Adults with autism overestimate the volatility of the sensory environment

Rebecca P Lawson[1-3] , Christoph Mathys[1,4-6] & Geraint Rees[1,2]

Insistence on sameness and intolerance of change are among the diagnostic criteria for autism spectrum disorder (ASD), but little research has addressed how people with ASD represent and respond to environmental change. Here, behavioral and pupillometric measurements indicated that adults with ASD are less surprised than neurotypical adults when their expectations are violated, and decreased surprise is predictive of greater symptom severity. A hierarchical Bayesian model of learning suggested that in ASD, a tendency to overlearn about volatility in the face of environmental change drives a corresponding reduction in learning about probabilistically aberrant events, thus putatively rendering these events less surprising. Participant-specific modeled estimates of surprise about environmental conditions were linked to pupil size in the ASD group, thus suggesting heightened noradrenergic responsivity in line with compromised neural gain. This study offers insights into the behavioral, algorithmic and physiological mechanisms underlying responses to environmental volatility in ASD.

When negotiating changeable real-world environments, humans face a set of learning problems involving different forms of uncertainty, in which the weighting of new evidence and prior expectations must be dynamically adjusted. Imagine opening your sock drawer and finding a pineapple inside. How surprised should you be? Under normal circumstances, you would expect to see socks, but if your four-year-old niece is visiting, you might adjust your expectations to suit a more volatile environment, thereby lessening any surprise. However, overestimating how volatile your bedroom is may result in compromised learning of the association between the cue (sock drawer) and outcome (socks) in the first place. In other words, aberrant representation of volatility may impair the dynamic formation of appropriate prior expectations, thereby rendering both the pineapple and the socks mildly surprising. Bayesian theories of perception in people with ASD (although we abide by the terminology of the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-5), we wish to acknowledge that the term 'autistic person' is preferred by many people on the spectrum)[1] propose that reduced weighting of prior expectations relative to sensory inputs leads to the perceptual atypicalities associated with the condition[2-7]; however, no studies to date have actually quantified the learning dynamics by which sensory expectations are formed in ASD. Here we sought to empirically address whether volatility learning is compromised in ASD[6,7].

Computationally, the amount of weight given to a surprising event is determined by its precision (inverse variance, proportional to learning rate α), with α determining the rate of integration over past events to predict future outcomes. Although computational studies of decision-making regarding rewards and punishments have shown that participants adapt their rate of learning about action–outcome contingencies in response to changes in environmental volatility[8-10], these models do not fit individual differences in volatility learning. However, knowing whether to disregard an unexpected outcome or to take it seriously (i.e., whether to adopt a high or low learning rate about cue–outcome probabilities) depends on the precision of one's beliefs about environmental change (i.e., whether one adopts a high or low learning rate regarding volatility). The recent application of hierarchical learning models has allowed for the quantification of individual learning about both probabilistic relationships and how these relationships change over time (volatility)[11-14], but no studies have applied these models to understand learning about uncertainty in ASD.

In a state in which uncertainty about one's beliefs is high (for example, in volatile conditions), top-down prior expectations should be suppressed relative to new bottom-up sensory evidence, to promote new learning about the current environmental context[15]. With their broad distribution and extensive connectivity, neuromodulatory systems are ideally positioned to facilitate the widespread changes in neural gain necessary to support such a function[16]. Noradrenaline (NA), in particular, is thought to signal contextual change, thus leading to enhanced bottom-up, thalamocortical transmission of sensory information[17-19]. Recent neurocomputational accounts of autism have proposed that aberrant signaling of volatility may result in pathological neural gain consistent with the cognitive and perceptual profile of autism, such as enhanced perceptual functioning, sensory overload and context insensitivity[4-6,20].

Here, we tested these computational and neurobiological hypotheses by examining how adults with ASD responded to experimentally

[1]Wellcome Trust Centre for Neuroimaging, University College London, London, UK. [2]Institute of Cognitive Neuroscience, University College London, London, UK. [3]Department of Psychology, University of Cambridge, Cambridge, UK. [4]Scuola Internazionale Superiore di Studi Avanzati (SISSA), Trieste, Italy. [5]Max Planck UCL Centre for Computational Psychiatry and Ageing, University College London, London, UK. [6]Translational Neuromodeling Unit, Institute for Biomedical Engineering, University of Zurich and Eidgenössische Technische Hochschule (ETH) Zurich, Zurich, Switzerland. Correspondence should be addressed to R.P.L. (rebecca.lawson@ucl.ac.uk).

1293

manipulated changes in their sensory expectations that independently assessed changes in the category of a stimulus, the informativeness of a cue predicting its appearance and changes in these associations over time. To do so, we used a hierarchical Bayesian model that allowed us to characterize each participant's learning 'fingerprint', specifically the simultaneous learning about multiple different sources of environmental uncertainty[11]. We hypothesized that adults with ASD would show reduced behavioral and neurophysiological responses in statistical contrasts of 'unexpected' (UE) and 'expected' (E) trials based on the experimental 'ground truth' (for example, reduced surprise when surprise would be expected for neurotypical (NT) individuals). This hypothesis is in line with results of previous studies showing a decreased distinction between repeated and novel stimuli in ASD[21-23]. However, we hypothesized that computational modeling of the actual learning process for each individual would demonstrate an increased tendency to represent and respond to environmental volatility in ASD, thus compromising learning about probabilistic relationships in the environment. Accordingly, we hypothesized that adults with ASD would show increased sensitivity of the pupil dilatory response to computational metrics of prediction error, which estimate when each individual is actually surprised.

RESULTS

We used a modified version of a common probabilistic associative learning task[24] to test the effects of learned expectations and sensory noise on behavior (reaction times (RT), error rates) and indices of phasic NA function (pupillometry)[25] in adults with ASD ($n = 24$) and age- and intelligence quotient (IQ)-matched NT adults ($n = 25$) (Online Methods).

Participants performed binary classification of images as either faces or houses, and images had high, medium or no noise added. A tone preceding each image was highly, weakly or not predictive of a given outcome, and these image–tone associations changed across time (**Fig. 1**), such that trials could be categorized as E, UE or neutral (N). This process created a ground-truth structure in the environment that participants had to implicitly learn. In contrast to reinforcement-learning[26,27], implicit-motor-learning[28] and serial-reaction-time[29] tasks that have examined sensitivity to probability manipulations in ASD, this task addresses perceptual associative learning and explicitly manipulated three different forms of uncertainty (categorical sensory uncertainty, probabilistic uncertainty and environmental uncertainty). When a participant received an unexpected outcome, the outcome might have reflected a probabilistically aberrant event, or it might have signaled that the environmental context had changed. To quantify individual learning about these different forms of uncertainty, RTs were modeled with a Bayesian belief-update scheme[11] (Online Methods). The model implied each participant's beliefs about these quantities, as reflected in the sequence of cue–outcome associations received by each participant and their trial-by-trial responses and response times.

Behavior

First, we examined behavioral responses, and E and UE trials were categorized according to the ground truth.

Reaction times. Reaction times (RTs) were subjected to a 3 × 3 mixed ANOVA with within-subject factors of expectedness (E, N and UE), noise (high, medium or none) and a between-participant factor of group (ASD and NT). There was a significant main effect of expectedness ($F_{(2,94)} = 25.48$, $P < 0.001$) and noise ($F_{(2,94)} =$
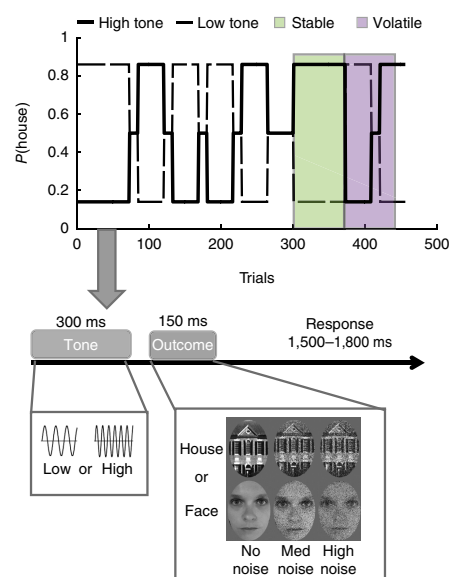

Figure 1 Task structure. Schematic of the task, showing the volatile environmental structure (top), for example, the probability of seeing a house (given the preceding high or low tone) across trials. The green area shows a 'stable' period of 72 trials in which the probabilities remained fixed, and the violet area shows a 'volatile' period of 72 trials in which the outcome probabilities switched three times. A single trial is also seen (bottom) showing example stimuli. Med, medium.

13.60, $P < 0.001$), thus indicating that RTs were slower for UE and high-noise stimuli relative to E and low-noise stimuli. A significant main effect of group ($F_{(1,47)} = 4.83$, $P = 0.03$) indicated that, in general, the participants with ASD were slower to respond than the NT participants. Crucially, only the expectedness × group interaction was significant in this analysis ($F_{(2,94)} = 4.47$, $P = 0.014$; **Fig. 2a**). The noise × group ($F_{(2,94)} = 0.06$, $P = 0.94$), noise × expectedness ($F_{(4,188)} = 0.47$, $P = 0.76$) and expectedness × noise × group interactions were not significant ($F_{(4,188)} = 1.31$, $P = 0.28$). These results suggested that for both groups, increasing sensory noise resulted in slower RT (**Supplementary Fig. 1**), but adults with ASD showed reduced modulation of RT as a function of learned expectations. This result was consistent with a decreased influence of prior information on perception and action in ASD[2], although future studies should explore how learned expectations affect perceiving structure in true noise or 50/50 composite images in which reliance on prior beliefs should be greater.

The results were unchanged when the identical analysis was carried out on log reaction times (**Supplementary Table 1**).

Subtracting the RTs to E outcomes from the RTs to UE outcomes provided a low-level index of 'surprise', which was significantly greater than zero in both groups (ASD, $t_{(23)} = 4.66$, $P < 0.001$; NT, $t_{(24)} = 7.25$, $P < 0.001$) but was attenuated in the ASD group relative to the NT group ($t_{(47)} = 3.51$, $P = 0.001$; **Fig. 2b**). These results suggested less distinction between UE and E outcomes in ASD, though this effect was conditioned upon adequate learning of the ground truth.

To ensure that the group difference in UE – E RT persisted over and above participants' mean 'baseline' RT and error rates, we conducted a linear regression to predict UE – E RT with group (ASD and NT), mean RT and mean errors as predictors. This model was significant overall ($F_{(3,48)} = 5.58$, $P = 0.002$), and the only significant predictor of the UE – E RT difference was group ($t = -2.87$, $P = 0.008$). Mean RT ($t = -1.08$. $P = 0.28$) and mean error ($t = 1.06$, $P = 0.29$) were not significant predictors. Importantly, this analysis demonstrated that the diminished effects

of behavioral surprise in participants with ASD persisted even when the variance associated with general response speeds and accuracy was included in the model.

Additional analyses confirmed that this key finding of group differences in UE – E RTs remained present when control analyses accounted for the effects of the speed–accuracy trade-off (**Supplementary Fig. 2**) and general group differences in caution of responding (**Supplementary Fig. 3**). We do, however, recognize that the slower overall responses (and higher accuracy) in the ASD group may indicate a tendency to manage uncertainty with increased response thresholds, which could be tested with drift diffusion models[30,31] in future studies in which error rates are higher by design.

Error rates. The same analysis as above was conducted for error rates. There was a significant main effect of expectedness ($F_{(1.5,70.5)} = 11.71$, $P < 0.001$) and a significant group × expectedness interaction ($F_{(2,94)} = 6.34$, $P = 0.003$), thus indicating that the NT group made more errors in UE relative to E trials, whereas the ASD group did not (**Fig. 2c**). The main effect of noise was not significant in this analysis ($F_{(1.7,78.8)} = 0.08$, $P = 0.92$), and neither were the noise × group ($F_{(2,94)} = 0.29$, $P = 0.75$), noise × expectedness ($F_{(4,188)} = 0.76$, $P = 0.55$) and expectedness × noise × group interactions ($F_{(4,188)} = 1.28$, $P = 0.28$).

The results were very similar when the identical analysis was carried out on log error rates (**Supplementary Table 1**).

Subtracting the percentage errors to E outcomes from those to UE outcomes provided a low-level index of surprise, which was significantly greater than zero in only the NT group (ASD, $t(23) = 1.11$, $P = 0.28$; NT, $t(24) = 3.65$, $P = 0.001$) and was attenuated in the ASD group relative to the NT group ($t(33.4) = 2.83$, $P = 0.007$; **Fig. 2d**).

Relation to symptoms. To explore the relationship between behavioral surprise and ASD symptom severity we conducted a multiple linear regression predicting the UE – E RT measure with the Autism Diagnostic Observation Scale (ADOS-2) communication, social reciprocal interaction scores and IQ as predictors. This model was significant ($F_{(3,23)} = 3.28$, $P = 0.04$), and communication score was the only significant predictor ($t = -2.57$, $P = 0.018$; **Fig. 3**). IQ ($t = 1.45$, $P = 0.16$) and social reciprocal interaction scores ($t = 0.95$, $P = 0.35$) did not predict behavioral surprise.

A second regression model that also contained baseline RT as a predictor narrowly missed the overall significance threshold ($F_{(6,23)} = 2.41$, $P = 0.07$), and communication score was once again the only significant predictor ($t = -2.81$, $P = 0.012$). A third regression model, in a reduced sample size (Online Methods), additionally included sensory sensitivity scores as a predictor of UE – E RT. This model was not significant ($F_{(4,21)} = 1.28$, $P = 0.32$), and the only predictor approaching significance was, again, communication score ($t = -1.89$, $P = 0.076$).

Communication, as measured by the ADOS-2, predominantly weights stereotyped and repetitive speech and conversational reciprocity, which arguably necessitate reflexive behavioral responses to change. Future studies should examine the specificity of this link between general behavioral adaptations to learned expectations and communication abilities, especially as measured by different instruments.

Nonclinical replication. Beyond the range of clinical phenotypes seen in people diagnosed with ASD, a wider continuum of social-communicative ability is expressed as autistic traits in the general population[32]. Encouragingly, the relationship between our behavioral measure of surprise (UE – E RT) and autistic tendency was replicated in an independent nonclinical sample ($n = 57$) of participants characterized according to expression of autistic traits (**Supplementary Fig. 4**). This result not
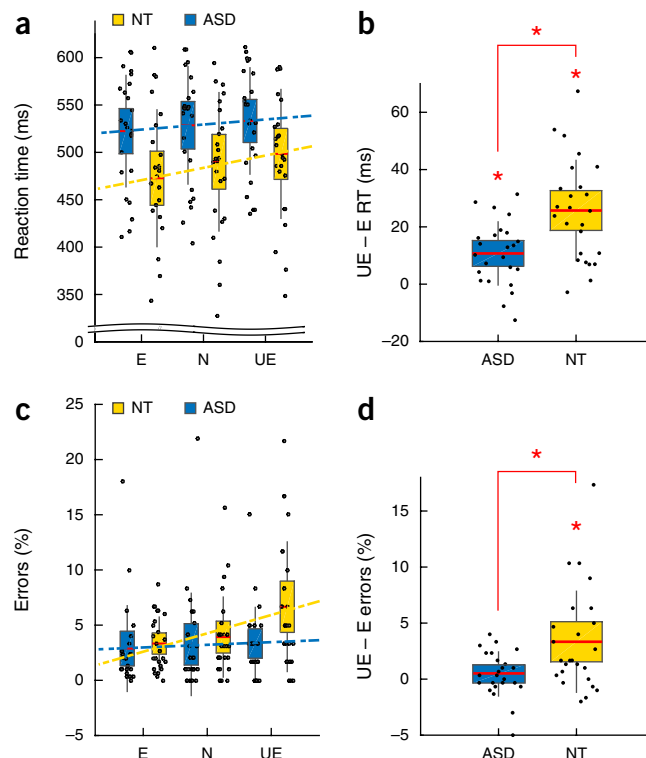

Figure 2 Behavioral results based on the ground truth. (**a**–**d**) Reduced modulation of reaction time (**a**,**b**) and error rates (**c**,**d**), as a function of expectation in adults with ASD ($n = 24$) relative to NT ($n = 25$). Dotted lines show linear fits. Data points indicate individual participants. Thick red horizontal line mean; shaded regions ± error bars, 95% confidence intervals ± 1 s.d. of the mean for each condition and group. *$P < 0.05$ for one-sample and paired t tests (two tailed).

only bolsters confidence in our clinical finding but additionally supports generalization of our findings to the broader autism spectrum in the wider population.

Responses to different stimulus types. Finally, control analyses indicated that there were no group differences in response time or accuracy across the face and house stimuli (**Supplementary Fig. 5**).

Computational modeling

To investigate learning about distinct kinds of uncertainty in ASD, we adopted a participant-specific Bayesian model to track the role of uncertainty on behavior (log RTs). In the hierarchical Gaussian filter (HGF)[11], beliefs are updated via prediction errors, with dynamic learning rates (α) at each level (i) influenced by uncertainty about the accuracy of current beliefs and environmental volatility (**Fig. 4a**). In the version of the HGF used here (introduced in ref. 33), learning occurs simultaneously on three coupled levels of an uncertainty hierarchy (x_1, x_2 and x_3). Level 1 (x_1) addresses uncertainty about outcomes (face or house), level 2 (x_2) addresses uncertainty about probabilities (cue–outcome contingencies), and level 3 (x_3) addresses uncertainty about environmental change (volatility). Additional model details can be found in Online Methods and **Supplementary Table 2**.

Model validation. First, to ensure that the HGF performed well as a model describing the behavior of our participants, we fit three alternative learning models to the data and compared them to the HGF with random-effects Bayesian-model selection (BMS). Relative

to simple reinforcement-learning (RL) models with fixed (RW) and dynamic (SK1) learning rates and a two-level HGF in which volatility updates were eliminated, the three-level HGF was the best model for explaining the data by a considerable margin (Online Methods and **Supplementary Fig. 6**). Importantly, BMS evaluates the relative plausibility of competing models in terms of their log evidences by quantifying the trade-off between the accuracy (fit) and complexity of a model, and it accounts for the possibility that the observed variability in log model evidence could be due to chance. Additionally, the three-level HGF model simulations captured the principal group differences in the behavioral effect of expectation on RT (Online Methods and **Supplementary Fig. 7**).

Predicting diagnostic status. A summary of group differences in each of the estimated model parameters is presented in **Supplementary Figure 8**. A binary logistic regression model predicting group status (ASD = 1, NT = 0), with all eight model parameters as predictors was significant ($\chi^2 = 26.83$, $P = 0.001$), and the overall prediction success was 81.6% (76% ASD, 88% NT), with cross-validated prediction success of 68%. The Wald statistic demonstrated that outcome uncertainty (β_2, $P = 0.043$), phasic volatility (β_4, $P = 0.006$) tonic volatility at the third level (ω_3, $P = 0.024$) and baseline log RT (β_0, $P = 0.007$) made a significant contribution to prediction (**Fig. 4b**).

Interestingly, these significant predictors predominantly pertained to the third level of the HGF, i.e., learning about environmental volatility. ω_3 can be understood as capturing 'metavolatility' (i.e., the tonic volatility of the phasic volatility, and the higher values in the ASD group indicated a belief in a world in which instability itself is instable (**Supplementary Fig. 8**). β_4 captures the modulation of log RT in response to phasic volatility; here, the smaller (negative) values in the NT group (**Supplementary Fig. 8**) indicated that when beliefs about volatility increased, participants became more attentive and responded faster. In contrast, the larger (positive) values in the ASD group (**Supplementary Fig. 8**) indicated that increased beliefs about volatility led to a slower reaction time. In general, these findings indicated difficulties in representing and responding to environmental change in ASD, specifically an increased tendency to expect the unexpected.

Learning-rate update in response to volatility. From the HGF, we inferred the trialwise rate of learning about two different sources of information: probabilistic outcomes (α_2) and the rate of learning about environmental change (α_3). When the environment is volatile, people should give more weight to recent sensory outcomes in building expectations about what they will see next (for example, adopt a high α); in contrast, they should give information from the distant past more weight when the environment is stable (for example, adopt a low α)[8,9]. To test the hypothesis that individuals with ASD have difficulties in flexibly updating their rate of learning (i.e., precision weighting) in response to environmental change, we examined the change (Δ) in α_2 (probability) and α_3 (environment) when switching from stable (violet in **Fig. 1**) to volatile (green in **Fig. 1**) periods of the task. We compared the change in α_2 and α_3 between these two periods, across the groups. This analysis revealed a trend toward a main effect of group ($F(1,47) = 0.26$, $P = 0.061$), a significant main effect of α type ($F(1,47) = 6.07$, $P = 0.017$) and crucially an α type × group interaction ($F(1,47) = 9.80$), $P = 0.003$). Follow-up with independent-sample t tests revealed that the ASD group did not update α_2 as much as the NT group ($t(47) = -2.37$, $P = 0.02$), whereas the ASD group updated α_3 more than the NT group ($t(47) = 3.16$, $P = 0.03$; **Fig. 4c**).
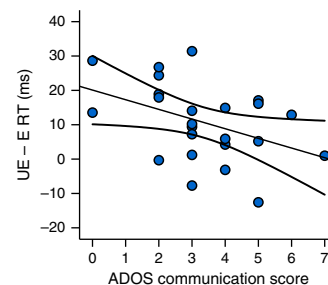

Figure 3 Relationship between behavioral surprise and symptoms. The magnitude of the reaction UE – E RT effect is predicted by communication symptoms in the ASD group (Pearson correlation, $r = -0.421$, $P = 0.04$, two tailed). Data points indicate individual participants with ASD ($n = 24$). Lines show linear fit and mean confidence intervals.

Average learning rates. To examine learning overall, we calculated average values for α_2 and α_3 for each participant. This analysis revealed no main effect of α type ($F(1,47) = 2.61$, $P = 0.11$), no main effect of group ($F(1,47) = 2.01$, $P = 0.16$) and no group × α type interaction ($F(1,47) = 2.54$, $P = 0.12$), thus suggesting that, in general, both groups were able to learn this task equally well.

Predicting learning-rate update from tonic volatility. Finally, because the HGF estimation does not fit α_2 and α_3 directly, we ran two linear regression models separately predicting $\Delta\alpha_2$ and $\Delta\alpha_3$ to determine which of the two ω parameters drove these differences in learning rate. In each case, the model was significant ($\Delta\alpha_2$, $F(2,48) = 68.94$, $P < 0.001$, $R^2 = 0.75$; $\Delta\alpha_3$, $F(2,48) = 102.53$, $P < 0.001$, $R^2 = 0.82$). The results indicated that $\Delta\alpha_2$ was positively predicted by ω_2 ($t = 2.72$, $P = 0.009$), thus suggesting that a tendency to believe that cue–outcome associations were unstable was associated with a larger update in $\Delta\alpha_2$ when switching from stable to volatile phases of the task. Interestingly, ω_3 negatively predicted $\Delta\alpha_2$ ($t = -8.89$, $P > 0.001$), thus indicating that a tendency to believe that instability was unstable drove a smaller update in α_2 in response to volatility. This result was consistent with our finding that the participants with ASD, who tended to have smaller $\Delta\alpha_2$ (**Fig. 4c**), showed reduced behavioral surprise (**Fig. 2b**) and also larger 'metavolatility' estimates (**Fig. 4b**). For the model predicting $\Delta\alpha_3$, both of the ω parameters were significant positive predictors (ω_2, $t = 7.88$, $P < 0.001$; ω_3, $t = 14.24$, $P < 0.001$). For the participants with ASD, who, on average, showed larger $\Delta\alpha_3$, this result was consistent with a tendency toward beliefs in the instability of both cue–outcome associations and instability itself.

Pupillometry

Predictive coding descriptions of ASD depart from normative Bayesian theories in that they make explicit predictions about the neurobiological basis of precision, namely the action of neuromodulators such as NA, which control the gain on cortical responses (prediction errors)[3,4,6]. Increased NA signaling in ASD has been suggested by elevated blood plasma levels[34] and increased arousal, i.e., heart rate variability[35], but no studies have examined phasic NA function in the context of learning about uncertainty in ASD. To do so, we acquired concurrent pupillometry in a reduced subset of the sample (Online Methods). The phasic pupil response to surprising outcomes (ground truth contrast of UE – E trials) revealed a significant increase in pupil size in NT participants (**Fig. 5a**), which was consistent with findings from many previous studies[25]. In agreement with the behavioral data (**Fig. 2b,d**), the ASD group did not show this distinction between UE and E trials (**Fig. 5a**). This pattern mirrored previous findings in the domains of electrophysiology (reduced mismatch negativity in ASD
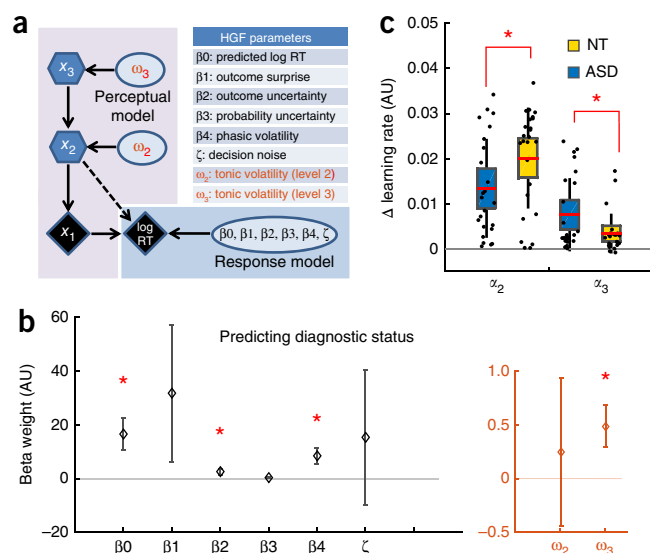
Figure 4 Computational-model details and results. (**a**) Schematic depiction of the three-level HGF. The perceptual model comprises three hierarchical states (x_1, x_2, and x_3). Participant-specific free parameters (ovals) are estimated from individual log RT data. Red parameters relate to the perceptual model, whereas black parameters relate to the response model. Diamonds, quantities that change over time (trials); hexagons quantities that change over time and that additionally depend on their previous state in time in a Markovian fashion. Details in main text. (**b**) Binary logistic regression. β weights for each of the HGF free parameters showing the contribution of each to predicting group status (ASD and NT) across all participants ($n = 49$). Significant predictors ($P < 0.05$ Wald statistic, two tailed) are denoted with asterisks. Error bars, s.e.m. for the β estimates. All parameters were included in the same model, but ω values are plotted on a separate scale (in red). Group differences in the model parameters at the level of individual subjects can be seen in **Supplementary Figure 8**. (**c**) Group differences in learning-rate update (i.e., change from stable to volatile periods of the task). Participants with ASD ($n = 24$) updated α_2 less than did NT participants ($n = 25$), whereas they updated α_3 more than did NT participants. *$P < 0.05$, two tailed, by independent-sample t tests. Data points indicate individual participants. Thick red horizontal line mean; shaded regions ± error bars, 95% confidence intervals ± 1 s.d. of the mean.

and smaller P300 (refs. 36,37), and blood oxygen level–dependent imaging (reduced functional magnetic resonance imaging repetition suppression in ASD[21,23])) but in the novel domain of pupillometry. However, this notion of surprise is conditioned upon adequate learning of the ground truth, and our computational analysis indicated that participants with ASD and the NT participants showed a dissociation in how they estimated volatility and adapted their learning rates in response to the changeability of the environment (**Fig. 4b,c**).

Computational pupillometry analysis. The HGF provided a nuanced and individualized trial-by-trial learning fingerprint and provided better characterization when participants were actually surprised as a function their personal learning process, namely 'high-level' precision-weighted prediction errors (PEs) about changes in cue–outcome contingency (ε_3). Here, the learning rate α_3 depended on the precision weight on the PE and was proportional to the update of environmental volatility (Online Methods). As such, ε_3 is a model-based measure of high-level surprise that is formally related to the dynamic learning about environmental change, in which we observed group differences (**Fig. 4c**). Applying multiple regression across every trial and every time point in the pupil time trace, we found a sustained positive relationship between

pupil size and precision-weighted PEs (ε_3) in the participants with ASD (**Fig. 5b**), and this relationship was significantly different from that in the NT group and from zero. Furthermore, these strong effects persisted when we controlled for the UE – E ground-truth contrast, trialwise differences in fixation compliance, mean RT and outcome image type (face/house), all of which were included in the model as covariates (Online Methods). Additional analyses revealed that the volatility learning rate (α_3) and the probability learning rate (α_2) were not encoded in the pupil response in either group (**Supplementary Fig. 9**; analysis examining the relationship between precision-weighted PEs in specific (volatile/stable) phases of the experiment in **Supplementary Fig. 10**).

Pupillometry control analyses. Given the possibility that people with ASD might look at face stimuli differently from people without ASD[38], the stimulus duration was purposefully made short (150 ms) to prevent saccades. Nonetheless, to ensure that there were no differences between the groups in fixation compliance across the stimulus types (faces and houses), we conducted a repeated-measures ANOVA on the mean absolute deviation (MAD) from fixation (in degrees of visual angle) across outcome image type (face and house) with a between-subject factor of group. All main effects and interactions in this analysis were nonsignificant (**Supplementary Table 1**).

To examine group differences in tonic pupil size (thought to be a measure of general noradrenergic tone[39]), we compared the average of the z-scored pupil measurement across all trials with an independent-samples t test. The results demonstrated no group differences in tonic pupil size in this sample ($t(23) = 0.36$, $P = 0.72$).

Finally, control analyses revealed that there were no group differences in fixation compliance across conditions (**Supplementary Fig. 11**) or the relationship between pupil size and simple behavior, such as trialwise RT (**Supplementary Fig. 12**). Raw pupil traces for each group are shown in **Supplementary Figure 13**.

DISCUSSION

In this study, behavioral (RT/error rates) and pupillometric results based on the experimental ground truth converged on the finding of a decreased distinction between unexpected and expected outcomes in ASD (**Figs. 2** and **5a**), in agreement with the results of many previous studies across a range of methods reporting decreased surprise in ASD[21,23,36,37]. Crucially, however, this low-level notion of E and UE trials assumed optimal or at least adequate learning of the ground truth. However, the statistical regularities that underlie the sensory world and shape expectations are changeable, and humans must learn about different kinds of uncertainty to adaptively adjust the weighting of prior expectations and sensory inputs. Knowing whether to disregard an unexpected outcome or to take it seriously (i.e., whether to adopt a high or low learning rate about cue–outcome probabilities (α_2)) depends on the precision of one's beliefs about environmental change (i.e., whether a high or low learning rate about volatility (α_3) is adopted). The present data extend previous work by specifically demonstrating that overestimating volatility in the face of environmental change—at the expense of learning about probabilistically aberrant events—characterizes the behavior of adults with ASD during perceptual inference (**Fig. 4c**).

Furthermore, computational pupillometry analyses indicated heightened encoding of trialwise surprise in phasic noradrenergic responses in ASD (**Fig. 5b**). Thus, under the assumption that pupil size is an index of NA release from the locus coruleus[40], these results suggested increased phasic neuromodulatory signaling in ASD. NA
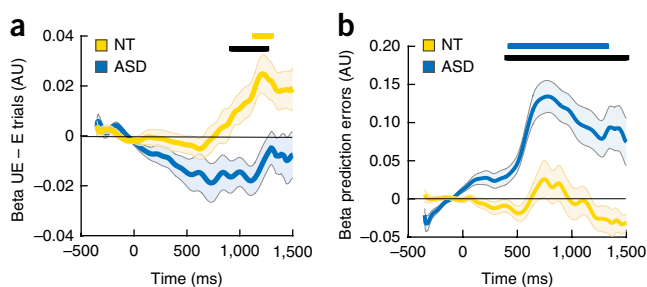
Figure 5 Pupillometry results. (**a**) Plot in which the solid yellow line shows the cluster of time points when the UE – E group difference was significantly positive in the NT participants, and the black solid line shows when the difference for NT participants was significantly greater than that of participants with ASD (2,000 permutations; familywise error (FEW) $\alpha = 0.05$, two tailed). (**b**) Plot in which the blue solid line indicates when participants with ASD showed a significant pupil response to precision-weighted prediction errors (ε_3) that was greater than zero, and the black solid line shows when this pupil response was significantly different in ASD compared with NT participants (2,000 permutations; FWE $\alpha = 0.05$, two tailed). NT, $n = 14$; ASD, $n = 11$. The x axis represents time since outcome. Shaded area indicates s.e.m. (across participants) for the beta at each time point. AU, arbitrary units.

is believed to change cortical gain in response to surprise, specifically salient events indicating that the global context has changed (cf. 'unexpected uncertainty'[15,25]). Here, our computational pupillometric analysis indicated a strong relationship between noradrenergic responsivity and precision-weighted prediction errors in participants with ASD. In agreement with our other model-based results (**Fig. 4b,c**), these findings again supported over-reactivity to environmental change in ASD, but in the context of physiological measures that index phasic neuromodulatory function. The NA system's signaling more high-level surprise in ASD may suggest atypical cortical gain during sensory processing, thus resulting in a state of disproportionate receptiveness to sensory inputs. Aberrant phasic NA (i.e., precision on prediction errors[4,6]) may alter the signal-to-noise ratio of cortical responses[41,42], broaden the tuning functions of sensory responses and subsequently improve discrimination behavior[43]. Thus, aberrant NA function may offer a neurobiological perspective on the profile of sensory-processing strengths and weakness experienced by people on the spectrum.

Importantly, these findings provide preliminary empirical evidence for neurobiologically informed Bayesian accounts of autism that emphasize aberrant representation of volatility and consequently inappropriate setting of gain (precision) on cortical responses (prediction errors)[4,6] under conditions of uncertainty. A recent pharmacological study using the HGF has indicated that noradrenaline antagonism selectively impairs volatility learning[13], in agreement with the raised pupillometric response to surprise about volatility reported here in the adults with ASD (**Fig. 5b**). We hypothesize that the noradrenergic locus coruleus and its coupling with the anterior cingulate cortex (ACC)[8,10] ratifies estimated volatility and that the downstream gain modulations act on the precision of cortical responses that are behaviorally relevant to the task at hand. Atypical social prediction error processing in the gyral surface of the ACC has recently been reported in autism[44], but whether differences in processing in the ACC region extend to nonsocial tasks with explicit computational models and manipulations of volatility remains to be seen. Carefully designed neuroimaging and neuropharmacology studies will be necessary to link these (presumed) noradrenergic effects, and the mathematical anatomy of uncertainty[11], to hierarchical processing in the brain[12].

Additionally, although we emphasize the role of noradrenaline here, we also acknowledge the likely importance of its direct precursor, dopamine, and the complementary relationship with acetylcholine and the signaling of expected uncertainty[15]. All three of these neuromodulators are likely candidates in the neurobiological mechanisms underlying responses to environmental change in ASD.

From a Bayesian perspective, the simplest way in which persistent overweighting of all sensory inputs (relative to prior expectations) might occur would be a generally larger outcome α, reflecting chronic and inflexible overweighting of recent, relative to past, sensory history. Such an explanation is implied by conservative interpretations of nonhierarchical Bayesian accounts of ASD[2] and predictive processing accounts that emphasize 'uniform' inflexibly high precision in sensory processing[3]. However, by logical extension, beyond a single ambiguous sensory event, all Bayesian accounts suggest that dynamic learning about structural regularities (i.e., the formation of priors) is probably compromised in ASD[2–7]. Under the aberrant precision account of ASD, problems with processing volatility under conditions of sensory ambiguity are hypothesized to underlie the difficulties faced by people on the spectrum[4–6]. For this reason, we designed a task to capture behavior under orthogonal manipulations of expectations and sensory noise and built a model equipped with the ability to reveal learning about volatility.

The recent proposal that nonhierarchical reinforcement-learning models can directly address predictive coding theories of ASD[3] is perhaps too simplistic, not least because predictive coding is largely regarded as a neural-process theory, and therefore behavioral or modeling results in the absence of a proxy for brain function can only speak to such an account but are not truly able to test it. Motivated by these claims, a relatively recent study has found no differences in learning-rate malleability in autistic children during a reward-learning task modeled with a delta learning rule[45]. Notably, however, that study found no group differences in simple behavior in that task. Here, we made a specific behavioral prediction on the basis of previous research (reduced surprise in ASD) and a specific computational prediction to explain this behavior (aberrant learning about volatility). We therefore designed a model sufficiently complex to address simultaneous hierarchical learning. Using Bayesian-model comparison, we found (**Supplementary Fig. 6**) that the simplest learning model (similar to that used previously in the context of reward learning[45]) performs most poorly in explaining participant behavior. Nonetheless, if it is the case that learning in the face of volatility is compromised in adults with autism (as reported here under conditions of sensory noise) but not children (as indicated previously in the context of reward learning[45]), this would be an important discovery. It will be crucial for future studies to use the same computational models and behavioral paradigms in adults and children to inform understanding of how autism affects cognition across the lifespan, especially because some features of the disorder can become more severe with age[46].

Conclusion

The surprise that you might experience after finding a pineapple in your sock drawer depends on the strength of your prior expectation to see socks. The results of this study suggest that adults with autism show a tendency to overestimate the volatility of the sensory environment, at the expense of learning to build stable expectations that lead to adaptive surprise. In other words, adults with autism may be mildly surprised by both the pineapple and the socks. Heightened encoding of prediction errors in pupil-size measures is consistent with neurobiologically focused Bayesian accounts of autism that emphasize neural-gain impairments due to aberrant neuromodulatory funcion[4–6]. The distinct but complementary results provided by the ground truth and

computational levels of analysis in our study indicate the utility of computational approaches in improving understanding of neurodevelopmental and psychiatric conditions with the aim of influencing clinical practice[47–49]. This study provides insights into the behavioral, algorithmic and physiological mechanisms that underlie learning about, and responses to, environmental change in ASD. Different patterns of learning may emerge when the environment is more or less changeable, when expectations are formed explicitly or when outcomes are not incidental but are instead tied to reward and/or social evaluation[10,50]. It will be important for future research to address these domains and to test volatility learning in larger cohorts and people with different intellectual abilities across the autism spectrum.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

AUTHOR CONTRIBUTIONS

R.P.L. conceived the study and collected and analyzed the data. R.P.L. and C.M. modeled the data. R.P.L., C.M. and G.R. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Kenny, L. *et al.* Which terms should be used to describe autism? Perspectives from the UK autism community. *Autism* **20**, 442–462 (2016).
2. Pellicano, E. & Burr, D. When the world becomes 'too real': a Bayesian explanation of autistic perception. *Trends Cogn. Sci.* **16**, 504–510 (2012).
3. Van de Cruys, S. *et al.* Precise minds in uncertain worlds: predictive coding in autism. *Psychol. Rev.* **121**, 649–675 (2014).
4. Lawson, R.P., Rees, G. & Friston, K.J. An aberrant precision account of autism. *Front. Hum. Neurosci.* **8**, 302 (2014).
5. Friston, K.J., Lawson, R. & Frith, C.D. On hyperpriors and hypopriors: comment on Pellicano and Burr. *Trends Cogn. Sci.* **17**, 1 (2013).
6. Lawson, R.P., Friston, K.J. & Rees, G. A more precise look at context in autism. *Proc. Natl. Acad. Sci. USA* **112**, E5226 (2015).
7. Palmer, C.J., Lawson, R.P. & Hohwy, J. Bayesian approaches to autism: towards volatility, action, and behavior. *Psychol. Bull.* **143**, 521–542 (2017).
8. Behrens, T.E., Woolrich, M.W., Walton, M.E. & Rushworth, M.F. Learning the value of information in an uncertain world. *Nat. Neurosci.* **10**, 1214–1221 (2007).
9. Browning, M., Behrens, T.E., Jocham, G., O'Reilly, J.X. & Bishop, S.J. Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nat. Neurosci.* **18**, 590–596 (2015).
10. Behrens, T.E.J., Hunt, L.T., Woolrich, M.W. & Rushworth, M.F.S. Associative learning of social value. *Nature* **456**, 245–249 (2008).
11. Mathys, C.D. *et al.* Uncertainty in perception and the Hierarchical Gaussian Filter. *Front. Hum. Neurosci.* **8**, 825 (2014).
12. Iglesias, S. *et al.* Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. *Neuron* **80**, 519–530 (2013).
13. Marshall, L. *et al.* Pharmacological fingerprints of contextual uncertainty. *PLoS Biol.* **14**, e1002575 (2016).
14. de Berker, A.O. *et al.* Computations of uncertainty mediate acute stress responses in humans. *Nat. Commun.* **7**, 10996 (2016).
15. Yu, A. & Dayan, P. Expected and unexpected uncertainty: ACh and NE in the neocortex. *Adv. Neural Inf. Process. Syst.* **15**, 173–180 (2003).
16. Berridge, C.W. & Waterhouse, B.D. The locus coeruleus-noradrenergic system: modulation of behavioral state and state-dependent cognitive processes. *Brain Res. Brain Res. Rev.* **42**, 33–84 (2003).
17. Hasselmo, M.E. & McGaughy, J. High acetylcholine levels set circuit dynamics for attention and encoding and low acetylcholine levels set dynamics for consolidation. *Prog. Brain Res.* **145**, 207–231 (2004).
18. Kobayashi, M. *et al.* Selective suppression of horizontal propagation in rat visual cortex by norepinephrine. *Eur. J. Neurosci.* **12**, 264–272 (2000).
19. Shepard, K.N., Liles, L.C., Weinshenker, D. & Liu, R.C. Norepinephrine is necessary for experience-dependent plasticity in the developing mouse auditory cortex. *J. Neurosci.* **35**, 2432–2437 (2015).
20. Lawson, R.P., Aylward, J., White, S. & Rees, G. A striking reduction of simple loudness adaptation in autism. *Sci. Rep.* **5**, 16157 (2015).
21. Ewbank, M.P. *et al.* Repetition suppression in ventral visual cortex is diminished as a function of increasing autistic traits. *Cereb. Cortex* **25**, 3381–3393 (2015).
22. Gomot, M. *et al.* Candidate electrophysiological endophenotypes of hyper-reactivity to change in autism. *J. Autism Dev. Disord.* **41**, 705–714 (2011).
23. Kleinhans, N.M. *et al.* Reduced neural habituation in the amygdala and social impairments in autism spectrum disorders. *Am. J. Psychiatry* **166**, 467–475 (2009).
24. den Ouden, H.E., Daunizeau, J., Roiser, J., Friston, K.J. & Stephan, K.E. Striatal prediction error modulates cortical coupling. *J. Neurosci.* **30**, 3210–3219 (2010).
25. Yu, A.J. Change is in the eye of the beholder. *Nat. Neurosci.* **15**, 933–935 (2012).
26. 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).
27. South, M., Newton, T. & Chamberlain, P.D. Delayed reversal learning and association with repetitive behavior in autism spectrum disorders. *Autism Res.* **5**, 398–406 (2012).
28. Nemeth, D. *et al.* Learning in autism: implicitly superb. *PLoS One* **5**, e11731 (2010).
29. Brown, J., Aczel, B., Jiménez, L., Kaufman, S.B. & Grant, K.P. Intact implicit learning in autism spectrum conditions. *Q. J. Exp. Psychol. (Hove)* **63**, 1789–1812 (2010).
30. Ratcliff, R., Smith, P.L., Brown, S.D. & McKoon, G. Diffusion decision model: current issues and history. *Trends Cogn. Sci.* **20**, 260–281 (2016).
31. Wiecki, T.V., Sofer, I. & Frank, M.J. HDDM: hierarchical Bayesian estimation of the drift-diffusion model in Python. *Front. Neuroinform.* **7**, 14 (2013).
32. 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).
33. Mathys, C., Daunizeau, J., Friston, K.J. & Stephan, K.E. A bayesian foundation for individual learning under uncertainty. *Front. Hum. Neurosci.* **5**, 39 (2011).
34. Lam, K.S., Aman, M.G. & Arnold, L.E. Neurochemical correlates of autistic disorder: a review of the literature. *Res. Dev. Disabil.* **27**, 254–289 (2006).
35. Daluwatte, C. *et al.* Atypical pupillary light reflex and heart rate variability in children with autism spectrum disorder. *J. Autism Dev. Disord.* **43**, 1910–1925 (2013).
36. Courchesne, E., Kilman, B.A., Galambos, R. & Lincoln, A.J. Autism: processing of novel auditory information assessed by event-related brain potentials. *Electroencephalogr. Clin. Neurophysiol.* **59**, 238–248 (1984).
37. Jeste, S.S. *et al.* Electrophysiological evidence of heterogeneity in visual statistical learning in young children with ASD. *Dev. Sci.* **18**, 90–105 (2015).
38. Falck-Ytter, T. & von Hofsten, C. How special is social looking in ASD: a review. *Prog. Brain Res.* **189**, 209–222 (2011).
39. Aston-Jones, G. & Cohen, J.D.A.N. An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.* **28**, 403–450 (2005).
40. Costa, V.D. & Rudebeck, P.H. More than meets the eye: the relationship between pupil size and locus coeruleus activity. *Neuron* **89**, 8–10 (2016).
41. Hasselmo, M.E., Linster, C., Patil, M., Ma, D. & Cekic, M. Noradrenergic suppression of synaptic transmission may influence cortical signal-to-noise ratio. *J. Neurophysiol.* **77**, 3326–3339 (1997).
42. Hirata, A., Aguilar, J. & Castro-Alamancos, M.A. Noradrenergic activation amplifies bottom-up and top-down signal-to-noise ratios in sensory thalamus. *J. Neurosci.* **26**, 4426–4436 (2006).
43. Martins, A.R.O. & Froemke, R.C. Coordinated forms of noradrenergic plasticity in the locus coeruleus and primary auditory cortex. *Nat. Neurosci.* **18**, 1483–1492 (2015).
44. Balsters, J.H. *et al.* Disrupted prediction errors index social deficits in autism spectrum disorder. *Brain* **140**, 235–246 (2017).
45. 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.* http://dx.doi.org/10.1111/desc.12435 (2016).
46. Happé, F.G. *et al.* Demographic and cognitive profile of individuals seeking a diagnosis of autism spectrum disorder in adulthood. *J. Autism Dev. Disord.* **46**, 3469–3480 (2016).
47. Haker, H., Schneebeli, M. & Stephan, K.E. Can Bayesian theories of autism spectrum disorder help improve clinical practice? *Front. Psychiatry* **7**, 107 (2016).
48. Corlett, P.R. & Fletcher, P.C. Computational psychiatry: a Rosetta Stone linking the brain to mental illness. *Lancet Psychiatry* **1**, 399–402 (2014).
49. Teufel, C. & Fletcher, P.C. The promises and pitfalls of applying computational models to neurological and psychiatric disorders. *Brain* **139**, 2600–2608 (2016).
50. Sevgi, M., Diaconescu, A.O., Tittgemeyer, M. & Schilbach, L. Social Bayes: using Bayesian modeling to study autistic trait–related differences in social cognition. *Biol. Psychiatry* **80**, 112–119 (2016).

ONLINE METHODS

Participants. 29 adults with ASD and 26 NT volunteers came to the UCL Institute of Cognitive Neuroscience during a testing day involving different researchers. Two adults with ASD did not complete this test, owing to time constraints or an inability to tolerate the sounds and/or to focus adequately on the test. After data examination, participants with more than 20% overall errors or mean reaction times (RTs) >2 s.d. from their respective group-mean RT were excluded from subsequent analysis to ensure the validity of the Bayesian modeling, thus yielding 24 participants in the ASD group (18 males and 6 females; mean age, 35.5 years; age range, 20–61 years) and 25 participants in the NT group (16 males and 9 females; mean age, 36 years; age range, 19–62 years). The ASD and NT groups were matched according to age ($t(47) = 0.54$, $P = 0.87$).

Participants with ASD had previously been diagnosed by an independent clinician, according to the DSM-IV[51] or ICD-10 criteria[52] (19 with Asperger syndrome; 3 with autistic disorder; 1 with high-functioning autism; and 1 with atypical autism). The Wechsler Adult Intelligence Scale (WAIS, third edition, UK) had previously been administered to assess IQ[53], and participants were matched on the basis of full-scale scores (ASD mean, 117; ASD range, 80–142; NT mean, 120; NT range, 99–145; $t(47) = -0.93$, $P = 0.36$). The Autism Diagnostic Observation Schedule (second edition)[54] assessment was completed by a qualified administrator to assess symptom severity in the participants with ASD. The mean ADOS total score was 9.9 (range 4–19). The mean scores for the communication and reciprocal social interaction subscores were 3.3 (range, 0–7) and 6.6 (range, 4–12), respectively.

An additional 57 NT volunteers were studied as part of a replication of our key behavioral results (25 males and 32 females; mean age, 27.1 years; age range, 19–50 years) and additionally completed the Autism Spectrum Quotient (AQ) questionnaire, a 50-item self-reported measure of autism traits[32,55]. The mean AQ score was 18.43 (median, 17; range, 5–45). All participants had normal or corrected-to-normal vision and provided written informed consent. We performed a median split on the data, such that participants were divided into high-AQ ($n = 26$) and low-AQ ($n = 31$) groups. The AQ score was significantly higher in the high-AQ group (mean, 27; s.d., 6.4; range, 18–45) relative to the low-AQ group (mean, 11.5; s.d., 3.4; range, 5–17; $t(55) = 11.28$, $P < 0.001$). The distribution of scores in the low-AQ group was almost exclusively below the mean range of neurotypical scores reported in a recent meta-analysis of 73 studies administering the AQ[56]. Importantly there was considerable overlap between the scores in the high-AQ group and the range reported, on average, for those with a diagnosis of ASD[56], even though these participants do not present with any clinical need.

No randomization was used to assign subjects or conditions. All participants provided written informed consent and were compensated financially for their time and travel expenses. The study was approved by the UCL Graduate School Ethics Committee (4357/001).

Stimuli. Auditory cues were either 330-Hz or 660-Hz pure tones generated in MATLAB R2012b (Mathworks) and presented with the Cogent toolbox (http://www.vislab.ucl.ac.uk/cogent_graphics.php/), via Sennheiser HD 201 headphones. Outcome images were either faces or houses. These stimuli were grayscale and comprised six different face identities (three male and female) or six different images of houses, masked by an ellipse and luminance matched with the SHINE toolbox[57]. Either medium or high Gaussian noise, with a mean of zero and a variance of 0.05 and 0.1, respectively, was added to outcome images by using the image processing toolbox (MATLAB R2012b). Example stimuli can be seen in **Figure 1**. The face in this image shows the likeness of author R.P.L. The actual stimuli used in the experiment were taken from the Nim Stim Face Identity Set (http://www.macbrain.org/resources.htm).

Procedure. Participants sat on a chair with their heads in a chin rest at a viewing distance of 80 cm. An example trial can be seen in **Figure 1**. Each trial began with the 300-ms presentation of a pure tone that was either high or low in pitch. After 200 ms, either a face or a house image was presented for 150 ms, to prevent saccades. The participants were told simply to respond to the image and indicate whether it was a face or a house (via left/right button press), and to be 'as fast and accurate as possible, trying to respond on every trial'. A variable response time of 1,500–1,800 ms followed the image, such that trials lasted 1,950–2,250 ms. Participants were instructed that the tone preceding each image was probabilistically associated with the likelihood of seeing a face or house and that these

probabilities would change over time. The probabilistic associations between the tones and the outcomes were highly predictive ($P = 0.84$), weakly predictive ($P = 0.16$) or nonpredictive ($P = 0.5$), and changed pseudorandomly across trials in blocks of 12, 36 or 72 trials (**Fig. 1**). All participants completed 456 trials over eight miniblocks with optional periods of rest between.

To ensure that participants' responses were not biased by learned expectations about the relative frequencies of the visual stimuli, the task was designed such that the marginal probabilities of faces and houses were identical at any point in time (**Fig. 1**), and each block contained equal numbers of randomly intermixed high- and low-tone trials. As with designs of previous studies[12,24], our design ensured that the a priori probability of a face (or house) occurring was always 50% in any given trial, before the tone was presented. Thus, any expectations about the visual stimulus could depend on only the preceding tone. Additionally, and uniquely to this study, equal numbers of high-, medium- and no-noise stimuli appeared in each of 12, 36 or 72 blocks of trials and across each cue type.

Data collection and analysis were not performed blind to the conditions of the experiment.

Pupillometry. To ensure fixation and to measure neuromodulatory responses, gaze direction and pupil size were measured with an infrared eye tracker (Cambridge Research Systems) tracking the left eye at 200 Hz. Calibration of the eye tracker was unsuccessful in all participants wearing glasses, and the eye tracker suffered a fatal technical failure before testing was completed; therefore, eye-tracking data are available for only 14 NT subjects and 11 subjects with ASD.

Hierarchical Gaussian filter. In the version of the HGF used here (introduced in ref. 33), learning occurs simultaneously on three coupled levels of an uncertainty hierarchy. The first level of the HGF (x_1) constitutes the outcome of any given trial (for example, face or house), the second level (x_2) represents the probabilistic associations between the tones and the outcomes (for example, the probability of seeing a house given that a high tone has just been heard), and the third level (x_3) quantifies the volatility of the probabilities (for example, the changeability of the environment). In each trial, the model provides an estimate for each level, before the outcome is seen, and the estimate updated accordingly. Predictions at each level are represented by a Gaussian distribution, described by its mean $\hat{\mu}_i$, and variance $\hat{\sigma}_i$. The variance $\hat{\sigma}_i$ represents the uncertainty of the estimate at each level. Updates of beliefs at each level occur via prediction errors that propagate upward and are precision weighted according to the ratio of the uncertainty of the level that generated them to the uncertainty of the level being updated. The manipulation of perceptual noise (for example, no, medium or high noise) is captured trial by trial as a fixed parameter representing the variance of the noise on the inputs.

For each participant, the perceptual-model parameters ω_2 and ω_3, the learning rates α_2 and α_3, and the response model parameters ($\beta_{0,...,4}$) were estimated from the trialwise log RT measures by using variational Bayes, as implemented in the HGF toolbox (http://www.translationalneuromodeling.org/tapas/). The ω values are the tonic log volatilities at their respective levels, according to the generative model

$$x_1^{(t)} \sim \text{Bernoulli}\left(s\left(x_2^{(t)}\right)\right)$$

$$x_2^{(t)} \sim \mathcal{N}\left(x_2^{(t-1)}, \exp\left(x_3^{(t)} + \omega_2\right)\right)$$

$$x_3^{(t)} \sim \mathcal{N}\left(x_3^{(t-1)}, \exp\left(\omega_3\right)\right)$$

with $s\left(x_2^{(t)}\right) := 1/\left(1 + \exp\left(-x_2^{(t)}\right)\right)$. Thus, the basic step size of the random walks in x_2 and x_3 is determined without taking into account phasic modulation by higher levels of the hierarchy. The learning rate α_2 represents, trial by trial, the size of the update in μ_2 (i.e., the mean of the belief on x_2) relative to the size of the prediction error δ_1, as expressed in terms of the update in predicted outcome probabilities $\hat{\mu}_i$:

$$\alpha_2^{(t)} := \frac{\hat{\mu}_1^{(t)} - \hat{\mu}_1^{(t-1)}}{\delta_1^{(t)}}$$

where $\widehat{\mu}_1^{(t)} := s\left(\mu_2^{(t-1)}\right)$. The learning rate α_3 is the equivalent quantity with respect to the size of the update in μ_3:

$$\alpha_3^{(t)} := \frac{\mu_3^{(t)} - \mu_3^{(t-1)}}{\delta_2^{(t)}}$$

Furthermore, α_3 is proportional to the precision weight on the prediction error $\varepsilon_3^{(t)}$:

$$\alpha_3^{(t)} \propto \frac{\widehat{\pi}_2^{(t)}}{\pi_3^{(t)}}$$

where $\pi_3^{(t)}$ is the posterior precision (inverse variance) at the third level, and $\widehat{\pi}_2^{(t)}$ is the precision (inverse variance) of the prediction at the second level. Accordingly, $\varepsilon_3^{(t)}$ is the precision-weighted prediction error at the second level, which serves to update the estimate of log volatility:

$$\varepsilon_3^{(t)} := \frac{\widehat{\pi}_2^{(t)}}{\pi_3^{(t)}} \delta_2^{(t)}$$

More details can be found in the supplementary material of ref. 6.

The β values are the coefficients of the response model, which describes how beliefs (i.e., the probability distributions on x_i, as represented by their sufficient statistics μ_i and σ_i) are translated into log reaction times. This is a straightforward linear model:

$$\log RT^{(t)}$$
$$\sim \mathcal{N}\left(\beta_0 + \beta_1 \times \text{surprise}^{(t)} + \beta_2 \times \text{unc1}^{(t)} + \beta_3 \times \text{unc2}^{(t)} + \beta_4 \times \text{volatility}^{(t)}, \zeta\right)$$

with the independent variables defined as follows:

$$\text{surprise}^{(t)} := \begin{cases} -\log_2\left(\widehat{\mu}_1^{(t)}\right) & \text{if } u^{(t)} = 1 \\ -\log_2\left(1 - \widehat{\mu}_1^{(t)}\right) & \text{if } u^{(t)} = 0 \end{cases}$$

$$\text{unc1}^{(t)} := \widehat{\sigma}_1^{(t)}$$

$$\text{unc2}^{(t)} := s\left(\mu_2^{(t)}\right)\left(1 - s\left(\mu_2^{(t)}\right)\right)\sigma_2^{(t)}$$

$$\text{volatility}^{(t)} := s\left(\mu_2^{(t)}\right)\left(1 - s\left(\mu_2^{(t)}\right)\right)\exp\left(\mu_3^{(t)}\right)$$

Here, $u^{(t)}$ is the outcome; $u^{(t)} = 1$ when the high-tone cue is followed by a face, or the low-tone cue is followed by a house, whereas $u^{(t)} = 0$ in the converse cases. Because $\widehat{\mu}_1^{(t)}$ is the predicted probability of $u^{(t)} = 1$ (and $1 - \widehat{\mu}_1^{(t)}$ of $u^{(t)} = 0$), the first independent variable is the Shannon surprise associated with the outcome. Uncertainty at the outcome level (i.e., the first) is the variance

$$\widehat{\sigma}_1^{(t)} = \widehat{\mu}_1^{(t)}\left(1 - \widehat{\mu}_1^{(t)}\right)$$

of the Bernoulli distribution over predicted outcomes. This parameter is the irreducible uncertainty associated with any kind of probabilistic prediction, which is referred to as risk in the economics literature. Uncertainty at the second level is the posterior variance σ_2 of the belief on x_2, expressed at the outcome level (hence the multiplication with the derivative of s taken at the current mean μ_2 of the belief on x_2; details on this transformation to the first level have been described in the supplementary material of ref. 12. This parameter is the informational uncertainty, so called because it quantifies the lack of information about the

quantity (here x_2) governing outcome probabilities. Volatility is the exponential of the phasic log volatility μ_3, which is also expressed at the outcome level.

The choice of these models was hypothesis driven. The reason for choosing the HGF as the learning model was twofold. First, because it reflects the hierarchical nature of changing environments in that it allows for volatility that is itself volatile, this model allowed us to test the hypothesis that participants with ASD differ from NT participants in how they address a hierarchy of uncertainties and specifically address learning about volatility. The response model was chosen because the log RTs followed an approximately Gaussian distribution, and a linear model allowed for the straightforward identification of the effects of all hypothesized modulating factors.

There were several reasons that we chose to fit reaction time over trialwise errors. First, reaction times are a sensitive behavioral-response measure that can have a range of values across trials, from fast to slow, and that empirically have been shown to vary with the uncertainty of participant responses in both detection and discrimination experiments[58]. Second, reaction times have previously been used in the application of Bayesian learning models to behavioral tasks very similar to ours, so modeling RT here increased the comparability across studies[24,59]. Third, error rates were very low in this study (~3% overall), and any logistic model attempting to explain such a small incidence of states coded as 1 (relative to 0) would require more trials than were performed in this study (increasing as a function of the explanatory variables in the model[60]). Fourth (and most pragmatically), some participants did not make any errors at all, so modeling RT maximized the number of participants included in the analysis. Finally, the group × probability interaction for percentage errors was not significant in our high- and low-AQ replication (**Supplementary Fig. 4c,d**), and so in modeling RT, we modeled the effect most comparable across both experiments in this work.

Sample size. In our NT participants, we sought a conceptual replication of the study by Den Ouden *et al.*[24], albeit with a modified design. We calculated a minimum sample size a priori on the basis of the low-probability minus high-probability RT difference previously reported (32 ms) and an assumed variance (actual s.d. not reported) of the same. This analysis indicated that we would need a minimum of 14 participants to achieve 95% power to detect a similar ($\alpha = 0.05$; two tailed) effect in the NT group. Given that initial effect sizes are often inflated[61] and that we sought power to detect a difference between two groups, we doubled this estimate and aimed to test ~28 participants in each group, with some attrition expected.

Because there is no prior precedent for detecting between-group differences in this specific task, we additionally assessed the required sample size to detect a medium effect size for a between-subject ANOVA with three levels and a between-subject factor of group. The results indicated that a total sample size of 48 participants would be necessary to have at least 90% power to detect an F-test effect size of 0.25.

For the pupil-size regression, in which it was not possible to calculate power a priori, the sample sizes and effect sizes (β values) reported for this particular analysis were in line with those reported in previous studies using the same methods[9]. *Post hoc* power calculations indicated that with 11 participants with ASD included in the actual analysis, we had 86% power to detect the mean positive β (slope = 0.72) that we observed in these participants ($\alpha = 0.05$; two tailed).

Statistics. *Behavioral data.* All statistical analyses of behavioral data were performed in MATLAB (Mathworks) and PASW Statistics 22 (SPSS/IBM). For the analysis of RTs, responses that were too fast or too slow (<100 or >1,000 ms) were excluded, and, including missing responses, there was no significant difference between the groups in the overall percentage of missing data (1.9% ASD, 2.3% NT, $t(47) = 0.45$, $P = 0.65$). To maximize the trial numbers per condition, we collapsed across face/house trials and, for correct trials only, subjected RTs to a mixed ANOVA with within-subject factors of expectedness (UE, N and E) and stimulus noise (high, medium and no noise), and a between-subject factor of group. We also quantified a behavioral measure of surprise, defined as the difference in RT between UE and E outcomes, on the basis of the ground truth, and compared this measure between the groups by using independent-sample t tests. An equivalent analysis was conducted for error rates and log transforms of both these measures. Percentage errors were calculated for each condition separately. Data distributions were assumed to be normal, but this assumption was not formally tested. When assumptions of heterogeneity of covariance were violated, the number of degrees of freedom was corrected with the Greenhouse–Geisser approach.

Eye-tracking data. All statistical analyses of eye-tracking data were performed in MATLAB (Mathworks) Only trials in which 80% or more of the samples were successfully tracked were included in the analysis. There was no significant difference in the mean number of included trials between the groups (mean good trials ASD = 298; NT = 261; $t(23) = 0.803$, $P = 0.43$). For pupil data, blinks were treated with linear interpolation, and the resulting pupil traces were low-pass-filtered and smoothed according to the conventions outlined in ref. 62. To explore phasic pupil responses for correct trials, traces were baseline corrected to the average response during the 500 ms preceding the outcome image. Tonic pupil responses were determined as the average of the *z*-scored pupil measurement across all trials. *z*-scoring accounts for individual differences in baseline pupil size and has previously been used in the literature[63,64]. The MAD from fixation (in degrees of visual angle) across groups and conditions was used to assess fixation compliance in each trial[65].

Regression analyses were conducted to examine the effects of surprise on the basis of the ground truth and volatility surprise (ε_3, trialwise precision-weighted prediction errors) on pupil dilation after outcome presentation. A similar approach has been used in recent studies examining the relationship between pupil dilation and computational-model parameters that vary across trials[9]. The postoutcome period for each trial was sampled with 370 5-ms bins. Regression analyses were conducted for each individual time bin, with HGF estimates of precision-weighted prediction errors (ε_3) and the ground-truth contrast of unexpected (1) minus expected (−1) included as regressors of interest, and the trial type (0 = face, 1 = house), fixation compliance (MAD) and RT for each trial entered as control regressors. The resultant time series of β weights (in other words, multiple regression conducted at every time point) provided estimates of the effects of ground-truth surprise and volatility surprise on the basis of pupil dilation across all trials.

At the group level, we then conducted *t* tests for the positive or negative effects of the regressors of interest, and the independent-samples difference between groups, corrected for multiple comparisons with a cluster-based permutation approach at 2,000 permutations (FWE α = 0.05, two tailed)[66]. This procedure allowed us to assess when our surprise metrics were significantly encoded in the pupil time series. This approach protected against false positives across correlated measurements (i.e., it maximized temporal sensitivity).

Learning-rate data. To test the hypothesis that individuals with ASD have difficulties in flexibly updating their rate of learning (precision weighting) in response to environmental change, we examined the change (Δ) in α_2 (probability) and α_3 (environment) during switching from stable to volatile periods of the task. We used the dynamic α trajectories estimated on the basis of all trials, but specifically interrogated a period of 72 trials (green in **Fig. 1**) in which the probabilistic association between tones and outcomes remained fixed and was followed by a period of 72 trials (violet in **Fig. 1**) in which the outcome probabilities switched three times. We compared the change in average α_2 and α_3 between these two periods, across the groups. Previous studies have examined learning about how reward probability changes in response to volatility in NT volunteers[8,10] and also in aversive environments[9]. In these studies, the participant responses were fit with a simple delta learning rule (i.e., Rescorla–Wagner[67]) separately in volatile and stable task phases, thereby annulling the elegance of the generative-model approach by imposing knowledge of the task structure. In contrast, we fit subject RTs across all 456 trials by using the HGF, and the two learning rates (α_2 and α_3) dynamically varied as a function of each participant's inferred beliefs about cue–outcome informativeness and changes in these associations over time. Whereas simpler models approximate participant outcome probability estimates, assuming that the participants are 'ideal' Bayesian observers, the HGF addresses what kind of Bayesian observer each participant actually is, thus making this method a more sensitive means of capturing individual differences in learning about uncertainty (additional information can be found in the 'HGF model validation' section; comparisons between the HGF and simpler reinforcement learning models can be found in **Fig. 3c**).

Bayesian-model selection. To disambiguate alternative explanations (models) for the participants' behavior, we used BMS. BMS evaluates the relative plausibility of competing models in terms of their log evidences, quantifying the tradeoff between the accuracy (fit) and complexity of a model. Here, we used a recently updated random-effect BMS method to account for potential interindividual variability in our sample, quantifying the protected posterior probabilities of four competing models[68]. The protected exceedance probabilities quantify the probability that any one model is more frequent than the others and also accounts for the possibility that the observed variability in (log) model evidences could be due to chance[68].

Regression analyses. To examine the relationship between the primary behavioral measure of surprise (UE – E RT) and the severity of autism symptoms, we conducted a multiple linear regression with ADOS-2 scores for communication and reciprocal social interaction, and IQ as predictors. A secondary regression model was also conducted in which a sensory sensitivity score (as measured on the basis of the adult sensory questionnaire[69]) was also included as a predictor. Sensory scores were available for only 21 of 24 participants with ASD; therefore, this analysis was conducted with a reduced sample size. In response to a reviewer request, we also conducted a third regression to predict UE – E RT that included baseline RT as an additional predictor. Because both communication scores ($r = -0.421$, $P = 0.04$) and mean RT ($r = -0.341$, $P = 0.017$) correlated with the UE – E RT difference in the participants with ASD, we created centered versions of these variables and their interaction effect in the regression model.

To assess the validity of the HGF-model parameters in predicting group status (ASD = 1, NT = 0) we conducted binary logistic regression (method = enter) in SPSS. The predictor variables in this analysis were the eight free parameters estimated by the HGF, namely the five response-model β values ($\beta_{0...4}$ plus decision noise (ζ) and the two omega parameters from the perceptual model (ω_2 and ω_3). Additionally, we recreated this analysis in R and used the cv.glm function in the boot package to perform leave-one-out cross-validation.

A **Life Sciences Reporting Summary** for this paper is available.

Code availability. We used the HGF toolbox (http://www.translationalneuromodeling.org/hgf-toolbox-v3-0/) for modeling learning. For pupillometry analysis, we used modified versions of the freely available preprocessing code available at http://www.tqmp.org/RegularArticles/vol10-2/p179/index.html, and analysis was conducted with code from the Mass Univariate Toolbox (http://openwetware.org/wiki/Mass_Univariate_ERP_Toolbox/). Code to control the low-level image properties of the stimuli used in this experiment is available at http://www.mapageweb.umontreal.ca/gosselif/SHINE/.

Data availability. The data that support the findings of this study are available upon reasonable request from the corresponding author, in accordance with local ethics rules.

51. American Psychiatric Association. *Diagnostic and Statistical Manual–Text Revision (DSM-IV-TRim, 2000)* (American Psychiatric Association, 2000).
52. World Health Organization. *International classification of diseases (ICD-10)* (World Health Organization, 2012).
53. Wechsler, D. & Hsiao-pin, C. *WASI-II: Wechsler Abbreviated Scale of Intelligence* (Pearson, 2011).
54. Lord, C. *et al.* The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J. Autism Dev. Disord.* **30**, 205–223 (2000).
55. Woodbury-Smith, M.R., Robinson, J., Wheelwright, S. & Baron-Cohen, S. Screening adults for Asperger Syndrome using the AQ: a preliminary study of its diagnostic validity in clinical practice. *J. Autism Dev. Disord.* **35**, 331–335 (2005).
56. Ruzich, E. *et al.* Measuring autistic traits in the general population: a systematic review of the Autism-Spectrum Quotient (AQ) in a nonclinical population sample of 6,900 typical adult males and females. *Mol. Autism* **6**, 2 (2015).
57. Willenbockel, V. *et al.* Controlling low-level image properties: the SHINE toolbox. *Behav. Res. Methods* **42**, 671–684 (2010).
58. Bonnet, C., Fauquet Ars, J. & Estaún Ferrer, S. Reaction times as a measure of uncertainty. *Psicothema* **20**, 43–48 (2008).
59. Vossel, S. *et al.* Spatial attention, precision, and Bayesian inference: a study of saccadic response speed. *Cereb. Cortex* **24**, 1436–1450 (2014).
60. Hosmer, D.W. Jr. & Lemeshow, S. *Applied Logistic Regression* (Wiley, 2004).
61. Ioannidis, J.P. Why most discovered true associations are inflated. *Epidemiology* **19**, 640–648 (2008).
62. Jackson, I. & Sirois, S. Infant cognition: going full factorial with pupil dilation. *Dev. Sci.* **12**, 670–679 (2009).
63. Kang, O. & Wheatley, T. Pupil dilation patterns reflect the contents of consciousness. *Conscious. Cogn.* **35**, 128–135 (2015).
64. Knapen, T. *et al.* Cognitive and ocular factors jointly determine pupil responses under equiluminance. *PLoS One* **11**, e0155574 (2016).
65. Schwarzkopf, D.S., Anderson, E.J., de Haas, B., White, S.J. & Rees, G. Larger extrastriate population receptive fields in autism spectrum disorders. *J. Neurosci.* **34**, 2713–2724 (2014).
66. Groppe, D.M., Urbach, T.P. & Kutas, M. Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. *Psychophysiology* **48**, 1711–1725 (2011).
67. Rescorla, R.A. & Wagner, A.R. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement. in *Classical Conditioning II: Current Research and Theory* 64–99 (Appleton-Century-Crofts, 1972).
68. Rigoux, L., Stephan, K.E., Friston, K.J. & Daunizeau, J. Bayesian model selection for group studies - revisited. *Neuroimage* **84**, 971–985 (2014).
69. Kinnealey, M., Oliver, B. & Wilbarger, P. A phenomenological study of sensory defensiveness in adults. *Am. J. Occup. Ther.* **49**, 444–451 (1995).

Corrigendum: Adults with autism overestimate the volatility of the sensory environment

Rebecca P Lawson, Christoph Mathys & Geraint Rees

In the version of this article initially published online, the first sentence of the Online Methods referred to "29 adults with ASD and 26 healthy NT volunteers." To avoid any implication that those with ASD are not healthy, this has been changed to "29 adults with ASD and 26 NT volunteers." Similarly, in the first paragraph of the "Learning-rate data" section, "healthy volunteers" has been changed to "NT volunteers." In the Life Sciences Reporting Summary, "healthy" has been changed to "neurotypical" in the first sentences of items 3 and 12. In the supplementary information originally posted online, the legend to Supplementary Figure 4 read "non-clinical healthy volunteers." This has been changed to "non-clinical volunteers." The errors have been corrected in the PDF and HTML versions of this article.

natureresearch

Corresponding author(s):   Rebecca P. Lawson

☐ Initial submission   ☐ Revised version   ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

▶ Experimental design

1. Sample size

Describe how sample size was determined.

In our NT participants we sought a conceptual replication of Den Ouden et al. 24, albeit with a modified design. We calculated a minimum sample size a priori on the basis of the low probability minus high probability RT difference that they report (32 ms) and an assumed variance (actual SD not reported) of the same. This analysis indicated that we would need a minimum of 14 participants to achieve 95% power to detect a similar ($\alpha = 0.05$; 2-tailed) effect in the NT group. Given that initial effect sizes are often inflated 61 and that we sought power to detect a difference between two groups, we doubled this estimate and aimed to test ~28 participants in each group with some attrition expected.
As there is no prior precedent for detecting between-groups differences using this specific task, we additionally assessed the required sample size to detect a medium effect size for a between-subjects ANOVA with three levels and a between-subjects factor of group. This indicated that a total sample size of 48 participants would be necessary to have at least 90% power to detect an F-test effect size of 0.25.
For the pupil size regression, where it was not possible to calculate power a priori, the sample sizes and effect sizes (β's) reported for this particular analysis are in line with previous studies employing the same methods9. Post-hoc power calculations indicate that with 11 ASD participants included in the actual analysis, we had 86% power to detect the mean positive β (slope=0.72) that we observed in these participants ($\alpha = 0.05$; 2-tailed).

2. Data exclusions

Describe any data exclusions.

Two adults with ASD did not complete this test owing to time constraints or an inability to tolerate the sounds and/or focus adequately on the test. Following data examination, participants with more than 20% overall errors or mean reaction times (RT's) > 2 standard deviations from their respective group mean RT were excluded from subsequent analysis to ensure the validity of the Bayesian modeling

3. Replication

Describe whether the experimental findings were reliably reproduced.

We replicated our primary behavioral results in a separate group of neurotypical volunteers (n=57) characterized according to high and low autistic traits. The results of this replication are presented in Supplemental Figure S4.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

No randomization was used to assign subjects or conditions.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Data collection and analysis were not performed blind to the conditions of the experiment.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	A statement indicating how many times each experiment was replicated
☐	☒	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒	The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒	A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

▶ Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

We used the HGF toolbox (http://www.translationalneuromodeling.org/hgf-toolbox-v3-0/) for modelling learning. For pupillometry analysis we used modified versions of the freely available pre-processing code available here (http://www.tqmp.org/RegularArticles/vol10-2/p179/index.html) and analysis was conducted using code after the Mass Univariate Toolbox (http://openwetware.org/wiki/Mass_Univariate_ERP_Toolbox). Code to control the low level image properties of the stimuli used in this experiment is available here: http://www.mapageweb.umontreal.ca/gosselif/SHINE/.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

▶ Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

No unique materials were used.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

No antibodies were used.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

No eukaryotic cell lines were used.

b. Describe the method of cell line authentication used.

No eukaryotic cell lines were used.

c. Report whether the cell lines were tested for mycoplasma contamination.

No eukaryotic cell lines were used.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

No eukaryotic cell lines were used.

▶ Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

> No research animals were used.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

> 29 adults with autism spectrum disorder (ASD) and 26 neurotypical volunteers (NTs) came to the UCL Institute of Cognitive Neuroscience as part of a testing day involving different researchers. Two adults with ASD did not complete this test owing to time constraints or an inability to tolerate the sounds and/or focus adequately on the test. Following data examination, participants with more than 20% overall errors or mean reaction times (RT's) > 2 standard deviations from their respective group mean RT were excluded from subsequent analysis to ensure the validity of the Bayesian modelling. This left 24 participants in the ASD group (18 males; mean age: 35.5, age range: 20-61) and 25 in the NT group (16 males; mean age: 36, age range: 19-62). The ASD and NT groups were matched on age $(t(47)=0.54, P=0.87)$.
>
> ASD participants had previously been diagnosed by an independent clinician, according to the DSM-IV[51] or ICD-10 criteria[52] [19 Asperger Syndrome, 3 Autistic Disorder, 1 High Functioning Autism, 1 Atypical Autism]. The Wechsler Adult Intelligence Scale (WAIS 3rd edition UK) had previously been administered to assess IQ [53] and participants were matched on full-scale scores (ASD mean: 117; range: 80-142; NT mean: 120, range: 99-145; $t(47)=-0.93, P=0.36$) The Autism Diagnostic Observation Schedule (2nd edition) [54] assessment was completed by a qualified administrator to assess symptom severity in the ASD participants. Mean ADOS total score was 9.9 (range 4-19). The mean scores for the communication and reciprocal social interaction sub scores were 3.3 (range: 0-7) and 6.6 (range 4-12), respectively.
>
> An additional 57 NTs were studied as part of a replication of our key behavioural result (25 male, 32 female; mean age: 27.1, age range: 19–50) and additionally completed the Autism Spectrum Quotient (AQ) questionnaire; a 50-item self-report measure of autistic traits [32,55]. Mean AQ score was 18.43 (median: 17, range: 5-45). All participants had normal or corrected to normal vision and gave written informed consent. We performed a median split on the data such that participants were divided into high AQ (n=26) and low AQ (n=31) groups. AQ score was significantly higher in the in the high AQ group (mean=27, SD=6.4, range=18-45), relative to the low AQ group (mean=11.5, SD=3.4, range=5-17; $t(55) = 11.28, P<0.001$). The distribution of scores the low AQ group falls almost exclusively below the mean range of neurotypical scores reported in a recent meta-analysis of 73 studies administering the AQ [56]. Importantly there is considerable overlap between the scores in the high AQ group and the range reported, on average, in those with a diagnosis of ASD [56] – even though these participants do not present with any clinical need.
>
> No randomisation was used to assign subjects or conditions. All participants provided written informed consent and were compensated financially for their time and travel expenses. The study was approved by the UCL Graduate School Ethics Committee (4357/001)