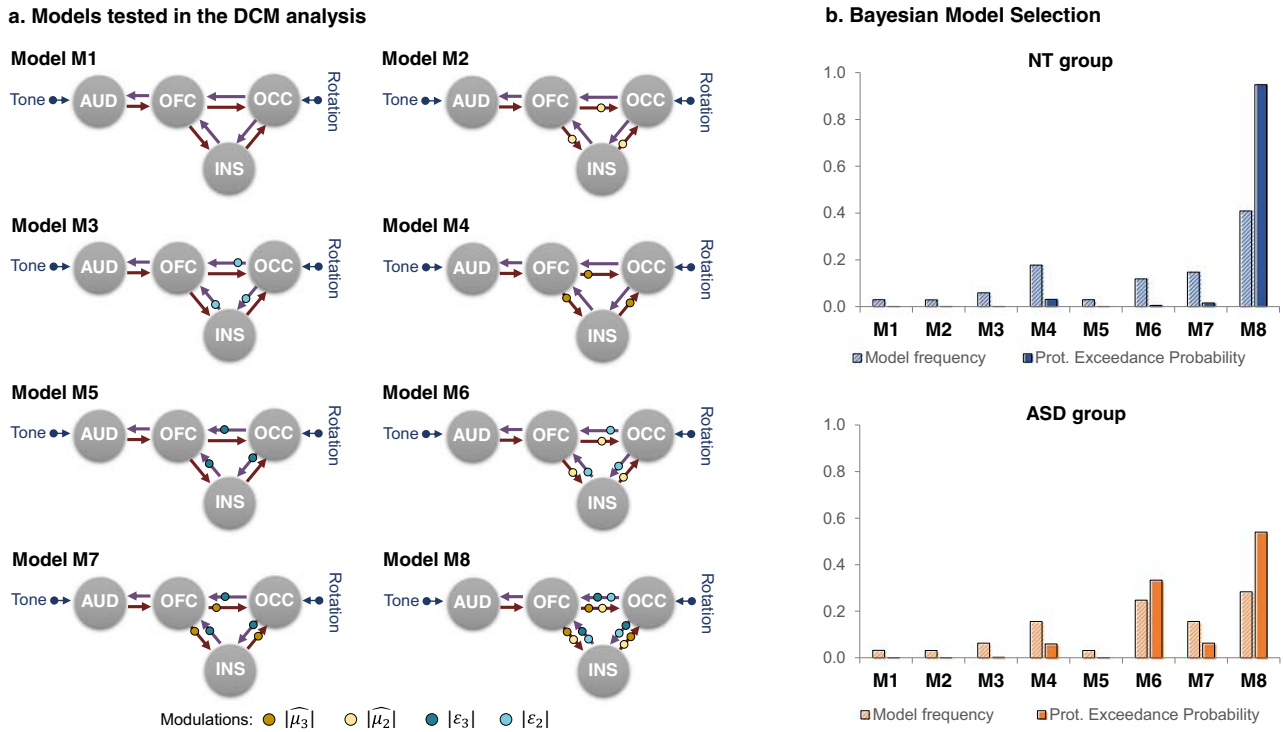


Fig. 5 | Dynamic causal modeling. **a** Description of the eight models used in the Bayesian Model Selection (BMS) assessing the effect of mid- and high-level priors and precision-weighted prediction errors on top-down and bottom-up connections. *AUD* left auditory cortex, *INS* left insula, *OFC* left orbitofrontal cortex, *OCC*

occipital cortex. To simplify the figure, the *OCC_L* and *OCC_R* were represented here as a single ROI, but the *OCC_L* and *OCC_R* were modeled separately. **b** Result of the BMS, displaying the model frequency and protected (prot.) exceedance probability.

Neural correlates of the prediction precision

The regions whose activity correlated with the precision of mid- and high-level predictions are listed in Supplementary Table S3 and shown in Supplementary Figure S1. The precision of high-level predictions $\hat{\pi}_3$ correlated with activity in the bilateral medial occipital cortex, middle and superior temporal cortex and temporo-parietal junction (TPJ) in both groups, with the bilateral middle cingulate cortex and right IFG in NT, and with the left postcentral gyrus, right central sulcus and bilateral cerebellum in ASD ($p_{FWE-corr} < 0.05$). Activity in the left superior frontal gyrus was more strongly correlated with $\hat{\pi}_3$ in NT than ASD.

The precision of mid-level predictions $\hat{\pi}_2$ did not yield any significant activity in any group. The only region that was close to significance level ($p < 0.001$ at voxel level, $p < 0.08$ at cluster level) was the right parahippocampal cortex in the ASD group. In comparison with NT, activity in the right superior parietal gyrus was more strongly correlated with $\hat{\pi}_2$ in ASD than NT.

Neural correlates of prediction errors

Finally, we were interested in characterizing the neural correlates of absolute precision-weighted prediction errors $|\varepsilon_3|$ and $|\varepsilon_2|$, and perceptual prediction errors $|\delta_j|$. The results are detailed in Supplementary Table S4 and shown in Fig. 4.

High-level precision-weighted prediction errors were associated with activity in the bilateral middle/posterior cingulate cortex in NT, and in the bilateral middle temporal gyrus, bilateral angular gyrus, and left superior frontal gyrus in ASD. There were no significant group differences.

Mid-level precision-weighted prediction errors correlated with activity in the bilateral SMA, middle and superior frontal gyrus, and inferior parietal lobe in both groups ($p_{FWE-corr} < 0.05$). In addition, they were related to the bilateral superior medial frontal gyrus, middle cingulate cortex, and precuneus in NT, and to the left insula in ASD

($p_{FWE-corr} < 0.05$). The correlation between $|\varepsilon_2|$ and the activity level in the left ACC and putamen was higher in ASD participants than NT. Note that the group difference in the ACC actually denotes an absence of correlation in the ASD group.

Perceptual prediction errors gave a similar pattern in both groups, with an involvement of the bilateral precentral gyrus, IFG, SMA, parietal lobe, insula, posterior inferior/middle temporal gyrus ($p_{FWE-corr} < 0.05$). There were no significant group differences.

Correlations between fMRI results and questionnaire scores

We assessed correlations between the contrast estimates in regions where there was a group difference and the scores of the questionnaires assessing autistic traits or symptoms. Given the presence of a group difference in these two domains, a correlation was expected, but this was performed in order to evaluate whether certain symptoms (e.g., sensory sensitivity) were especially related to one of the group differences. Medium to large positive correlations were found with most of the questionnaires, especially with the IU and AQ, and the results are presented in Supplementary Table S5.

DCM analyses

We assessed whether top-down and bottom-up connections were modulated by high-level and/or mid-level predictions and/or precision-weighted prediction errors, respectively (Fig. 5a). The BMS (Fig. 5b) revealed that M8 was the best model in the NT group (protected exceedance probability: 0.95, model frequency: 0.41). In the ASD group, the results were less clear, with M8 being the best model (protected exceedance probability: 0.54, model frequency: 0.28), followed by M6 (protected exceedance probability: 0.33, model frequency: 0.25).

A BMA on the posterior parameters showed that all the intrinsic connections were significantly different from zero (p values < 0.001 in

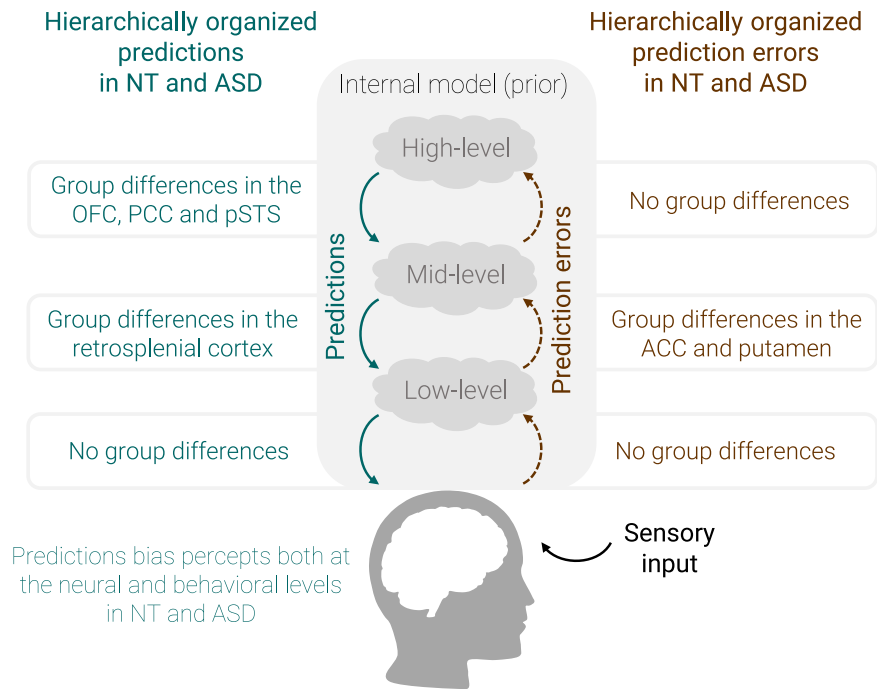


Fig. 6 | Summary figure. Using an associative learning task and model-based fMRI, we characterized the neural correlates of hierarchically structured predictions and prediction errors in neurotypical (NT) and autistic (ASD) adults at the low, mid, and high levels of the hierarchy. Both groups managed to make accurate predictions and were biased by their expectations both at the behavioral and neural levels. The neural networks identified in the two groups were globally similar, but group differences were found in a set of brain regions. High-level predictions were following

activity levels in the left orbitofrontal cortex (OFC), bilateral posterior cingulate cortex (PCC), and right posterior superior temporal sulcus (pSTS) more closely in ASD than NT. The right retrosplenial cortex was more strongly correlated with mid-level predictions in ASD than NT. Mid-level prediction errors were correlated with activity in the anterior cingulate cortex (ACC) in NT but not in ASD, and were more strongly correlated with activity in the left putamen in ASD.

both groups), except for the connection from the INS to the OFC (NT: $p = 0.10$, ASD: $p = 0.52$). In NT, there were modulatory effects by $|\varepsilon_3|$ on the connections from the OCC_L and OCC_R to the INS (p values < 0.005), and by $|\varepsilon_2|$ on the connections from the OCC_R to the INS (p values < 0.05). In NT, there was a modulatory effect by $|\widehat{\mu}_2|$ on connections from the OFC to the OCC_L ($p < 0.01$), and non-significant trends for modulations by $|\widehat{\mu}_3|$ on connections from the OFC to the OCC_L ($p = 0.067$) and by $|\widehat{\mu}_2|$ on connections from the INS to the OCC_L ($p = 0.099$). In ASD, there were modulatory effects by $|\varepsilon_3|$ on the connections from the OCC_L to the INS ($p < 0.05$) and to the OFC ($p < 0.05$). In ASD, there was a modulatory effect by $|\widehat{\mu}_3|$ on connections from the OFC to the INS ($p < 0.005$), and a non-significant trend toward a modulation by $|\widehat{\mu}_2|$ on this connection ($p = 0.062$). Between-group comparisons only showed a non-significant trend toward a group difference on the modulation by $|\varepsilon_2|$ on the connection from the OCC_L to the INS ($t(48) = 1.7$, $p = 0.89$, NT: $0.02 \text{ Hz} \pm 0.18$, ASD: $-0.16 \text{ Hz} \pm 0.48$).

Discussion

In this study, we used model-based fMRI to investigate the neural correlates of predictive mechanisms in ASD. We were particularly interested in characterizing the set of brain regions involved in signaling predictions and prediction errors in ASD, and in potentially revealing group differences to shed light on Bayesian theories of ASD. At the behavioral level, the two groups managed to make accurate predictions and were biased by their expectations. At the neural level, we identified the correlates of hierarchical predictive processes within group and found a rather large overlap of brain regions between groups, but also some group differences at the higher levels of the hierarchy. The main findings of this study are summarized in Fig. 6.

The predictive skills of ASD participants appeared to be typical in this associative learning task. Indeed, both groups managed to learn predictions and were biased by their expectations, which indicates that

autistic individuals do not have uniformly weak priors. In this volatile environment, autistic adults managed to flexibly use prior knowledge to make perceptual inference. In the same task with a more uncertain context (i.e., lower probabilistic association)²⁴, autistic adults were also able to learn predictions, but failed to update them after a change in contingency. As compared to NT, autistic adults had a decreased precision of associative learning (π_a) and were less biased by priors²⁴. This is in contrast with the current study where no such group differences were found, even though more autistic traits in NT were also associated with lower π_a . It suggests that autistic adults are able to handle some degrees of expected and unexpected uncertainty, but might be impaired above a certain level of uncertainty. This is in line with a recent review²⁰ concluding that associative learning may be particularly impaired when predictive features have low consistency. Similarly, a probabilistic reinforcement learning task⁴⁷ showed that autistic adults managed to learn stimulus pairs that were frequently reinforced (80%), but were impaired when the reinforcement was less frequent (70%). In addition, we found no group differences in learning rates, consistently with the behavioral experiment cited above²⁴. In some other associative learning tasks, the learning rates did not differ either in autistic vs. typically developing children²⁵, or in adults with more or less autistic traits⁴⁸.

The neural correlates of predictions were identified in each group, and to the best of our knowledge, these regions had not been characterized in ASD before. First, we observed a hierarchical organization of regions encoding predictions in each group. While low-level predictions only involved the occipital cortex, higher-level predictions elicited activity in a broader set of supramodal regions. The regions identified along this hierarchy resemble the principal gradient of macroscale organization in the human connectome⁴⁹, as higher-level predictions involve regions belonging to the principal gradient, i.e., transmodal regions including the default mode network and

frontoparietal regions, while low-level predictions resemble the map of the second functional gradient. Regions of the default mode network are at the upper end of the topographical hierarchy and integrate information from several sensory modalities to get abstract representations⁴⁹. This functional gradient from primary sensory regions to transmodal regions is therefore in line with the hierarchical neural correlates of predictions that we identified. Interestingly, a recent study investigated the microstructural profiles of these regions⁵⁰. They identified gradual cytoarchitectural variations that globally followed the functional gradient, but the transmodal cortices had a less hierarchical organization⁵⁰. The rich and less-constrained interconnectivity of these transmodal regions would be central to underlie flexible cognitive functions⁵⁰, which may allow representing higher-level abstract representations, for instance about the stability of predictions over time. Interestingly, at the lower-level of the hierarchy, we found that hearing a tone elicited activity in the retinotopic areas representing the CW or CCW rotations that was modulated specifically by the rotation direction predicted by the tone. This effect is consistent with findings in NT³³, and reveals proactive mechanisms in both NT and autistic individuals. Together with the prior bias observed at the behavioral level, this fMRI result indicates a similar influence of priors on sensory processing at low hierarchical levels in autistic individuals as in NT. In particular, using fMRI allowed to point out that sensory processing itself is biased by expectations in ASD, resolving an ambiguity, which often plagues behavioral studies, on whether priors bias perceptual or decisional processes. Hence, in the context of this experiment, the findings do not support the weak prior hypothesis of ASD¹¹. This result for low-level predictions is in line with a recent study showing that the illusory triangle induced in the primary visual cortex by the Kanizsa figure did not differ between NT and ASD⁴⁶, suggesting an intact perceptual bias in ASD. Future studies including more ambiguous trials could perform decoding analyses (e.g., such as in ref. 51) to assess if the perceptual bias induced in ambiguous trials generates activity in motion area MT.

The pattern of brain regions associated with mid-level predictions is highly consistent with the literature^{34,36,52–54}, in particular with results in NT using the same paradigm³³ and showing an involvement of the OFC, hippocampus, insula, precuneus, and medial frontal gyrus. This pattern is also consistent with regions that are processing contextual associations, such as the retrosplenial complex and parahippocampal cortex, and that contribute to eliciting prediction-related representations in the OFC¹. In particular, the activity level in a cluster of the right retrosplenial complex followed the strength of mid-level predictions more closely in ASD than NT. This region plays a fundamental role in learning contextual associations between sensory stimuli^{1,55–57}. One potential explanation for this group difference could be that the tone-rotation association was learned more strongly or more rigidly, in ASD. This stronger correlation might also be interpreted as ASD participants needing a greater involvement of this region to reach a similar performance. Future studies could use a causal approach to further corroborate these interpretations.

High-level predictions about the stability of the association over time evoked activity in regions such as the middle/posterior cingulate gyrus, IFG, pSTS or SMA in both groups. Yet, the pattern of regions seemed a bit more distinct between groups than for mid-level predictions. In particular, there were stronger correlations between high-level predictions and activity in the left OFC, bilateral posterior cingulate cortex and right pSTS in ASD, relative to NT. A plausible hypothesis could be that these regions might encode the volatility more precisely in ASD, which may be related to the idea of an over-estimation of the environmental volatility in ASD²⁶. It may also contribute to a higher intolerance of uncertainty and more autistic traits, as suggested by the correlations with the questionnaires. The stronger correlations with high-level predictions in these regions might entail different predictive processes in ASD. For instance, the posterior

cingulate cortex signals the expected environmental information and contributes to increase routines⁵⁸. Particularly, the OFC plays a key role in encoding prior uncertainty, maintaining contextual priors and updating perceptual beliefs^{1,34,59,60}. In ASD and NT, the model best explaining the data was the one where both mid- and high-level predictions modulating top-down connections from the OFC to the visual cortex and insula. We can hypothesize that the stronger activation of the OFC in ASD may be due to more top-down prior knowledge sent from this region in ASD, which might suggest a difference in the specialization of the OFC to encode priors. Nevertheless, note that the extent to which predictions modulated top-down connections from the OFC did not differ significantly between groups.

Regarding prediction errors, the neural correlates of low-level perceptual prediction errors encompassed regions such as the IFG, insula, precentral gyrus, inferior parietal and the middle temporal gyrus in both groups, consistently with findings in NT³³. As reported in a recent meta-analytic study³⁶, the IFG and insula were found to be involved in signaling both predictions and prediction errors. The absence of group differences suggests that low-level perceptual prediction errors may be processed similarly in NT and ASD. Mid-level precision-weighted prediction errors involved, among others, middle and superior frontal regions, the caudate nucleus, the insula and inferior parietal regions in both groups, which is again in line with results in NT^{31,33}. In ASD compared to NT, there was a stronger correlation between mid-level prediction errors and activity levels in the putamen, a region encoding prediction errors^{41,61}. A plausible interpretation could be a higher weight of mid-level prediction errors in ASD, in line with the HIPPEA theory¹³ suggesting that highly precise prediction errors would cause overwhelming sensations of surprise, leading them to avoid unpredictable environments. Furthermore, we observed a negative correlation between neural activity in the ACC and mid-level precision-weighted prediction errors in NT, but not in ASD. Interestingly, in a social context, the gyral surface of the ACC showed a deactivation in response to social prediction errors in NT, but not in ASD²⁸. Moreover, they found that the activation level of the ACC was positively correlated with the social symptoms of ASD assessed with a sub-scale of the ADOS²⁸, while we also found positive correlations between the ACC activity level and autistic traits or social difficulties across groups. The encoding of prediction errors in the ACC might therefore be altered across multiple domains in ASD, and could play a key role in the atypical predictive learning mechanisms observed in ASD. Finally, even though no significant group differences were found, high-level precision-weighted prediction errors were associated with distinct sets of regions in each group. The DCM analysis showed that both high- and mid-level precision-weighted prediction errors modulated bottom-up connections similarly in each group.

The current study has several limitations. First, the DCM analysis only focused on a reduced set of regions and might have shown a different pattern of results if we would have selected other candidate regions such as the IFG, ACC, caudate nucleus, or putamen. Furthermore, in the fMRI analyses, we decided to focus on positive effects of the model quantities on brain activity, which seems easier to interpret, but negative effects may also provide additional information. Using a design with shorter TRs could have revealed variability across different regions of the cortex in the temporal shape of the HRF⁶², with faster responses in sensory regions than in associative areas⁶³. We chose to rely on a dual-report paradigm with both prediction and perception responses to get an explicit measure of prediction learning (where tone could not be simply ignored) and to get an implicit measure of perceptual bias, but using such paradigm might bias responses as participants often tend to ensure self-consistency⁶⁴. Finally, some of our autistic participants had comorbidities or were taking

Table 1 | Demographic characteristics and questionnaire scores of the participants

	NT group	ASD group	p
Number of participants	26	26	-
Male/female number	13/13	13/13	ns
Age (years)	30.9 (±8.3)	32.2 (±9.5)	ns
Left/right-handed	2/24	5/21	ns
Total IQ score	113.9 (±12.3)	112.1 (±16.5)	ns
AQ score	13.0 (±6.2)	32.1 (±8.7)	***
IU score	27.9 (±8.4)	42.7 (±6.6)	***
GSQ score	34.5 (±16.7)	53.3 (±21.2)	***

The table presents the group means (±standard deviations). *IQ* Intelligence Quotient (Wechsler Adult Intelligence Scale IV), *AQ* Autism spectrum Quotient, *IU* Intolerance of Uncertainty, *GSQ* Glasgow Sensory Questionnaire. Note that IQ data from four NT participants are missing as they were not native Dutch speakers. One of the ASD participants was included in the behavioral analyses but not in the fMRI analyses. The two groups were compared using two-sided Student *t* tests (i.e., age, total IQ, AQ, IU, and GSQ scores) and chi-square tests (i.e., proportions of male/female, left/right-handed). ns non-significant, ****p* < 0.001.

medications, which may have influenced their behavior and neural responses. Yet, we believe that including such participants is more representative of the ASD population, as having comorbidities or taking some medication is highly frequent in ASD.

In conclusion, we found that autistic adults can learn priors and tend to encode priors hierarchically, like NT. Yet, they showed alterations in the neural processing of mid- and high-level predictions that may be interpreted as a stronger encoding of these predictions. In addition, as autistic individuals had their percepts biased by their expectations both at the behavioral and neural levels, it does not support the weak prior hypothesis of ASD¹¹. Correlates of mid-level precision-weighted prediction errors may be in line with stronger prediction errors in ASD, consistent with the HIPPEA hypothesis¹³. Overall, the predictive mechanisms appear to be relatively intact in ASD, but subtle differences at higher levels of the hierarchy might contribute to the difficulties in predictive learning observed in certain contexts. These results suggest that the Bayesian account of ASD should be refined to better describe predictive mechanisms in ASD. Finally, planning comparative neuroimaging studies with other clinical populations hypothesized to have atypical perceptual inference processes (e.g., schizophrenia or dyslexia) would help to better identify the specificities of each condition regarding predictive skills.

Methods

Participants

Twenty-six neurotypical (NT) adults and 26 autistic adults participated in the MRI experiment. Their demographic characteristics are shown in Table 1. The two groups were matched for age, sex ratio, handedness ratio, and total intelligence quotient. One autistic participant completed the behavioral experiment but fMRI scans were not acquired as she was too sensitive to the noise of the scanner when we started the fMRI acquisition. Hence, the sample size is 26 in each group for the behavioral/model analyses, but 25 in the ASD group, and 26 in the NT group for the fMRI analyses. Note that for the localizer analyses, the sample size is 25 ASD and 25 NT participants, as there was a technical failure during the acquisition of the localizer of one NT participant.

Inclusion criteria were being between 18 and 50 years old, reporting normal or corrected-to-normal hearing and vision. Exclusion criteria were having contra-indication for MRI, having a total intelligence quotient below 70 at the Wechsler Adult Intelligence Scale IV⁶⁵, or scoring above 32 at the Autism Spectrum Quotient (AQ)⁶⁶ for NT participants. Autistic participants received their diagnoses from a multidisciplinary Expertize Center for Autism (University Hospitals of KU Leuven) in a standardized way according to

the criteria of the DSM-5¹⁵, and had idiopathic ASD. Autistic adults were recruited via this expertize center and via the LAuRes (Leuven Autism Research) consortium website. NT participants were recruited via the University of Leuven or acquaintances. None of the NT participants reported having a comorbidity or being under medication. Five ASD participants reported having one or several comorbidities (ADHD (4), dyslexia (2), Gilles de la Tourette (1)). These participants who had comorbidities are included in the analyses, but note that the results of the behavioral and fMRI analyses did not change after removing these five participants. Eleven ASD participants reported taking one or several medications (Abilify (2), Asaflow (1), Bufonix (1), Celecoxib (1), Deanxit (1), Depakine (1), Escitalopram (1), Fluoxetine (1), Fluoxone (1), Hydrea (1), L-Thyroxine (3), Medikinet (1), Melatonin (1), Montelukast (1), Notrilen (1), Redomex (1), Ritalin (1), Trazadone (1), Venlafaxine (1), Welbutrin (1)).

The study was approved by the medical Research Ethical Committee UZ / KU Leuven. All participants provided written informed consent, according to the Declaration of Helsinki.

Overall procedure

Prior to the experiment, participants filled out online questionnaires (in Dutch or English, depending on their native language): AQ^{66,67}, short version of the Intolerance of Uncertainty scale^{68,69} and Glasgow Sensory Questionnaire^{70,71}. Participants started with a short training (7 trials), before being installed in the MR scanner. After a T1-weighted anatomical scan, participants performed another short training and five fMRI runs, followed by a localizer run. Magnetic resonance spectroscopy data were also acquired during this session, as part of another study⁷².

Data were acquired on a 32 head coil 3 T Philips Achieva system at the University Hospital of Leuven. During the fMRI acquisition, stimuli were projected on a screen behind the scanner, and reflected through a mirror mounted on the head coil. Stimuli were presented using the Psychtoolbox (version 3) in Matlab (version 2019a), and auditory stimuli were presented binaurally via headphones.

Experimental paradigm

Main task

Trial structure. Participants performed an associative learning task (Fig. 1, based on³³) where tones were probabilistically associated with the rotation direction of two dots. Participants were instructed that there was an underlying association between the tone and the rotation direction, and that this association could change. A high (576 Hz) or low (352 Hz) tone was presented for 500 ms and followed by a *prediction* response screen for 1000 ms (jitter of 100 to 300 ms between the tone and the prediction screen). The prediction screen displayed a right and a left arrow, and participants had to click on the right or left button of the MRI response box if they thought that the tone was predictive of a clockwise (CW - right) or counterclockwise (CCW - left) rotation, respectively. The arrow selected by the participant turned red. Then, two dots appeared at their vertical position for 600 ms, made a CW or CCW rotation within 33 ms and remained at their horizontal position for 600 ms. This was followed by a *perception* response screen showing a right and a left double arrow, displayed for 1000 ms. Participants had to report whether they perceived a CW or CCW rotation using the right or left button, respectively (the selected double-arrow turned red). The inter-trial interval lasted for 1000 to 2000 ms (uniform distribution).

There were two types of trials: *unambiguous trials* where the two dots rotated, and *ambiguous trials* where the pair of dots simply appeared in their vertical and then horizontal positions (no rotation). While unambiguous trials allowed learning the contingency, ambiguous trials allowed assessing whether participants were biased by their expectations (i.e., if they would report a rotation consistent with their expectations, despite the absence of rotation).

Run structure. Participants completed five runs of 72 trials (total of 360 trials). In each run, there were 12.5% of ambiguous trials (9 trials) and 87.5% of unambiguous trials (63 trials). Among the unambiguous trials, 75% were expected (i.e., main tone-rotation association) and 12.5% were unexpected (i.e., least frequent association). The trial order was pseudo-randomized so that these percentages (75% unambiguous expected, 12.5% unambiguous unexpected, 12.5% ambiguous) remained the same across eight successive trials. In each run, the contingency reversed after 16, 24, or 32 successive trials.

Localizer task. The localizer task was designed for retinotopic mapping of the CW and CCW dot trajectories. Checkerboards covering the dot CW or CCW trajectory were displayed for 15 s, and flickered at a frequency of 8 Hz. The checkerboards did not overlap the initial vertical or final horizontal dot position but covered the upper-right and lower-left quadrants in CW trials, and the upper-left and lower-right quadrants in CCW trials. There were 4 CW trials and 4 CCW trials interleaved, separated by 5 s of fixation. To maintain attention, participants had to fixate on a central cross and click on the left button when the cross changed of color (white or red).

Post-fMRI confidence rating task. After finishing the fMRI experiment, participants completed a short computer task as in³³ to assess the perceptual quality of the ambiguous trials. The structure of this task was the same as the main task, but trials included a third response screen showing the options “1. *Very sure*”, “2. *Quite sure*”, “3. *Quite unsure*”, “4. *Very unsure*” (displayed for 2600 ms). Participants used the numbers 1 to 4 to indicate how confident they were about their perception response. There were 48 trials, including 50% of ambiguous trials and 50% of unambiguous trials. To calculate a mean confidence rating, the 1 (*Very sure*) to 4 (*Very unsure*) scale was transformed into a 100% to 0% certainty scale.

MRI acquisition

Anatomical scan. A high-resolution T1-weighted anatomical scan was acquired with a MPRAGE sequence (200 contiguous coronal slices, voxel size = $1 \times 1 \times 1$ mm³, TR = 9.7 ms, TE = 4.6 ms, field of view = $256 \times 240 \times 200$ mm³, acquisition matrix = 256×238 , acquisition time = 4 min 35 s).

Functional MRI. Whole-brain T2*-weighted echo-planar imaging sequences were collected (voxel size = $2 \times 2 \times 2$ mm³, TR = 2 s, TE = 30 ms, flip angle = 90°, FOV = $224 \times 224 \times 132$ mm³, 60 ascending transverse slices, multi-band factor: 2). In each run of the main task, 198 volumes were acquired (6 min 44 s per run, total of 990 volumes). In the localizer run, 160 volumes were collected (4 min 52 s).

Behavioral analyses

Statistical analyses. All the results are presented as mean (\pm standard deviation). Demographic data (Table 1) of the NT and ASD groups were compared using Student's *t* tests and chi-square tests. The percentage of correct predictions was compared to chance level using one-sample *t* tests with $\mu = 0.50$. Between-group comparisons on accuracy levels were performed using two-sample *t* tests. Effect sizes are reported as Cohen's *d*: very small ($d = 0.01$), small ($d = 0.20$), medium ($d = 0.50$), large ($d = 0.80$), or very large ($d > 1.20$) effect sizes^{73,74}. Correlations were assessed using Pearson's correlation test, and were corrected for multiple comparisons using False Discovery Rate (FDR) correction. Statistical analyses were performed using R (version 4.0.3, <http://www.r-project.org/>). All Student's *t* tests were two-tailed. The threshold for statistical significance was set at $p < 0.05$.

Computational modeling. The behavioral data were modeled using the Bayesian modeling approach of Weilhhammer and colleagues³³ (see the Supplementary Note S1 for the full mathematical model

description). The model consists of a *contingency model* and a *perceptual model* (Fig. 2a–b). The mean and variance of the priors used in this model are shown as Supplementary Table S1 and are the same as in ref. 33.

The *contingency model* infers the associations between the tone and the rotation direction to determine the prediction response. It relies on a three-level Hierarchical Gaussian Filter (HGF)^{40,75} with low-, mid-, and high-level priors. Low-level priors correspond to the subjectively estimated chance of having a CW or CCW rotation (binomial), and to the conditional probability of a certain rotation direction given the tone. Mid-level priors capture the belief about the probabilistic association between the tone and the rotation direction. High-level priors model the stability of this association over the time course of the experiment (i.e., this association might remain stable or reverse).

The *contingency model* is coupled with a *perceptual model* which assesses whether the perceptual responses are influenced by associative learning (i.e., influence of the current hidden contingency), priming (i.e., influence of the preceding trial) and/or sensory memory (i.e., influence of the preceding ambiguous trial on the next ambiguous trial).

The behavioral data (prediction and perception responses) were fitted by eight models that included associative learning (A), priming (P) and/or sensory memory (S) in the perceptual model (i.e., none: 0, A, P, S, AP, AS, PS, APS). Model inversions were performed separately for each run. We used the HGF for binary inputs with the quasi-Newton Broyden-Fletcher-Goldfarb-Shanno minimization as optimization algorithm, implemented in the HGF 4.0 toolbox (TAPAS toolbox – Translational Algorithm for Psychiatry-Advancing Science, translationalneuromodeling.org/tapas/) in Matlab (R2020b version). To identify which of the eight models best fitted the data in each group, we used random-effect Bayesian Model Selection (BMS)⁷⁶ in SPM12 (<http://www.fil.ion.ucl.ac.uk/spm/>). This model comparison relies on the log-model evidence, which is calculated as the negative variational free energy. The model parameters were estimated using Bayesian Model Averaging (BMA) and compared between groups using Student's *t* tests. As an additional way to check the model fit, behavioral data (percentages of correct predictions) were correlated with the estimated model parameters using Pearson correlation tests. Model and parameter recoveries are presented in Supplementary Note S2.

fMRI analyses

The fMRI data were preprocessed and analyzed using Matlab 2020b and Statistical Parametric Mapping (SPM), version 12.

fMRI preprocessing. Prior to the preprocessing the anatomical image was manually reoriented to be centered on the middle anterior commissure. The preprocessing consisted of realigning the functional volumes of the five runs and of the localizer, slice timing correction (on timings, with the middle slice as reference), coregistration of the anatomical scan, segmentation, normalization (based on the standard template of the Montreal Neurological Institute, MNI) and smoothing with a Gaussian kernel of 6 mm (full width at half maximum). The Artifact detection toolbox (ART: art-2015-10 release, https://www.nitrc.org/projects/artifact_detect/) with a motion threshold of 2 mm was used to detect the outliers that were added as unique regressors to discard the identified outlier scans. On average, 0.5% and 1.5% of scans were excluded in the NT and ASD groups, respectively. Runs with more than 15% of outliers were excluded, resulting on the exclusion of the fourth and fifth runs of one ASD participant (i.e., all participants had five runs included, except for one ASD participant who had three runs included).

Finally, we conducted a denoising step using the CONN functional connectivity toolbox v17 (<http://www.nitrc.org/projects/conn>). Using

aCompCor⁷⁷, a principal component analysis on the cerebrospinal fluid and white matter masks was performed to remove physiological confounds from the signal.

Whole-brain analysis. In order to identify the neural correlates of predictions and prediction errors in individuals with and without ASD, we conducted a model-based fMRI analysis. We defined a General Linear Model (GLM) using two regressors (tone and rotation) and nine parametric modulators that were model quantities from the inverted behavioral model. These parametric modulators mostly corresponded to the subject-specific trajectories of prediction errors, prior mean, and precision.

The regressor *tone* modeled an event starting at the onset of the tone and was parametrically modulated by five model parameters: 3rd level prior precision $\hat{\pi}_3$ and mean $|\hat{\mu}_3|$, 2nd level prior precision $\hat{\pi}_2$ and mean $|\hat{\mu}_2|$, and low-level predictions μ_a (i.e., inferred conditional probability of a CW rotation). The 3rd level represents the belief about the volatility of the environment, while the 2nd level corresponds to the belief about the main contingency. The absolute prediction mean (e.g., $|\hat{\mu}_2|$) models the strength of the belief (e.g., an individual with a high or low estimated $|\hat{\mu}_2|$ believes that the tone is highly or not predictive of the rotation direction, respectively). The prediction precision denotes how quickly these estimates change across time. The regressor *rotation* modeled an event starting 600 ms after the presentation of the two vertical dots, i.e., at the onset of the rotation in unambiguous trials, and at the onset of the presentation of the two horizontal dots in ambiguous trials. The regressor *rotation* was parametrically modulated by four model parameters: 3rd level precision-weighted prediction errors $|\varepsilon_3|$, 2nd level precision-weighted prediction errors $|\varepsilon_2|$, perceptual prediction error $|\delta_q|$ (i.e., lower-level prediction error defined as $\delta_q = P(\Theta_j) \cdot y_{\text{perception}}$) and the posterior probability of perceiving a CW rotation $P(\Theta_j)$. All regressors were convolved with a canonical hemodynamic response function. Note that the definition of the GLM was based on the study on NT³³, but third-level model quantities were added to the GLM.

In addition, the potential confounds were modeled as separate regressors of the GLM matrix and consisted of the six motion parameters, the ART-based outliers (if any) and the 10 first principal components identified with aCompCor in the denoising step.

At the first level, contrast images were computed for each regressor or parametric modulator at the individual level using t-statistics. At the second level, the mean of the contrasts across participants was compared to zero using a one-sample Student's *t* test. Activation patterns were compared between the two groups using independent two-sample *t* tests. When a significant group difference was found, we used Marsbar (release 0.44)⁷⁸ to extract the mean contrast estimate in the cluster ($p < 0.001$ at the voxel level, $p < 0.05$ at cluster level) and plot the distribution of contrast estimates across participants.

Voxel-based level thresholding was set at $p < 0.001$, and cluster-level extent thresholding was set at $p < 0.05$ (cluster-extent thresholds are estimated by Gaussian Random Field method implemented in SPM12). In the Results section, we report whether these results remain significant after Family Wise Error (FWE) correction at cluster level. The SPM toolbox bspmview (version 20161108)⁷⁹ was used to illustrate the brain activation patterns in Figs. 3, 4 and S1.

To assess if the perceptual bias induced in ambiguous trials generated activity in V5/MT, we conducted additional fMRI analyses, presented in Supplementary Note S3 and Supplementary Figure S2.

Localizer-based analysis. In addition to the whole-brain analysis, we conducted a region of interest (ROI)-based analysis, after having defined the ROIs using the localizer and the main task. The aim was to investigate whether the low-level prediction (μ_a or $1 - \mu_a$) about the rotation direction of the two dots (CW or CCW) would trigger

activation in the retinotopic representation of the CW or CCW trajectories at the onset of the tone (i.e., whether the brain was proactive). The three-step procedure for the selection of the ROIs was the same as in the study by Weilhammer and colleagues³³.

First, we specified a GLM with two box-car regressors in the localizer run: a *CW regressor* corresponding to the presentation of checkerboards over the upper-right and lower-left trajectories, and a *CCW regressor* corresponding to the checkerboard presentation over the lower-right and upper-right quadrants. Each of these regressors started at the onset of appearance of the checkerboard and lasted for 15 s. At the subject-level, we used the contrasts CW > CCW and CCW > CW, thresholded at $p < 0.05$ to identify the clusters responding to CW and CCW trajectories and used Marsbar to extract these clusters.

Second, in order to only select the voxels that were highly specific for the dot CW and CCW rotations in the main task, we constructed a second single-subject GLM. This GLM defined in the main task contained three regressors, all modeled as events (0 s duration): *CW regressor* at the onset of the CW rotation, *CCW regressor* at the onset of the CCW rotation and *Ambiguous regressor* at the onset of appearance of the dots in their horizontal position in ambiguous trials. At the subject-level, we used the contrasts CW > CCW and CCW > CW, thresholded at $p < 0.05$ (uncorrected) to identify the clusters responding to the CW and CCW rotations and used Marsbar to extract these clusters.

Third, we defined the CW and CCW ROIs by selecting the intercept of the masks defined in the two previous steps. At a single-subject level, these ROIs were used to mask the contrast μ_a (parametric modulator of the tone) in the GLM described in the "Whole-brain analysis". The individual contrast estimates were again extracted using Marsbar.

DCM analysis. We conducted a DCM⁸⁰ analysis to assess how mid-level and high-level priors and precision-weighted prediction errors influenced top-down and bottom-up connections, respectively, in each group. Using DCM has the particularity to make inferences about the causal relationships of activity patterns.

Definition of the GLM. The GLM used in the DCM analysis was the same as in the main analysis, but only included the regressors *Tone*, *Rotation*, $|\hat{\mu}_3|$, $|\hat{\mu}_2|$, $|\varepsilon_3|$ and $|\varepsilon_2|$.

Definition of the regions of interest and extraction of the time series. The DCM analysis involved the bilateral occipital cortex, left auditory cortex, the left OFC, and the left insula. The OFC has been identified as a key region encoding contextual priors^{33,34,39,81,82}. We decided to include the insula among two other main candidates (i.e., the ACC and the caudate nucleus) as it was reported in articles investigating prediction errors across multiple levels of the hierarchy^{31–33,41,83} and because a recent meta-analytical study³⁶ highlighted the role of the left insula in encoding prediction errors.

The regions of interest (ROIs) were identified using a two-step procedure: we first identified the peak of maximum activity across groups, and then at the individual level (same methods as in^{84,85}). First, we defined the peak of activation across groups with $p < 0.001$ at voxel level and $p_{\text{FWE-corr}} < 0.05$ at cluster level. The contrast *Tone* vs. *baseline* was used to identify the peak of the auditory ROI (AUD: $x = -58, y = -14, z = 4$). Note that the maximum activity was in the left hemisphere and we decided to restrict to one region as the same auditory input was sent in each ear. We used the contrast *Rotation* vs. *baseline* masked with the CW or CCW mask from the localizer to get the peaks of the left occipital (OCC_L: $x = -32, y = -88, z = 4$) and right occipital (OCC_R: $x = 34, y = -84, z = 4$) ROIs. We used the contrast $|\hat{\mu}_2|$ vs. *baseline* to select the peak coordinates of the left orbitofrontal cortex (OFC: $x = -26, y = 34, z = -10$). Using the contrast $|\varepsilon_2|$ vs. *baseline*, we extracted the peak activity in the left anterior insula (INS: $x = -36, y = 14,$

$z = -2$). As participants included in the DCM analysis needed to have similar numbers of runs, the ASD participant who had only 3 runs instead of 5 was excluded from this analysis (i.e., $n = 26$ NT, and $n = 24$ ASD in this analysis).

Then, we created a 12.5 mm-radius spheres around the maxima described above using Marsbar. Within this sphere, we extracted the coordinates of the peak activity of each participant using the same contrasts, with $p < 0.01$ for the OCC ROIs, and $p < 0.05$ for the other ROIs. For each contrast of interest, we then extracted the first eigenvector across all voxels that were above the threshold within a 6 mm-radius sphere centered on the individual peak coordinates.

DCM specification and estimation, and model comparison. The inputs were the sound (regressor *Tone*) entering through the AUD, and the visual input (regressor *Rotation*) entering through the OCC_L and OCC_R.

We first specified a DCM with no modulatory influence to identify the intrinsic connections. All the models included bidirectional connections between the AUD and OFC. In addition, in M1_{intrinsic}, the OFC was connected to the OCC and INS, in M2_{intrinsic}, the OCC was connected to the INS and OFC, in M3_{intrinsic}, the INS was connected to the OCC and OFC, and in M4_{intrinsic}, these regions were all connected (i.e., OFC with the OCC and INS, and OCC connected with the INS). BMS showed that M4_{intrinsic} was the best model in each group (protected exceedance probability in both NT and ASD: 1.00). So, we selected this model and added modulatory influences, as described below.

We considered modulatory influences of top-down connections from the OFC to the OCC and INS, and from the INS to the OCC by the prior mean at a high- and/or mid-levels, and modulatory influences of bottom-up connections by precision-weighted prediction errors at high- or mid-levels from the OCC to the OFC and INS, and from the INS to the OFC. We specified and estimated eight models to assess if the connections were modulated by priors and/or prediction errors at the mid and/or high levels of the hierarchy. The eight models were specified as follow: M1 without any modulation, M2 with a modulation by $|\widehat{\mu}_2|$, M3 with a modulation by $|\varepsilon_2|$, M4 with a modulation by $|\widehat{\mu}_3|$, M5 with a modulation by $|\varepsilon_3|$, M6 with modulations by $|\widehat{\mu}_2|$ and $|\varepsilon_2|$, M7 with modulations by $|\widehat{\mu}_3|$ and $|\varepsilon_3|$, and M8 with modulations by $|\widehat{\mu}_2|$, $|\varepsilon_2|$, $|\widehat{\mu}_3|$ and $|\varepsilon_3|$ (Fig. 5).

Finally, we used BMS to select the model that best fitted the fMRI data at the group level. We report both the protected exceedance probability and the model frequency for each model. We then used BMA to estimate the values of the posterior parameters of the DCM analyses (i.e., connections and modulation of connections).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The behavioral data generated in this study have been deposited in the Zenodo database under accession code (<https://doi.org/10.5281/zenodo.7808070>). The raw fMRI data are protected and are not available due to data privacy laws, but group-level fMRI results are available by request to the corresponding author. Source data are provided with this paper.

Code availability

The codes of the behavioral and fMRI analyses of this study are available in a Zenodo deposit (<https://doi.org/10.5281/zenodo.7808070>).

References

- Bar, M. The proactive brain: using analogies and associations to generate predictions. *Trends Cogn. Sci.* **11**, 280–289 (2007).
- Friston, K. Learning and inference in the brain. *Neural Netw.* **16**, 1325–1352 (2003).
- Rao, R. P. & Ballard, D. H. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* **2**, 79–87 (1999).
- Clark, A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* **36**, 181–204 (2013).
- Friston, K. The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* **11**, 127–138 (2010).
- Friston, K. A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **360**, 815–836 (2005).
- Feldman, H. & Friston, K. Attention, uncertainty, and free-energy. *Front. Hum. Neurosci.* **4**, 215 (2010).
- Brock, J. Alternative Bayesian accounts of autistic perception: comment on Pellicano and Burr. *Trends Cogn. Sci.* **16**, 573–574 (2012).
- Lawson, R. P., Rees, G. & Friston, K. J. An aberrant precision account of autism. *Front. Hum. Neurosci.* **8**, 302 (2014).
- Palmer, C. J., Lawson, R. P. & Hohwy, J. Bayesian approaches to autism: towards volatility, action, and behavior. *Psychol. Bull.* **143**, 521–542 (2017).
- Pellicano, E. & Burr, D. When the world becomes ‘too real’: a Bayesian explanation of autistic perception. *Trends Cogn. Sci.* **16**, 504–510 (2012).
- Sinha, P. et al. Autism as a disorder of prediction. *Proc. Natl. Acad. Sci. USA.* **111**, 15220–15225 (2014).
- Van de Cruys, S. et al. Precise minds in uncertain worlds: predictive coding in autism. *Psychol. Rev.* **121**, 649–675 (2014).
- Gomot, M. & Wicker, B. A challenging, unpredictable world for people with autism spectrum disorder. *Int. J. Psychophysiol.* **83**, 240–247 (2012).
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders: Dsm-5*. (American Psychiatric Publishing). (2013)
- Maenner, M. J. Prevalence of autism spectrum disorder among children aged 8 years — autism and developmental disabilities monitoring network, 11 sites, United States, 2016. *MMWR Surveill. Summ.* **69**, 1–12 (2020).
- Jenkinson, R., Milne, E. & Thompson, A. The relationship between intolerance of uncertainty and anxiety in autism: a systematic literature review and meta-analysis. *Autism Int. J. Res. Pract.* 1362361320932437 <https://doi.org/10.1177/1362361320932437> (2020).
- Soulières, I., Mottron, L., Giguère, G. & Laroche, S. Category induction in autism: slower, perhaps different, but certainly possible. *Q. J. Exp. Psychol.* 2006 **64**, 311–327 (2011).
- Haker, H., Schneebeli, M. & Stephan, K. E. Can bayesian theories of autism spectrum disorder help improve clinical practice? *Front. Psychiatry* **7**, 107 (2016).
- Cannon, J., O'Brien, A. M., Bungert, L. & Sinha, P. Prediction in autism spectrum disorder: a systematic review of empirical evidence. *Autism Res.* <https://doi.org/10.1002/aur.2482> (2021).
- Croydon, A., Karaminis, T., Neil, L., Burr, D. & Pellicano, E. The light-from-above prior is intact in autistic children. *J. Exp. Child Psychol.* **161**, 113–125 (2017).
- Van de Cruys, S. et al. Structural and contextual priors affect visual search in children with and without autism. *Autism Res.* <https://doi.org/10.1002/aur.2511> (2021).
- Lieder, I. et al. Perceptual bias reveals slow-updating in autism and fast-forgetting in dyslexia. *Nat. Neurosci.* **22**, 256–264 (2019).
- Sapey-Triomphe, L.-A., Weilhhammer, V. A. & Wagemans, J. Associative learning under uncertainty in adults with autism: Intact learning of the cue-outcome contingency, but slower updating of priors. *Autism* <https://doi.org/10.1177/13623613211045026> (2021).

25. Manning, C., Kilner, J., Neil, L., Karaminis, T. & Pellicano, E. Children on the autism spectrum update their behaviour in response to a volatile environment. *Dev. Sci.* **20**, e12435 (2016).
26. Lawson, R. P., Mathys, C. & Rees, G. Adults with autism overestimate the volatility of the sensory environment. *Nat. Neurosci.* **20**, 1293–1299 (2017).
27. Goris, J. et al. Sensory prediction errors are less modulated by global context in autism spectrum disorder. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **3**, 667–674 (2018).
28. Balsters, J. H. et al. Disrupted prediction errors index social deficits in autism spectrum disorder. *Brain J. Neurol.* **140**, 235–246 (2017).
29. Mosner, M. G. et al. Neural mechanisms of reward prediction error in autism spectrum disorder. *Autism Res. Treat.* **2019**, e5469191 (2019).
30. Kinard, J. L. et al. Neural mechanisms of social and nonsocial reward prediction errors in adolescents with autism spectrum disorder. *Autism Res.* **13**, 715–728 (2020).
31. Iglesias, S. et al. Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. *Neuron* **80**, 519–530 (2013).
32. Iglesias, S. et al. Cholinergic and dopaminergic effects on prediction error and uncertainty responses during sensory associative learning. *NeuroImage* **226**, 117590 (2021).
33. Weinhhammer, V. A., Stuke, H., Sterzer, P. & Schmack, K. The neural correlates of hierarchical predictions for perceptual decisions. *J. Neurosci.* **38**, 5008–5021 (2018).
34. Vilares, I., Howard, J. D., Fernandes, H. L., Gottfried, J. A. & Kording, K. P. Differential representations of prior and likelihood uncertainty in the human brain. *Curr. Biol.* **22**, 1641–1648 (2012).
35. Ide, J. S., Shenoy, P., Yu, A. J. & Li, C. R. Bayesian prediction and evaluation in the anterior cingulate cortex. *J. Neurosci.* **33**, 2039–2047 (2013).
36. Ficco, L. et al. Disentangling predictive processing in the brain: a meta-analytic study in favour of a predictive network. *Sci. Rep.* **11**, 16258 (2021).
37. Bell, A. H., Summerfield, C., Morin, E. L., Malecek, N. J. & Ungerleider, L. G. Encoding of stimulus probability in macaque inferior temporal cortex. *Curr. Biol.* **26**, 2280–2290 (2016).
38. Bar, M. Visual objects in context. *Nat. Rev. Neurosci.* **5**, 617–629 (2004).
39. Bar, M. & Aminoff, E. Cortical analysis of visual context. *Neuron* **38**, 347–358 (2003).
40. Mathys, C. D. et al. Uncertainty in perception and the hierarchical gaussian filter. *Front. Hum. Neurosci.* **8**, 825 (2014).
41. den Ouden, H. E. M., Daunizeau, J., Roiser, J., Friston, K. J. & Stephan, K. E. Striatal prediction error modulates cortical coupling. *J. Neurosci.* **30**, 3210–3219 (2010).
42. Corbett, J. E., Venuti, P. & Melcher, D. Perceptual averaging in individuals with autism spectrum disorder. *Front. Psychol.* **7**, 1735 (2016).
43. Ego, C. et al. Behavioral characterization of prediction and internal models in adolescents with autistic spectrum disorders. *Neuropsychologia* **91**, 335–345 (2016).
44. Van de Cruys, S., Vanmarcke, S., Van de Put, I. & Wagemans, J. The use of prior knowledge for perceptual inference is preserved in ASD. *Clin. Psychol. Sci.* <https://doi.org/10.1177/2167702617740955> (2017).
45. Van de Cruys, S., Vanmarcke, S., Steyaert, J. & Wagemans, J. Intact perceptual bias in autism contradicts the decreased normalization model. *Sci. Rep.* **8**, 12559 (2018).
46. Utzerath, C., Schmits, I. C., Kok, P., Buitelaar, J. & de Lange, F. P. No evidence for altered up- and downregulation of brain activity in visual cortex during illusory shape perception in autism. *Cortex* **117**, 247–256 (2019).
47. Solomon, M., Smith, A. C., Frank, M. J., Ly, S. & Carter, C. S. Probabilistic reinforcement learning in adults with autism spectrum disorders. *Autism Res.* **4**, 109–120 (2011).
48. Goris, J. et al. Autistic traits are related to worse performance in a volatile reward learning task despite adaptive learning rates. *Autism* <https://doi.org/10.1177/1362361320962237> (2020).
49. Margulies, D. S. et al. Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proc. Natl. Acad. Sci.* **113**, 12574–12579 (2016).
50. Paquola, C. et al. Microstructural and functional gradients are increasingly dissociated in transmodal cortices. *PLOS Biol.* **17**, e3000284 (2019).
51. Schmack, K., Rothkirch, M., Priller, J. & Sterzer, P. Enhanced predictive signalling in schizophrenia. *Hum. Brain Mapp.* **38**, 1767–1779 (2017).
52. Ranganath, C., Cohen, M. X., Dam, C. & D'Esposito, M. Inferior temporal, prefrontal, and hippocampal contributions to visual working memory maintenance and associative memory retrieval. *J. Neurosci.* **24**, 3917–3925 (2004).
53. Aminoff, E., Gronau, N. & Bar, M. The parahippocampal cortex mediates spatial and nonspatial associations. *Cereb. Cortex N.Y.* **17**, 1493–1503 (2007).
54. Payzan-LeNestour, E., Dunne, S., Bossaerts, P. & O'Doherty, J. P. The neural representation of unexpected uncertainty during value-based decision making. *Neuron* **79**, 191–201 (2013).
55. Robinson, S., Keene, C. S., Iaccarino, H. F., Duan, D. & Bucci, D. J. Involvement of retrosplenial cortex in forming associations between multiple sensory stimuli. *Behav. Neurosci.* **125**, 578–587 (2011).
56. Gilad, A. & Helmchen, F. Spatiotemporal refinement of signal flow through association cortex during learning. *Nat. Commun.* **11**, 1744 (2020).
57. Fournier, D. I., Eddy, M. C., DeAngeli, N. E., Huszár, R. & Bucci, D. J. Retrosplenial cortex damage produces retrograde and anterograde context amnesia using strong fear conditioning procedures. *Behav. Brain Res.* **369**, 111920 (2019).
58. Barack, D. L. & Platt, M. L. Neuronal activity in the posterior cingulate cortex signals environmental information and predicts behavioral variability during trapline foraging. *J. Neurosci.* **41**, 2703–2712 (2021).
59. Schmack, K. et al. Delusions and the role of beliefs in perceptual inference. *J. Neurosci.* **33**, 13701–13712 (2013).
60. Farovik, A. et al. Orbitofrontal cortex encodes memories within value-based schemas and represents contexts that guide memory retrieval. *J. Neurosci.* **35**, 8333–8344 (2015).
61. Cooper, J. C., Dunne, S., Furey, T. & O'Doherty, J. P. Human dorsal striatum encodes prediction errors during observational learning of instrumental actions. *J. Cogn. Neurosci.* **24**, 106–118 (2012).
62. Taylor, A. J., Kim, J. H. & Ress, D. Characterization of the hemodynamic response function across the majority of human cerebral cortex. *NeuroImage* **173**, 322–331 (2018).
63. Chaudhuri, R., Knoblauch, K., Gariel, M.-A., Kennedy, H. & Wang, X.-J. A large-scale circuit mechanism for hierarchical dynamical processing in the primate cortex. *Neuron* **88**, 419–431 (2015).
64. Stocker, A. A. & Simoncelli, E. P. A bayesian model of conditioned perception. *Adv. Neural Inf. Process. Syst.* **2007**, 1409–1416 (2007).
65. Wechsler D. Wechsler Adult Intelligence Scale, Fourth Edition (2008).
66. Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J. & Clubley, E. The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *J. Autism Dev. Disord.* **31**, 5–17 (2001).
67. Hoekstra, R. A., Bartels, M., Cath, D. C. & Boomsma, D. I. Factor structure, reliability and criterion validity of the Autism-Spectrum

- Quotient (AQ): a study in Dutch population and patient groups. *J. Autism Dev. Disord.* **38**, 1555–1566 (2008).
68. Carleton, R. N., Norton, M. A. P. J. & Asmundson, G. J. G. Fearing the unknown: a short version of the Intolerance of Uncertainty Scale. *J. Anxiety Disord.* **21**, 105–117 (2007).
69. de Bruin, G. O., Rassin, E., van der Heiden, C. & Muris, P. Psychometric properties of a Dutch version of the Intolerance of Uncertainty Scale. *Neth. J. Psychol.* **62**, 87–92 (2006).
70. Robertson, A. E. & Simmons, D. R. The relationship between sensory sensitivity and autistic traits in the general population. *J. Autism Dev. Disord.* **43**, 775–784 (2013).
71. Kuiper, M. W., Verhoeven, E. W. & Geurts, H. M. The dutch glasgow sensory questionnaire: psychometric properties of an autism-specific sensory sensitivity measure. *Autism* **23**, 922–932 (2019).
72. Sapey-Triomphe, L.-A., Temmerman, J., Puts, N. A. J. & Wagemans, J. Prediction learning in adults with autism and its molecular correlates. *Mol. Autism* **12**, 64 (2021).
73. Cohen, J. *Statistical power analysis for the behavioral sciences.* (Routledge, 1988).
74. Sawilowsky, S. S. New effect size rules of thumb. *J. Mod. Appl. Stat. Methods* **8**, 597–599 (2009).
75. Mathys, C., Daunizeau, J., Friston, K. J. & Stephan, K. E. A bayesian foundation for individual learning under uncertainty. *Front. Hum. Neurosci.* **5**, 39 (2011).
76. Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J. & Friston, K. J. Bayesian model selection for group studies. *NeuroImage* **46**, 1004–1017 (2009).
77. Behzadi, Y., Restom, K., Liu, J. & Liu, T. T. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage* **37**, 90–101 (2007).
78. Brett, M., Anton, J.-L., Valabregue, R. & Poline, J.-B. Region of interest analysis using an SPM toolbox. Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2–6, 2002, Sendai, Japan. (2002).
79. Bob Spunt. *spunt/bspmview: BSPMVIEW v.20161108.* (Zenodo). <https://doi.org/10.5281/zenodo.168074> (2016).
80. Friston, K. J., Harrison, L. & Penny, W. Dynamic causal modelling. *NeuroImage* **19**, 1273–1302 (2003).
81. Sherman, M. T., Seth, A. K. & Kanai, R. Predictions shape confidence in right inferior frontal gyrus. *J. Neurosci.* **36**, 10323–10336 (2016).
82. Trapp, S. & Bar, M. Prediction, context, and competition in visual recognition. *Ann. N. Y. Acad. Sci.* **1339**, 190–198 (2015).
83. Allen, M. et al. Anterior insula coordinates hierarchical processing of tactile mismatch responses. *NeuroImage* **127**, 34–43 (2016).
84. Cardin, V., Friston, K. J. & Zeki, S. Top-down modulations in the visual form pathway revealed with dynamic causal modeling. *Cereb. Cortex* **21**, 550–562 (2011).
85. Sapey-Triomphe, L.-A. et al. Ventral stream hierarchy underlying perceptual organization in adolescents with autism. *NeuroImage Clin* **25**, 102197 (2020).

Acknowledgements

We would like to thank Naomi Couder and Joke Temmerman for their help in recruiting participants and collecting data. We also thank Ronald Peeters for his help in preparing the MRI exam card. Finally, we thank all the participants for their time. LAST was supported by a postdoctoral fellowship of the H2020 Marie Skłodowska-Curie Actions (PreCoASD). The research was financed by grants from a long-term structural funding from the Flemish Government (METH/14/02) to JW.

Author contributions

All authors contributed to the conception of the study. L.A.S.T. and L.P. piloted the experiment and collected the data. L.A.S.T. and L.P. analyzed the behavioral data. L.A.S.T. analyzed the fMRI data. L.A.S.T., V.W., and P.S. modeled the data. J.W. was the main supervisor of the project. L.A.S.T. wrote the first draft of the article, and J.W. edited it. All authors approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-38580-9>.

Correspondence and requests for materials should be addressed to Laurie-Anne Sapey-Triomphe.

Peer review information *Nature Communications* thanks Boris Bernhardt and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023