Atypical cortical feedback underlies failure to process contextual information in the superior colliculus of Scn2a$^{+/-}$ autism model mice

Authors: Leiron Ferrarese[1], Hiroki Asari*[1]

Affiliations:
1. Epigenetics and Neurobiology Unit, EMBL Rome, European Molecular Biology Laboratory, Monterotondo, 00015, Italy

***Corresponding author:** Hiroki Asari, E-mail: asari@embl.it

Acknowledgments: This work was supported by research grants from EMBL (H.A.) and SFARI (#875447; H.A.); and EMBL Interdisciplinary, International and Intersectorial Postdoctoral Fellowship (EI3POD; L.F.). The EMBL Gene Editing and Embryology Facility is acknowledged for the mouse line rederivation; EMBL Genetic and Viral Engineering Facility for virus production; EMBL IT Support for provision of computer and data storage servers; and the LAR facility for taking care of animals. We thank all the Asari lab members as well as Cornelius Gross, Eleonora Satta and Michael Lombardo for many useful discussions.

Author contributions: L.F. and H.A. designed the study; L.F performed experiments and analyzed the results; L.F. and H.A. wrote the manuscript.

Competing interests: The authors declare no competing financial interests.

Key words: autism spectrum disorder; superior colliculus; implicit visual learning; Bayesian inference; hierarchical Gaussian filter.

Acronyms: AAV, adeno-associated virus; ANOVA, analysis of variance; AP, anterior-posterior; ASD, autism spectrum disorder; CNMF, constrained non-negative matrix factorization; CNO, clozapine N-oxide; DAPI, 4',6-diamidino-2-phenylindole; DREADD, designer receptors exclusively activated by designer drugs; DV, dorsal-ventral; ES, early stable; FDR, false discovery rate; GECI, genetically-encoded calcium indicator; hM4Di, modified human M4 muscarinic acetylcholine receptor for inhibitory DREADD; HGF, hierarchical Gaussian filter; IR, infrared; ISI,: interstimulus interval; L5, layer 5; LS, late stable; ML, medial-lateral; PFA, paraformaldehyde; PBS, phosphate buffer saline; RGC, retinal ganglion cell; SC, superior colliculus; Scn2a, sodium voltage-gated channel alpha subunit 2; V1, primary visual cortex; VL, volatile; WT, wild-type.

Abstract

Sensory abnormalities in autism are thought to arise from atypical Bayesian inference processes in the brain. It remains elusive, though, which neuronal circuits are involved. Here we focused on the superior colliculus and performed *in vivo* two-photon calcium imaging while the mice were engaged in an implicit visual learning task. Following the Bayesian learning principles, wild-type mice showed contextual modulation of the visual responses in the superior colliculus and pupil dynamics, but not in the signals from the retina. In contrast, Scn2a-haploinsufficient autism-model mice exhibited anomalous modulation patterns. By chemogenetically blocking the feedback input from the primary visual cortex to the superior colliculus selectively, we identified that this top-down signaling mediated the observed contextual modulations, both at the neurophysiological and behavioral levels. These results indicate that the corticotectal input is critical for contextual sensory integration in the superior colliculus, and that its anomaly underlies atypical learning in autism spectrum disorder.

Introduction

Bayesian inference provides a theoretical framework for understanding how the brain combines incoming sensory information with prior knowledge to shape perceptual experiences and guide behavior (Friston, 2012). In particular, a hierarchical Bayesian learning model has been successfully applied to explain atypical behaviors observed in individuals with autism spectrum disorder (ASD; Pellicano and Burr, 2012; Chrysaitis and Seriès, 2023): e.g., how overlearning and overestimation of the environmental volatility result in a reduced response to unexpected events in ASD (Lawson et al., 2017). However, the current experimental evidence supporting this hypothesis is limited in both sample sizes and computational models (Chrysaitis and Seriès, 2023).

How do neuronal circuits generate such differences in learning behavior between subjects with and without ASD? The superior colliculus (SC) is a promising target to address this question. It is a midbrain hub for the visual system, implicated in atypical visual processing in ASD (Marco et al., 2011; Johnson et al., 2016; Jure, 2022). For instance, gaze management and attentional focus are two major functions of SC that are often altered in ASD (Lomber et al., 2002; Kato et al., 2011). SC also plays a role in the initial automatic bias toward global visual processing (Sridharan et al., 2017). Individuals with ASD, however, typically focus too much on details (Jure, 2022), suggesting anomalous processing in SC. Moreover, extensive connections between SC and various brain regions (May, 2006) support that SC activates emotional and motor networks, allowing animals to perceive and respond appropriately to sensory stimuli with attentional prioritization, such as the innate avoidance response to an overhead dark looming stimulus in mice (Dean et al., 1989; Yilmaz and Meister, 2013; Wei et al., 2015; De Franceschi et al, 2016). The involvement of SC in these processes highlights its relevance to the neurophysiological mechanisms underlying ASD (Jure, 2022).

From the Bayesian inference perspective, SC is expected to be central for a hierarchical recursive system of visual inference in the brain (**Fig. 1A**; Friston, 2012; Jure, 2022). Firstly, neurons in the superficial layers of SC receive direct visual input from the retina, and play a role in the processing of visual motion and orientation (Lomber et al., 2001; Feinberg and Meister, 2015; Inayat et al., 2015; Li et al., 2020). These retinal inputs can be considered as "*likelihood*" of the sensory evidence. Secondly, superficial SC neurons also receive a prominent feedback signal from the primary visual cortex (V1; Wang and Burkhalter, 2013; Zingg et al., 2017; Benavidez et al., 2021), which is implicated to help integrate sensory and perceptual information. Recent studies have revealed that the V1-to-SC projection provides contextual gain modulation, enhancing the response amplitude of SC neurons to looming stimuli, while also acting as a negative gain modulator for the saliency of complex visual scenes (Zhao et al., 2014; Ahmadlou

et al., 2017). Moreover, the corticotectal synapses show nonlinear properties, suggesting their role in regulating sensory-guided innate behavior and modulating attention and learning processes (Zhao et al., 2014; Ruediger and Scanziani, 2020; Tang and Higley, 2020). Therefore, although some controversies still remain on the modulatory role of the corticotectal inputs (Hoffmann and Straschill, 1971; Ogasawara et al., 1984; Wang et al., 2010), one can assume that they contribute to the formation of "*prior"* beliefs. Finally, given the convergence of retinal inputs and cortical feedback onto SC, one can further assume that local SC neurons are involved in the computation of "*posterior*" perception.

To better understand how ASD affects inference processes involving SC circuit computations, we developed an implicit visual learning task that followed the Bayesian update scheme based on the hierarchical Gaussian filter (HGF) model, much as those behavioral paradigms in previous human psychophysics studies (Mathys et al., 2014; Lawson et al. 2017). Using *in vivo* two-photon calcium imaging, we then recorded the population activity of SC neurons in head-fixed wild-type (WT) and ASD-model mice engaged in this task. Specifically, here we investigated the impact of *sodium voltage-gated channel alpha subunit 2 (Scn2a)* mutation, a genetic variant strongly associated with ASD (Ben-Shalom et al., 2017). The *Scn2a* constitutes the $Na_V1.2$ sodium channels and regulates dendritic excitability of mature cortical pyramidal neurons (Planells-Cases et al., 2000), including those forming the corticotectal pathway (Zingg et al., 2017; Spratt et al., 2019). Haploinsufficiency of $Na_V1.2$ reduces the backpropagation of action potentials into dendrites, impairing synaptic plasticity and strength in layer 5 (L5) corticofugal pyramidal neurons (Spratt et al., 2019). Therefore, *Scn2a*$^{+/-}$ mice are expected to have atypical corticotectal projections, leading to abnormal neuronal modulations within the SC circuitry.

Consistent with the Bayesian learning principles, here we report that WT mice exhibited a robust environmental contextual modulation in SC as well as the pupil size, but not in the input from the retina. In particular, the activity of SC neurons followed posterior expectations on the

environmental volatility rather than cue-outcome contingencies *per se*. In contrast, we found anomalous contextual modulation in the SC of Scn2a$^{+/-}$ mice and no modulation in their pupil dynamics, indicating abnormal integration of sensory information and prior beliefs. These ASD-model mice showed no adaptation to increased outcome predictability in stable environments, and a larger number of SC neurons represented low-level outcome expectations in the HGF model framework. We further demonstrated that chemogenetically inhibiting the visual cortical feedback to SC fully abolished contextual modulations in both WT and Scn2a$^{+/-}$ mice, both at the physiological and behavioral levels. These results suggest that the corticotectal feedback carries information about a prior belief on the state of the world, and that animals direct their attention to sensory features through top-down modulation of the sensory responsiveness. Atypical top-down sensory attenuation can thus lead to abnormalities in sensory attention in ASD.

Results

Implicit visual learning paradigm to test behavioral and physiological anomalies in autism model mice

Individuals with autism spectrum disorder (ASD) often struggle to learn about probabilistically aberrant events (Lawson et al., 2017). Understanding the neuronal processes underlying implicit learning and the integration of contextual information will thus provide valuable insights into the neurobiological mechanisms of ASD (Chrysaitis and Seriès, 2023). In this study, we focused on the mouse visual system to study a) how neuronal circuit computation contributes to implicit learning about visual environmental contexts; and b) how it is altered in a mouse model of ASD (Scn2a$^{+/-}$). We thus began by designing an implicit visual learning paradigm based on Bayesian principles (Mathys, 2011), i.e., probabilistic associative learning of cue-outcome stimulus sequences.

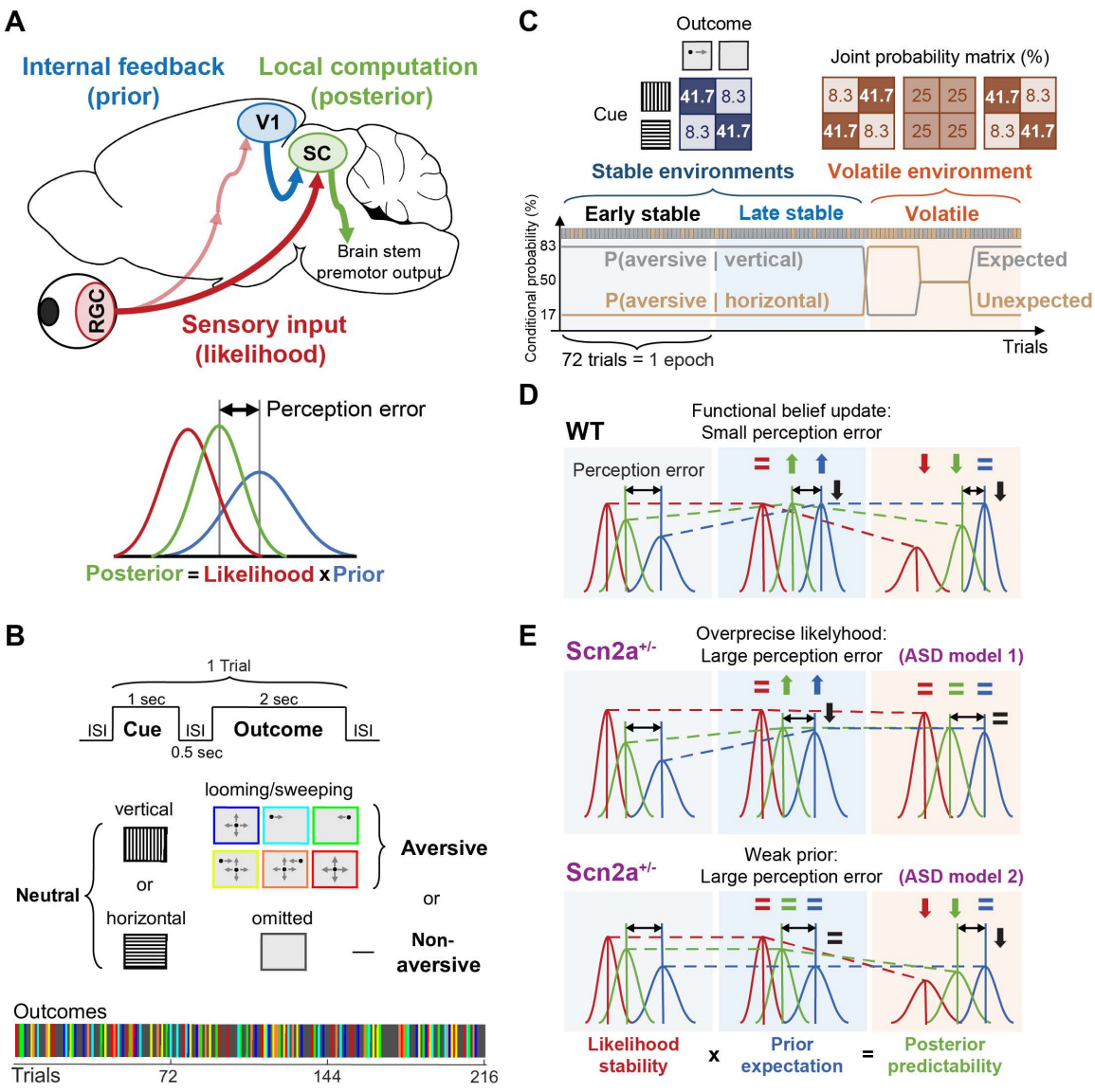


Figure 1: Bayesian implicit learning task to compare the visual responses of the superior colliculus between wild-type and autism model mice.

A: Rationale of experimental design. The superficial superior colliculus (SC) receives direct sensory input from reginal ganglion cells (RGCs) and feedback from the primary visual cortex (V1). In a Bayesian inference framework, one can assume that the retinal input represents the likelihood (dark red), while V1 provides information about the animal's prior belief (blue). Neurons in SC can then be considered to compute the posterior contextual perception (light green). **B**: Schematic of the implicit visual learning task. In each trial, a neutral cue (vertical or horizontal gratings; 1 s) is followed by an aversive outcome (looming and/or sweeping dark spot, randomly chosen from six different types; 2 s) or non-aversive one (omitted; 2 s) with an inter-stimulus interval (ISI) of 0.5 s. In total, the task consisted of 216 trials, including 108 aversive outcome trials (color-coded) and 108 non-aversive ones (dark gray). **C**: The conditional probability of cue-outcome contingencies (expected, gray; unexpected, light brown) during stable and volatile environments. The stable environment is twice as long as the volatile one (orange), and divided into early- and late-stable epochs (72 trials each; light gray and light blue, respectively). **D**: The task aims to let animals form expectations on cue-outcome probabilities during the stable environment, leading to increased prior and posterior beliefs in the late-stable epoch. In the volatile epoch, however, the precision of the posterior decreases as the likelihood reduces. **E:** In Scn2a$^{+/-}$ mice, sensory hypersensitivity (top, model 1) and/or imprecise priors (bottom, model 2) lead to the posterior perception dominated by sensory input (likelihood).

116 Our implicit learning task was carefully tailored with visual stimuli to elicit robust responses
117 in the mouse superior colliculus (SC; Cang et al., 2018), such as static gratings with different
118 orientations (Basso and May, 2017; De Franceschi and Solomon, 2018) and looming and
119 sweeping stimuli that trigger SC-mediated arousal and innate defensive behavior (Dean et al.,
120 1989; Yilmaz and Meister, 2013; Wei et al., 2015; De Franceschi et al., 2016). Specifically, we
121 used vertical and horizontal gratings as neutral cues; and a combination of sweeping and looming
122 dark spots as aversive outcomes while a gray screen (omitted stimulus) as a non-aversive
123 outcome (**Fig. 1B**). Since the task did not involve any reward or punishment, we exploited visual
124 threats in the outcome stimulus to motivate animals to recognize stimulus sequences. To increase
125 unpredictability, we introduced uncertainty by randomizing the type and direction of the visual
126 threat at each presentation (**Fig. 1B**; see Methods for details).

127 The task followed a Bayesian update scheme modeled by a hierarchical Gaussian filter
128 (HGF; Mathys, 2011; Lawson et al., 2017), consisting of a stable environment followed by a
129 volatile one (**Fig. 1C**). During the stable environment (144 trials in total), we presented aversive
130 outcome stimuli more frequently after one neutral cue stimulus than after another. Mice were then
131 expected to form probabilistic associations for these cue-outcome stimulus sequences. At a small
132 probability, trials with expected outcomes were interspersed with unexpected ones. This allowed
133 animals to determine whether an unexpected outcome indicates a deviation from a stable stimulus
134 pattern or a change in the pattern itself. In the stable environment, the stability and precision of
135 the sensory inputs (likelihood) remained high. Thus, the prior expectation should increase in the
136 early stable epoch (ES; initial 72 trials) and remain high in the late-stable epoch (LS; last 72 trials);
137 and so does the posterior belief about the environment, or the predictability, that can be given by
138 the product of the likelihood and the prior (**Fig. 1D**). In the following volatile epoch (VL, 72 trials in
139 total), we then introduced variations in the probability of aversive stimuli following each neutral
140 stimulus. This challenged the animal's expectations on the stimulus sequences. The environment

141 became less precise (i.e., likelihood decreased) while the prior belief remained relatively precise
142 because updating the prior is slow in general. Thus, the posterior precision should decrease and
143 shift towards the prior in wild-type (WT) animals.

144 In Scn2a$^{+/-}$ mice (**Fig. 1E**), we instead expect that their posterior is rather dominated by
145 the likelihood due to sensory hypersensitivity and/or imprecise prior belief formation as implicated
146 in human studies (Pellicano and Burr, 2012; Haker et al., 2016; Lawson et al., 2017). On the one
147 hand, hypersensitivity to sensory inputs will lead to an overly high likelihood regardless of the
148 stability of the environment. The posterior probability will then follow the changes in the prior,
149 hence its precision will increase in the LS epoch, but stay unchanged in the VL epoch (model 1).
150 On the other hand, if the prior expectations are incoherent and not updated during the task, the
151 posterior predictability will follow the changes in the likelihood. Thus, its precision will remain in
152 the LS epoch but decrease in the VL epoch (model 2). In either case, the prior and posterior
153 predictions will not become close enough in ASD, leading to a larger perception error than in WT.

154 Distinct contextual modulation patterns in SC between WT and Scn2a$^{+/-}$ mice

155 If neurons in SC compute contextual-dependent predictions about the environment, they should
156 reflect the posterior predictability distribution, while inputs from retinal ganglion cells (RGCs) and
157 the primary visual cortex (V1) represent the likelihood and the animal's prior belief, respectively.
158 To test this hypothesis, we conducted experiments on awake, head-fixed mice expressing
159 genetically-encoded calcium indicators (GECIs) in SC somata delivered via adeno-associated
160 viral vectors (AAVs; see Methods for details; **Fig. 2A,B**). Specifically, we exposed the mice to the
161 implicit visual learning task as described above, and simultaneously performed *in vivo* two-photon
162 calcium imaging of superficial SC through a cranial window, together with pupillometry for
163 behavioral monitoring (Dombeck et al., 2007; Feinberg and Meister, 2015; Molotkov et al., 2023;
164 **Fig. 2A,B; Supplementary Fig. 1A**).

On average we recorded 123±37 cells per animal (mean ± standard deviation; 27 mice) and estimated their spiking activity using CaImAn (Pnevmatikakis et al., 2019; $Z(t)$ in ***z***-score, see Methods for details). As expected (Cang et al., 2018; De Franceschi and Solomon, 2018), superficial SC neurons generally responded well to the onset and/or offset of the static gratings (**Supplementary Fig. 1**). About a half of them were responsive to the looming and/or sweeping stimuli (60±19 cells per recording), whereas in most cells we found virtually no response to the omitted stimulus (gray screen; e.g., **Supplementary Fig. 1B**). Here we considered cells responsive if the trial average of their visual response dynamics exceeded a threshold adjusted for each recording (e.g., $\max[z(t)] > 1$, where $z(t) = \langle Z(t) \rangle_{\text{trials}}$; see Methods for details).

To examine the SC dynamics during the implicit learning task, we compared the aversive stimulus responses at the population level across the three analysis epochs in our paradigm (ES, LS, and VL; **Fig. 1C**). For each responsive cell, we calculated the average response during the stimulus presentation period in each epoch ($\bar{z} = \langle z(t) \rangle_t$; e.g., **Fig. 2C**), and subtracted the mean across the epochs for normalization ($\Delta z = \bar{z} - \langle \bar{z} \rangle_{\text{epochs}}$; e.g., **Fig. 2C-E**). Using this normalized response, Δz, we found that SC neurons in WT animals responded significantly weaker to the aversive stimuli during the LS epoch than the ES or VL epoch (**Fig. 2E**, 1,037 cells in total from 17 mice; *p*(ES-LS)=0.007, *p*(ES-VL)=0.8, *p*(LS-VL)=0.04, Tukey-Kramer test). This suggests that the responses of SC neurons in WT mice are modulated by the visual contexts in two ways. First, when the stimulus pattern is reliably predictable during the stable environment, the responses to aversive stimuli become progressively attenuated (adaptation). Second, when the stimulus pattern deviates from an animal's expectation during the volatile environment, the response of SC neurons is reset to the baseline and becomes as high as in the initial ES epoch before adaptation.

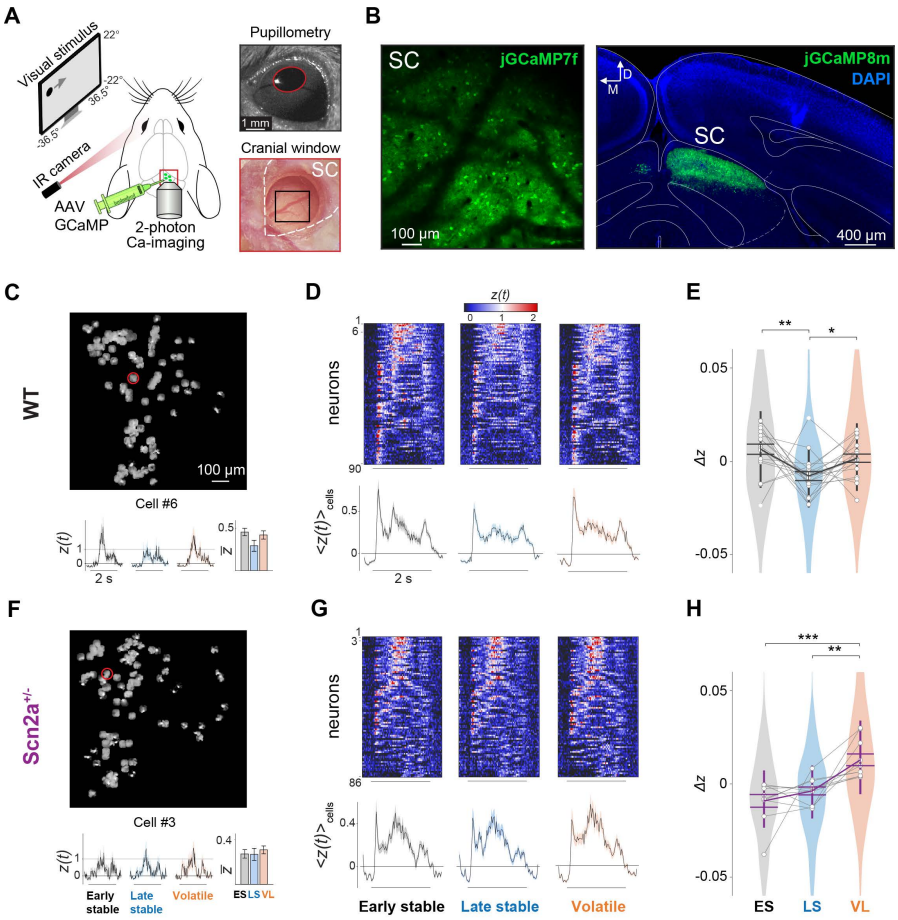
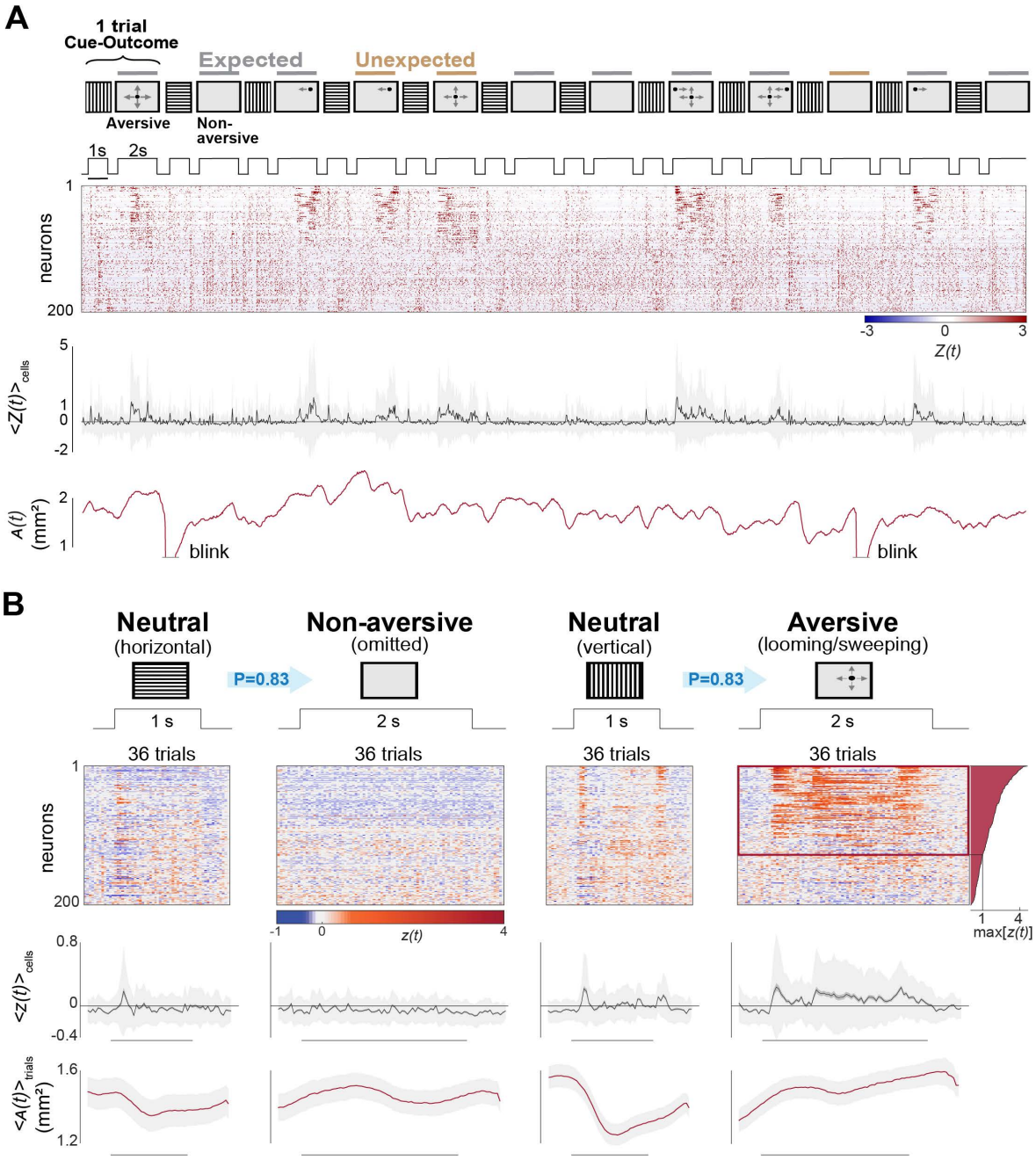


Figure 2: Neurons in the superior colliculus show contextual adaptation to a stable environment in wild-type mice but not in Scn2a$^{+/-}$ mice.

A: Experimental setup. *Left*: The schematic illustrates a head-fixed mouse pre-injected with adeno-associated viruses (AAVs) carrying genetically encoded calcium indicators (e.g., jGCaMP7f). Visual stimulus was presented to the left eye. *Right:* The pupil dynamics of the left eye were recorded with an infrared (IR) camera (top; pupil outline, red). Two-photon calcium imaging of the right superior colliculus (SC) was performed through a glass window implanted with a silicone plug to expose SC (bottom). **B**: *Left:* An example frame of motion-corrected two-photon calcium imaging in the mouse SC. *Right*: A coronal slice image taken two months after AAV injections, demonstrating a high expression of jGCaMP8m (green) in the right SC, along with 4',6-diamidino-2-phenylindole stain (DAPI, blue). **C**: Deconvoluted spatial components extracted using CaImAn in a sample wild-type (WT) mouse displaying the topographic position of SC neurons responsive to aversive stimuli (top). Representative aversive stimulus responses of an SC neuron (circled in red on top), averaged over 36 aversive outcome trials for each analysis epoch (bottom: $z(t) = \langle Z(t) \rangle_{\text{trials}}$; shaded area, s.e.m.). The average over the stimulus period ($\overline{z} = \langle z(t) \rangle_t$) and s.e.m. are shown on the right for each epoch (from left to right: early stable (ES), gray; late stable (LS), light blue; volatile (VL), orange). **D**: Heatmap ($z(t)$) and the population average ($\langle z(t) \rangle_{\text{cells}}$) of the aversive stimulus responses in each epoch for the responsive SC neurons (sorted with $\max[z(t)] > 1$), recorded from a representative WT mouse (90 cells from C; shaded area, s.e.m.). **E**: Distributions of the normalized SC responses across epochs ($\Delta z = \overline{z} - \langle \overline{z} \rangle_{\text{epochs}}$) from all WT animals in the dataset (1,037 cells from 17 mice): violin plots, distribution of all cells; gray lines with open circles, average for each mouse; vertical bars, interquartile range; thick line graph with error bars, mean ± s.e.m.; ***, p<0.001; **, p<0.01; *, p<0.05 with Tukey-Kramer test for pairwise comparisons. See Methods for details. **F-H**: Corresponding data for Scn2a$^{+/-}$ mice (F, simultaneously recorded SC cell location in the image plane and a representative cell's responses; G, representative data from all simultaneously recorded responsive cells (86 cells); H, population data for Scn2a$^{+/-}$ mice, 630 cells in total from 10 mice).



Supplementary Figure 1: Neuronal activity of the mouse superior colliculus and pupil dynamics in response to cue-outcome stimuli in the implicit visual learning task.
A: Trial-by-trial response example during the early stable (ES) epoch, with 83% expected and 17% unexpected outcomes. *From top to bottom*: The spiking activity of the superior colliculus (SC) neurons ($Z(t)$ in z-score) estimated from recorded calcium dynamics; corresponding population average response ($\langle Z(t)\rangle_{\text{cells}}$; gray shade, ±1 standard deviation); and concurrent pupil area dynamics ($A(t)$). **B**: Responses of SC neurons to each of the four stimulus types presented in the ES epoch (from left to right: neutral cue stimulus associated with non-aversive outcome, non-aversive omitted stimulus, neutral cue stimulus associated with aversive outcome, and aversive stimuli). *From top to bottom*: the trial average responses of individual SC cells ($z(t) = \langle Z(t)\rangle_{\text{trials}}$), sorted by the peak aversive stimulus response values ($\max[z(t)]$, rightmost histogram; red rectangle, responsive cells with $\max[z(t)]$>1); the average population response ($\langle z(t)\rangle_{\text{cells}}$); and the average pupil dynamics ($\langle A(t)\rangle_{\text{trials}}$). Gray shade, ±1 standard deviation.

226 In contrast, Scn2a$^{+/-}$ mice displayed different context-dependent modulations in SC during
227 the implicit visual learning task: no adaptation in the stable environment, but increased neuronal
228 responses during the volatile environment (**Fig. 2F-H**, 630 cells in total from 10 mice; *p*(ES-
229 LS)=0.7, *p*(ES-VL)=5e-4, *p*(LS-VL)=0.009, Tukey-Kramer test). These response patterns of SC
230 neurons were significantly different between WT and Scn2a$^{+/-}$ mice (*p*(epoch x group) = 0.002, 3-
231 way nested analysis of variance (ANOVA); see Methods for details). This suggests that, unlike
232 WT animals, SC neurons in Scn2a$^{+/-}$ mice fail to attenuate their responses even when the stimulus
233 pattern is kept stable and predictable. They are, however, still sensitive to increased volatility, and
234 thus their response during the VL epoch exceeded the baseline activity in the ES epoch. This is
235 consistent with the observations on ASD subjects who tend to overestimate environmental
236 volatility during stable environments, and overreact to increased unpredictability during volatile
237 environments (Lawson et al., 2017).

238 Context-dependent modulation is absent in retinal output responses

239 How do SC neurons achieve such context-dependent modulations? One possibility is that the
240 contextual modulation in SC is inherited from RGCs. To investigate the contribution of bottom-up
241 RGC input to SC, we performed *in vivo* two-photon calcium imaging of RGC axon terminals in SC
242 (Molotkov et al., 2023) while the animals were engaged in the implicit visual learning task (**Fig.**
243 **3A**). On average we recorded 240±34 putative axonal boutons per animal (mean ± standard
244 deviation; 5 mice), of which 131±29 were responsive to the looming and/or sweeping stimuli
245 ($\max[z(t)] > 1$). These axons of RGCs projected to SC showed consistent responses to the
246 aversive stimuli regardless of the visual context (**Fig. 3B**). Across populations, we found no
247 significant contextual modulation across ES, LS, and VL epochs (**Fig.3C**; 655 putative RGC
248 boutons in total from 5 mice; *p*(ES-LS)=0.8, *p*(ES-VL)=0.6, *p*(LS-VL)=0.9, Tukey-Kramer test).
249 This excludes the possibility that the contextual modulation in SC arises from the retina.

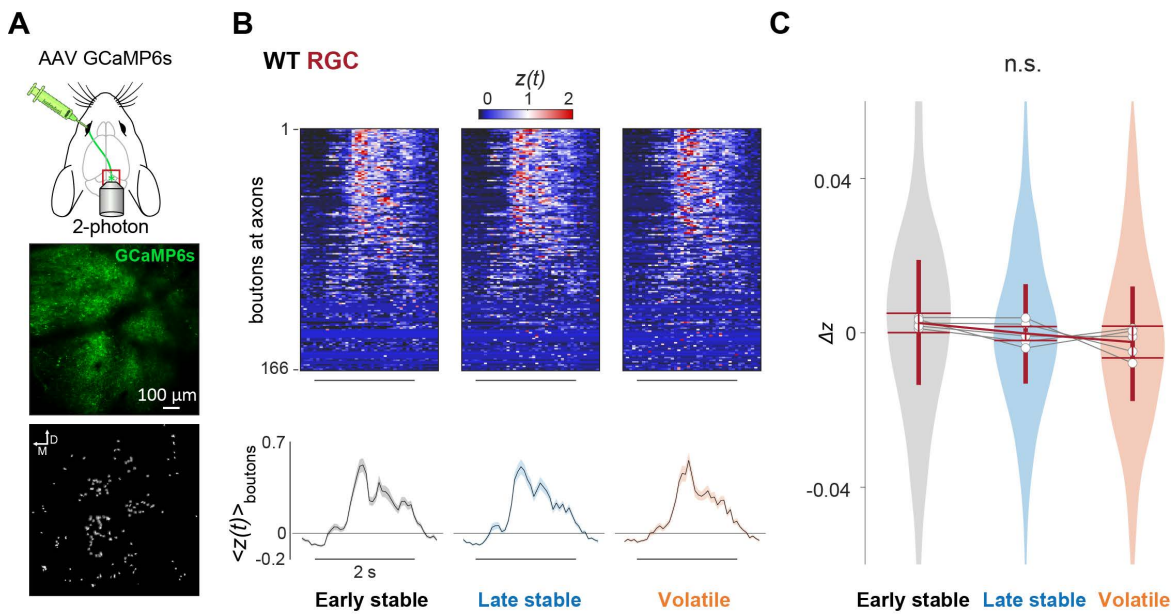


Figure 3: Retinal ganglion cells projected to the superior colliculus do not show any context-dependent modulation during the implicit learning task.
A: Experimental setup. Animals were intraocularly injected with adeno-associated viruses (AAVs) carrying GCaMP6s and underwent cranial window surgery over the right superior colliculus (SC) to monitor the activity of retinal ganglion cell (RGC) axons *in vivo* (top). An example frame of RGC axonal imaging in SC (middle) and deconvoluted spatial components by CaImAn (bottom). **B**: Aversive stimulus responses of simultaneously recorded RGC putative boutons (top: heatmap of $z(t) = \langle Z(t) \rangle_{\text{trials}}$, sorted by $\max[z(t)]$) and the population average (bottom, $\langle z(t) \rangle_{\text{boutons}}$; 166 boutons in total from A; shaded area, s.e.m.) in each analysis epoch (shown in the same format as in **Fig. 2D,G**). **C**: Distributions of the normalized RGCs responses across epochs ($\Delta z = \overline{z} - \langle \overline{z} \rangle_{\text{epochs}}$, where $\overline{z} = \langle z(t) \rangle_t$; 655 boutons from 5 WT mice), shown in the same format as in **Fig. 2E,H**. Violin plots, distribution of all boutons; gray lines with open circles, average for each mouse; vertical bars, interquartile range; thick line graph with error bars, mean ± s.e.m. No significant difference was found by pairwise comparisons (Tukey-Kramer test).

Chemogenetically inhibiting corticotectal projection abolishes context-dependent modulation in the superior colliculus

Another possible source for the contextual modulation in SC is the corticotectal projection (**Fig. 1A**). Besides sensory input from RGCs, the superficial layer of SC receives substantial feedback signals from layer 5 (L5) pyramidal neurons in V1 (Zingg et al., 2017). These V1 L5 neurons integrate feedforward visual information through their basal dendrites and feedback contextual information via the apical dendrites (Larkum, 2013), and provide output to SC for controlling the response gain (Zhao et al., 2014; Ahmadlou et al., 2017; Tang and Higley, 2020). Cortical pyramidal neurons in Scn2a$^{+/-}$ mice have impaired synaptic functions due to a reduced expression

of the $Na_V1.2$ sodium channel (Spratt et al., 2019). This may result in a failure to incorporate visual and contextual information in V1, which in turn may lead to an anomalous modulation of SC as observed in our paradigm (**Fig. 2**).

To examine the role of V1 feedback in the contextual modulation of SC, we employed a chemogenetic approach to selectively inhibit the V1 L5 corticotectal neurons using a designer receptor exclusively activated by designer drug (DREADD; Armbruster et al., 2007). Specifically, we achieved selective expression of modified human M4 muscarinic acetylcholine receptors for inhibitory DREADD (hM4Di) in the target population by injecting Cre-dependent AAV8 carrying the hM4Di receptor into V1 and AAV2retro carrying Cre-recombinase into SC (**Fig. 4A,B**). Systemic administration of the DREADD actuator, clozapine N-oxide (CNO), then activated hM4Di, hence effectively inhibited the V1 neurons projecting to SC. As before, AAVs carrying GECIs were also injected into SC for monitoring the visual responses of SC neurons with *in vivo* two-photon microscopy.

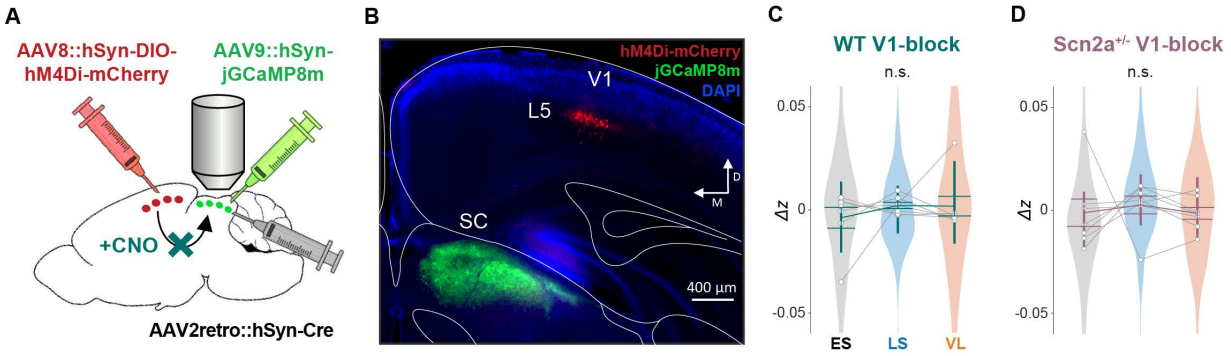
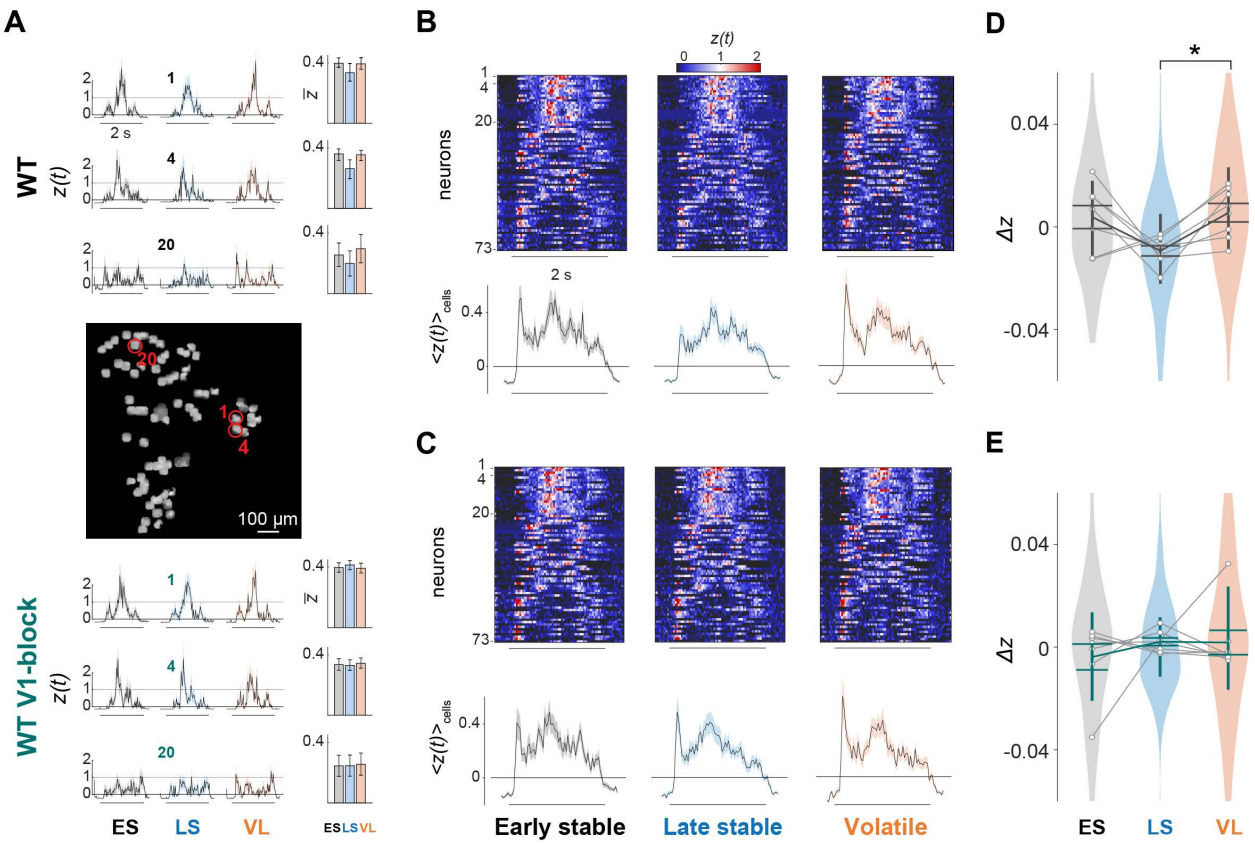


Figure 4: Inhibition of the corticotectal projection abolishes context-dependent modulation in the superior colliculus.
A: The schematic of the viral vector strategy for chemogenetic inhibition of the projection from the primary visual cortex (V1) to the superior colliculus (SC). AAV8::hSyn-DIO-hM4Di-mCherry was injected into V1, while AAV2retro::hSyn-Cre and AAV9::hSyn-jGCaMP8m were co-injected in SC. Systemic administration of clozapine N-oxide (CNO) activated the hM4Di receptor, hence inhibiting the V1-to-SC projection. **B**: Histological verification of the hM4Di receptor expression (with mCherry, red) in the layer 5 (L5) cells in V1 projecting to SC. Blue, 4',6-diamidino-2-phenylindole (DAPI); green, SC neurons expressing jGCaMP8m. **C,D**: Distributions of the normalized SC responses (Δz) across epochs in WT (C, 355 neurons in total from 7 mice) or $Scn2a^{+/-}$ mice (D, 433 neurons in total from 7 mice) after chemogenetically blocking V1-to-SC projection. Data are shown in the same format as in **Fig. 2E,H**. No significant contextual-modulation was found in either case (Tukey-Kramer test).



Supplementary Figure 2: Context-dependent modulation in the superior colliculus is abolished by chemogenetic inhibition of the corticotectal projection.
A: Representative aversive stimulus responses ($z(t) = \langle Z(t) \rangle_{\text{trial}}$; gray shades, s.e.m.) of three SC neurons in a representative WT mouse expressing hM4Di in SC-projecting V1 neurons with intraperitoneal injection of saline (top; WT) or the DREADD actuator, CNO (bottom; WT V1-block). The average over the stimulus presentation period ($\bar{z} = \langle z(t) \rangle_t$) is shown on the right for each cell for each epoch (early stable, ES; late stable, LS; volatile, VL). Spatial arrangement of the cells, extracted from imaging data by CaImAn, is shown in the middle with red circles indicating the representative cells (#1, #4, and #20). **B,C:** Heatmap ($z(t)$; 73 cells) and the population average ($\langle z(t) \rangle_{\text{cells}}$) of the aversive stimulus responses in each epoch from the representative WT mouse after saline (B) or CNO (C) injections. **D,E**: Distributions of the normalized SC responses across epochs ($\Delta z = \bar{z} - \langle \bar{z} \rangle_{\text{epochs}}$) with saline (D, 355 cells in total from 7 mice; WT dataset) and CNO injections (E, 348 cells in total from 7 mice; WT V1-block dataset, reproduced from **Fig.4C**). Data are shown in the same format as **Fig. 2D,E**. *, p<0.05 with Tukey-Kramer test for pairwise comparisons.

Chemogenetic inhibition of the projection from V1 to SC abolished the context-dependent modulation of SC in both WT and Scn2a$^{+/-}$ mice (**Fig. 4C,D; Supplementary Fig. 2**). After the intraperitoneal injection of CNO, SC neurons did not show adaptation during the LS epoch or enhanced responses during the VL epoch (e.g., **Supplementary Fig. 2**). As a result, the overall SC response level did not significantly change throughout the implicit visual learning task (**Fig.**

318 **4C**; 355 cells in total from 7 WT mice: p(ES-LS)=0.8, p(ES-VL)=0.7, p(LS-VL)=1, Tukey-Kramer
319 test; **Fig. 4D**; 433 neurons from 7 Scn2a$^{+/-}$ mice: p(ES-LS)=0.7, p(ES-VL)=1, p(LS-VL)=0.9). Such
320 response patterns across epochs in the chemogenetically-manipulated animals were significantly
321 different from those in intact animals (p(epoch x group) = 0.004, 3-way nested ANOVA; see
322 Methods for details). These results support that feedback input from V1 plays a key role in
323 modulating the aversive stimulus response of SC neurons in a context-dependent manner. In
324 particular, we suggest that V1 projections provide SC with contextual information necessary for
325 achieving adaptation to expected outcomes during the stable environment.

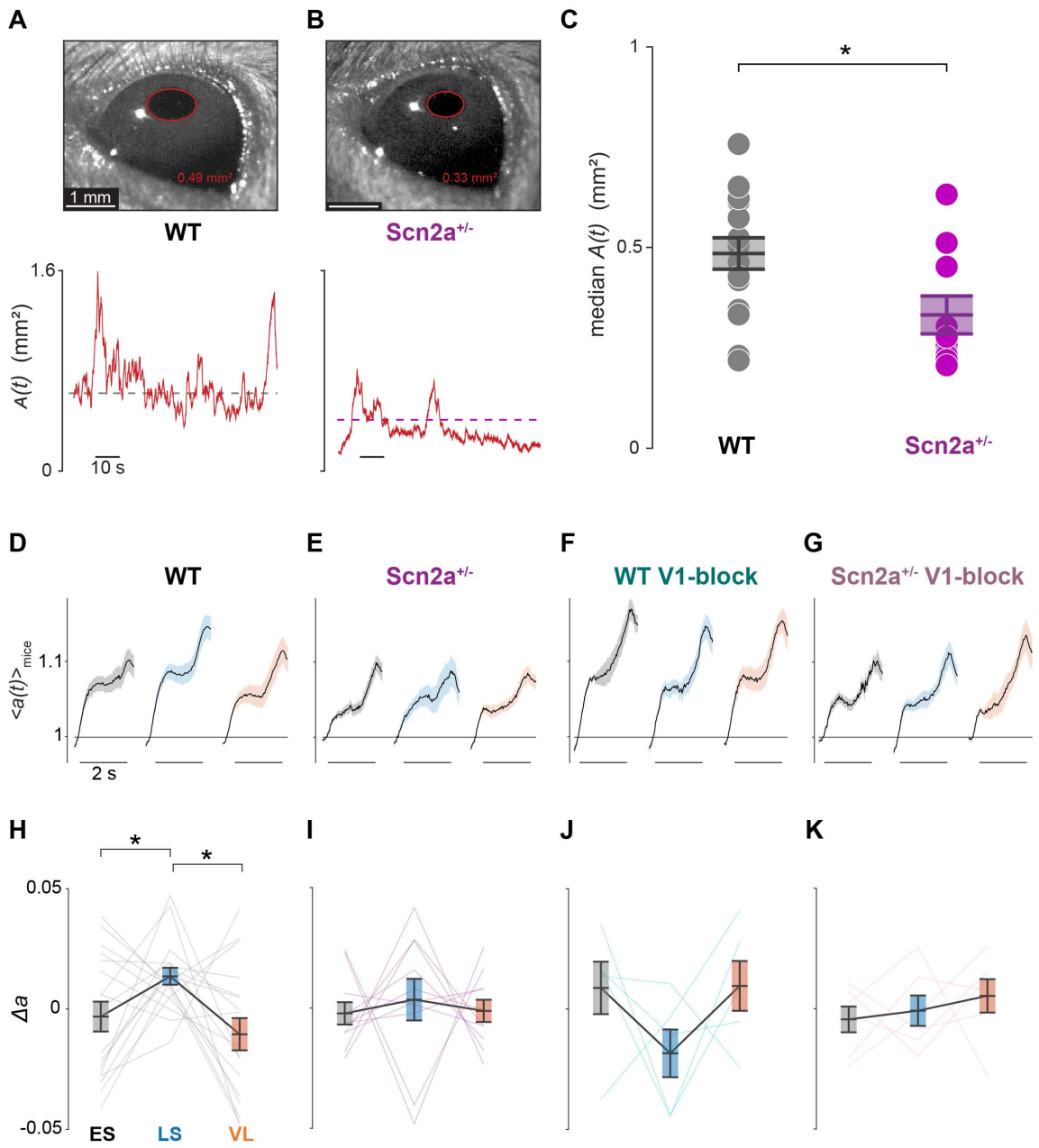
326 Pupil dynamics reflect context-dependent stimulus pattern recognition in WT but 327 not in Scn2a$^{+/-}$ mice

328 Besides neuronal activities, we collected pupillometric data (**Fig. 5**) as a proxy to monitor cognitive
329 processes of the animals (Loewenfeld, 1958). Here we in particular interpreted pupil dynamics as
330 a reflection of an animal's stimulus pattern recognition (Suzuki et al., 2018; Ganea et al., 2020;
331 Dobbins, 2021). Indeed, all mice exhibited stereotypical pupil dynamics during the implicit visual
332 learning task: i.e., pupil contraction in response to the cue stimuli, and pupil dilation during the
333 outcomes, especially when the aversive stimuli were presented (**Supplementary Fig. 1B**,
334 *bottom*). Here we focused our analysis on the pupil dynamics during the aversive outcome trials
335 across the three analysis epochs (ES, LS, and VL; **Fig. 1C**) as we did for the neuronal dynamics
336 (**Figs. 2-4**).

337 In WT animals, we found a greater pupil dilation during the LS epoch than the ES or VL
338 epoch (**Fig. 5D,H;** q(ES-LS) = 0.04, q(ES-VL) = 0.1, q(LS-VL) = 0.02, Storey's false discovery
339 rate (FDR) from Wilcoxon signed-rank test; 17 mice). This supports that pupil dilation is associated
340 with an animal's recognition of the stimulus pattern (Suzuki et al., 2018; Ganea et al., 2020;
341 Dobbins, 2021). This context-dependent modulation of the pupil dynamics was abolished when

342 the feedback signal from V1 to SC was chemogenetically blocked (**Fig. 5F,J;** *q*(ES-LS)=0.1, *q*(ES-
343 VL)=0.1, *q*(LS-VL)=0.1; 5 mice). In contrast, Scn2a$^{+/-}$ animals generally showed a smaller pupil
344 size than WT animals throughout the task (10 Scn2a$^{+/-}$ mice, 0.33 ± 0.05 mm^2, mean ± s.e.m of
345 the median pupil area; 17 WT mice, 0.49 ± 0.04 mm^2; p = 0.037, Mann-Whitney *U*-test; **Fig. 5A-**
346 **C**), and displayed no context-dependent modulation (**Fig. 5E,I;** *q*(ES-LS)=0.2, *q*(ES-VL)=0.1,
347 *q*(LS-VL)=0.1; 10 mice). Chemogenetic inhibition of the corticotectal projection did not affect the
348 pupil dynamics in Scn2a$^{+/-}$ mice (**Fig. 5G,K;** *q*(ES-LS)=0.1, *q*(ES-VL)=0.1, *q*(LS-VL)=0.1; 5 mice).
349 The modulation pattern of the pupil dynamics across epochs was significantly different between
350 WT and the other three experimental groups (*p*(epoch x group) = 0.04; 2-way ANOVA; see
351 Methods for details). These findings indicate that the influence of the corticotectal signals
352 extended beyond the neuronal activity in SC, likely playing a role in the contextualization of the
353 visual scene even at the perceptual level; and that Scn2a$^{+/-}$ mice have altered V1-to-SC signaling,
354 resulting in a failure to properly recognize the stimulus patterns during the implicit learning task.

355 Taken together, we found that the modulation of the pupil dynamics (**Fig. 5**) was well linked
356 to that of the neuronal dynamics in SC (**Figs. 2-4**): i.e., the greater the pupil dilation was, the
357 weaker the SC responses were in our experimental paradigm, much as was demonstrated in
358 previous studies (Wang et al., 2012; Joshi et al., 2016). Most importantly, Scn2a$^{+/-}$ mice showed
359 impaired neurophysiological and behavioral phenotypes, both due to anomalous signaling from
360 V1 to SC. This highlights a crucial role of the top-down corticotectal pathway in forming perceptual
361 associations to recognize and properly react to visual contexts (Tang and Higley, 2020), and a
362 possible relationship between ASD and dysfunctions of V1 L5 projection neurons (Marco et al.,
363 2011; Johnson et al., 2016; Jure, 2022).



364

365 **Figure 5: Volatility-dependent pupillometric response patterns in wild-type but not in Scn2a$^{+/-}$ mice.**
366 **A,B**: Representative image frame of the left eye (top; red, pupil outline) of WT (A) and Scn2a$^{+/-}$ mice (B)
367 and the pupil area dynamics $A(t)$ during visual stimulation (bottom; dashed line, median). **C**: median pupil
368 size of 17 WT and 10 Scn2a$^{+/-}$ mice during the implicit visual task (error bar, mean ± s.e.m): *, $p < 0.05$,
369 Mann-Whitney *U*-test. **D-G**: Pupil size dynamics of mice (average proportional change in pupil area,
370 $\langle a(t)\rangle_{\text{mice}}$; shaded area, s.e.m.) in response to the aversive stimulus in each analysis epoch (from left to
371 right: early stable, ES; late stable, LS; volatile, VL): D, 17 WT mice; E, 10 Scn2a$^{+/-}$ mice; F, 5 WT mice with
372 chemogenetic inhibition of V1-to-SC projection (WT V1-block); and G, 6 Scn2a$^{+/-}$ mice with chemogenetic
373 inhibition of V1-to-SC projection (Scn2a$^{+/-}$ V1-block). **H-K**: Normalized change in pupil size upon aversive
374 stimulus presentation in each analysis epoch ($\Delta a = \overline{a} - \langle\overline{a}\rangle_{\text{epochs}}$): thin line, individual animal; error bars,
375 population average and s.e.m.; H-K, from corresponding groups in D-K, respectively. *, $q < 0.05$, Storey's
376 FDR from Wilcoxon signed-rank test.

SC neurons represent expectations on environmental volatility in WT mice, while those on outcomes in Scn2a$^{+/-}$ mice

Thus far we have analyzed the overall activity level of SC populations during the implicit visual learning task, and demonstrated that their dynamics were largely consistent with the behavior of the posterior predictability from the Bayesian inference viewpoint (**Figs. 2-4**). What do then individual neurons in SC represent? To address this question, we simulated how an ideal Bayesian observer should respond to the sequence of the cue-outcome contingencies in our experimental paradigm (Mathys, 2011; Friston, 2012), and compared the dynamics of the underlying model parameters with those of the observed neuronal responses (**Fig. 6**).

As an ideal observer model, here we employed an HGF model similar to the one used to explain the differences between subjects with and without ASD in their behavioral responses to environmental uncertainty (Lawson, et al., 2017). Using variational Bayes inversion processes, the model infers hidden states that represent beliefs about the sensory input at three hierarchical levels (**Fig. 6A**; see Methods for details). The lowest level concerns directly on the outcome (aversive versus non-aversive outcomes; X_1); the next level on the outcome probability (cue-outcome contingencies; X_2); and the highest level on the likelihood of environmental changes (stability versus volatility; X_3). Inference at each level i depends on that at a higher level, and involves the following five parameters: prior predictions on the outcome ($\hat{\mu}_i$); variance of prediction belief ($\hat{\sigma}_i$); learning rates (α_i); precision-weighted prediction errors (ε_i); and uncertainties (unc_i). In total, the HGF model has 13 parameters that are dynamically updated on a trial-to-trial basis (as the two are redundant; **Fig.6B**; see Methods for details). Each of these parameters provides an insight into distinct aspects of the Bayesian inference process at different levels of uncertainty associated with the sensory input.

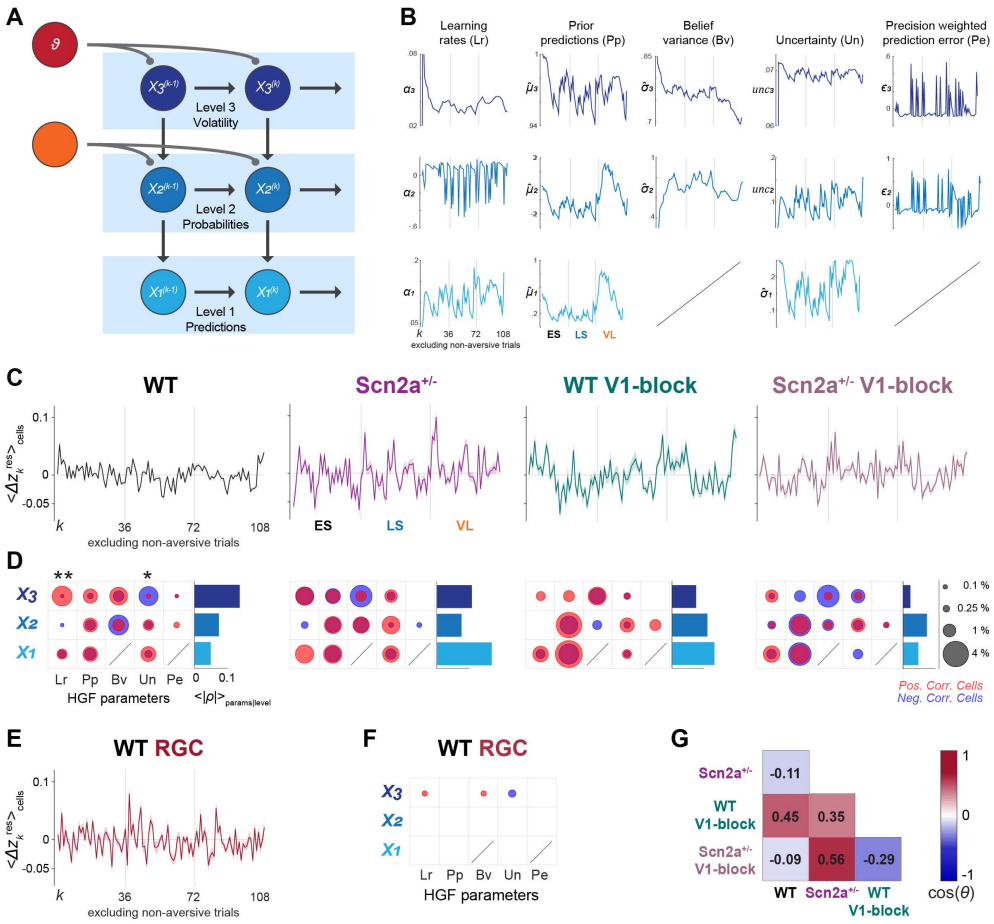


Figure 6: Superior colliculus neurons in wild-type mice compute expectations about the volatility of the visual environment while those in Scn2a$^{+/-}$ mice focus instead on outcome expectations. **A:** Schematic representation of the generative HGF model as a Bayesian network. The perceptual model tracks a mouse's learning of the task structure across three levels. Level 1 (X_1) represents trial-wise stimulus transitions from one stimulus to the next; X_2 the probability of the transitions of contingencies; and X_3 the environment volatility, where k is the current trial number. Mice hold and update beliefs about the true quantities at each level, with a mean μ and a variance σ; ϑ and ω are mouse-specific parameters that couple the levels and determine the respective speed of belief updating about phasic volatility and transition contingencies. **B**: HGF model parameter trajectories, generated with the cue-outcome contingencies of the implicit visual learning task. Non-aversive outcome trials were excluded to match neuronal trajectories (in C). Learning rates (Lr) at X_1-X_3; Prior predictions (Pp) at X_1-X_3; Believe variances (Bv) at X_2 and X_3; Outcome uncertainty (Un) at X_1; Information uncertainty at X_2; Volatility uncertainty at X_3; Precision-weighted prediction errors (Pe) at X_2 and X_3. See Methods for details. **C**: Population average of the normalized neuronal trajectories for each experimental group ($\langle \Delta z_k^{\mathrm{res}} \rangle_{\mathrm{cells}}$). From left to right: WT, 1150 neurons from 19 mice; Scn2a$^{+/-}$, 523 neurons from 9 mice; WT V1-block, 289 neurons from 6 mice; Scn2a$^{+/-}$ V1-block, 386 neurons from 6 mice. **D**: *Left*, Neural representation of HGF parameters for each experimental group (as in C), given as the percentage of SC neurons whose response trajectory was significantly correlated with each HGF model parameter trajectory (from B): circle area, percentage of cells with q<0.05 from Storey's FDR; red, positive correlation; blue, negative correlation. A bias of the correlation polarity for each HDR parameter was examined with a shuffling test with FDR correction: *, q<0.05; **, q<0.01. *Right*, absolute correlation between the population average neuronal trajectories ($\langle \Delta z_k^{\mathrm{res}} \rangle_{\mathrm{cells}}$ from C) and HGF trajectories (from B), averaged for each level, $\langle |\rho| \rangle_{\mathrm{params|level}}$. **E,F:** Population average trajectory of RGC axonal responses (E as in C) and the percentage of RGC axonal boutons whose trajectories were significantly correlated with each HGF parameter trajectory (F as in D). **G**: Pairwise cosine similarity index of neural representations of HGF parameters between different experimental groups. See Methods for details.

To make a fair comparison between the model and neuronal outputs, we focused only on the trials with aversive outcomes because SC neurons were generally not responsive to the non-aversive stimulus in our paradigm (i.e., the omitted stimulus; see, e.g., **Supplementary Fig. 1B**). Thus, after running the simulation with the entire trial sequence, we concatenated the output of the aversive outcome trials alone to generate the trajectories of the HGF model parameters (**Fig. 6B**). To generate corresponding trajectories of the neuronal signals ($\Delta z_k^{\mathrm{res}}$; **Fig. 6C**), we normalized the observed responses of individual neurons for each of the six different types of aversive stimuli, and further eliminated slow non-stationary components (see Methods for details). We then conducted pairwise Pearson's correlation tests (with FDR correction) between the neuronal and model-parameter trajectories to characterize neuronal representations of the HGF model in the SC of WT and Scn2a$^{+/-}$ mice.

We obtained two conclusions from this analysis. First, most of the HGF parameters were represented by some subsets of SC neurons, evidenced by statistically significant correlations between the two trajectories (**Fig. 6D**, *left*). This was the case with both WT and Scn2a$^{+/-}$ animals, regardless of the chemogenetic inhibition of the corticotectal projection. In contrast, hardly any HGF parameter was well represented by the incoming signals from the retina (**Fig. 6E,F**).

Second, SC neurons in WT mice had a distinct representation of the HGF parameters, compared to those in Scn2a$^{+/-}$ animals or when corticotectal projections were chemogenetically blocked (**Fig. 6D,G**). Specifically, WT mice showed a significant bias in the representation of two HGF parameters at level 3 (learning rate α_3, 3.4% of positively and 0.1% of negatively correlated cells, q = 0.002, shuffling test; uncertainty unc_3, 0.1% of positively and 3.2% of negatively correlated cells, q = 0.01). In contrast, Scn2a$^{+/-}$ mice showed a tendency towards a positive correlation bias at lower levels (level 1 learning rate α_1, 2.8% of positively and 0.0% of negatively correlated cells; level 2 uncertainty unc_2, 2.7% of positively and 0.2% of negatively correlated cells). Blocking V1 inputs to SC generally attenuated the bias of the parameter representations in

both WT and Scn2a$^{+/-}$ animals (**Fig. 6D**). This trend was also clear when we computed Pearson's correlation, ρ, between the population average of the neuronal trajectories ($\langle \Delta z_k^{\mathrm{res}} \rangle_{\mathrm{cells}}$; **Fig. 6C**) and each of the HGF parameter trajectories (**Fig. 6B**). The average of the absolute correlation values over the model parameters at each level ($\langle |\rho| \rangle_{\mathrm{params|level}}$; **Fig. 6D**, *right*) showed that SC population trajectories in WT were on average more correlated with X_3 parameters, while those in Scn2a$^{+/-}$ or after blocking V1-to-SC projections were with X_1 parameters. Furthermore, we used a cosine similarity to measure how close the HGF parameter representations in SC were between different experimental groups (see Methods for details), and found that the representations in WT and Scn2a$^{+/-}$ animals were dissimilar between each other, but became rather similar when V1 projections to SC were chemogenetically inhibited in WT mice (**Fig. 6G**).

Taken together, these results provide additional evidence that SC neurons in WT animals are involved in the computation of the overall contextual uncertainty in the visual environment by incorporating information from V1. In contrast, SC neurons in Scn2a$^{+/-}$ mice focus primarily on "lower levels" of uncertainty from recent sensory experiences due to anomalous V1 feedback signaling. This is consistent with an oversensitivity to immediate outcomes observed in ASD (Bertone et al., 2005; Baron-Cohen et al. 2009; Lawson et al., 2017), highlighting the importance of feedback signaling from V1 in properly processing visual environment context.

Discussion

Many species have evolved innate behaviors such as flight or freeze responses as essential survival strategies to evade predators (Tinbergen, 1951; LeDoux, 2012; Branco & Redgrave, 2020). Mice, for example, respond rapidly and robustly to overhead shadows as they often imply a threat, such as an aerial predator (Yilmaz and Meister, 2013; De Franceschi et al., 2016). This instinctive response is controlled by the superior colliculus (SC; Sahibzada et al., 1986; Evans et al., 2018; Shang et al., 2018), a midbrain structure that directly transforms visual signals from the

475 retina into motor commands, hence bypassing higher cognitive processing in the neocortex (Cang
476 et al., 2018). This helps animals react fast to potential threats, prioritizing survival over conscious
477 vision. This can, however, do more harm than good to animals if their response flexibility is too
478 limited, much as in autism spectrum disorder (ASD; Marco et al., 2011; Johnson et al., 2016; Jure,
479 2022). In autumn, for instance, looming shadows arise frequently from falling leaves; however,
480 they are harmless and thus animals do not need to respond. Adaptive learning allows animals to
481 tailor their behavior according to the specific context, determining properly when to react to or
482 ignore potential threats, such as looming shadows, to ensure survival without unnecessary
483 reactions.

484 In this study, we dissected the neurophysiological basis of such contextual learning in wild-
485 type (WT) and Scn2a-haploinsufficient ASD-model mice. Following the Bayesian inference
486 principle (Pellicano and Burr, 2012; Friston, 2012), we developed an implicit visual learning task
487 and examined the visual responses of SC neurons in animals engaged in the task, with the
488 assumption that SC represents the posterior belief on the stimulus while inputs from retinal
489 ganglion cells (RGCs) and the primary visual cortex (V1) reflect the likelihood and the prior belief,
490 respectively (**Fig. 1**). Consistent with this assumption, WT mice showed contextual modulations
491 in both SC (**Fig. 2**) and pupil dynamics (**Fig. 5**): the smaller the SC responses, the larger the pupil
492 area, and the more precise the posterior prediction (i.e., the narrower the posterior distribution;
493 **Fig. 1D**). Contrary to our initial hypothesis, however, we did not observe any contextual
494 modulation in RGC input to SC (**Fig. 3**). This indicates that signals from the retina do not directly
495 carry information on the visual environmental context or the likelihood. We instead suggest that
496 the feedback signals from V1 should carry both the likelihood and the prior. Indeed,
497 chemogenetically blocking V1 feedback fully eliminated the contextual modulations in SC (**Fig. 4**)
498 as well as in pupil dynamics (**Fig. 5**). This underscores a central role of the cortical feedback in
499 the adaptive responses at both neurophysiological and behavioral levels.

500 As expected from previous studies on ASD subjects (Lawson et al., 2017), Scn2a$^{+/-}$
501 animals exhibited anomalous responses to the implicit learning task at both neurophysiological
502 and behavioral levels. In particular, SC neurons demonstrated sustained responses (no
503 adaptation) in a stable environment, and an over-enhanced activity in a volatile environment (**Fig.**
504 **2**). This is consistent with our weak prior model of ASD, where we expect no change in the
505 posterior precision during the stable epochs but a reduction during the volatile epoch (**Fig. 1E**,
506 model 2). This implies that posterior computation in Scn2a$^{+/-}$ is primarily driven by changes in the
507 likelihood (sensory input variability). Notably, such an atypical modulation pattern in Scn2a$^{+/-}$ was
508 abolished by blocking V1 input (**Fig. 4**). Taken together with the results from WT, we then suggest
509 that the feedback signals from V1 have two counteracting roles in modulating SC responses (Tang
510 and Higley, 2020) – suppression linked with stability and facilitation with volatility – and their
511 imbalance underlies the anomaly in Scn2a$^{+/-}$. In WT mice, on the one hand, V1 suppresses SC
512 responses to predictable aversive stimuli in a stable environment. A milder behavioral response
513 will then follow so that animals can better evaluate the situation. When the environment becomes
514 more volatile and unpredictable, V1 then facilitates SC to regain its response. This will help trigger
515 a stronger behavioral response, prioritizing the animal's immediate safety. In Scn2a$^{+/-}$ animals, on
516 the other hand, the suppressive function of V1 feedback seems primarily impaired, leading to an
517 imbalance towards facilitation (**Fig. 2**). It has been shown that reduced expression of $Na_V1.2$
518 channels in Scn2a$^{+/-}$ affects signal integration properties of these layer 5 (L5) corticofugal
519 pyramidal cells (Spratt et al., 2019). They may then fail to properly integrate top-down cortico-
520 cortical signals at the apical dendrites that carry contextual information, especially about
521 environmental stability.

522 The above-described scenario was further supported by our hierarchical Gaussian filter
523 (HGF) model analysis (**Fig. 6**). In both WT and Scn2a$^{+/-}$ animals, HGF parameters across all levels
524 were widely represented by subpopulations of SC neurons (**Fig. 6D**). We found, however, a

representation bias at different hierarchical levels, in particular for the learning rate α: level 3 for WT, while level 1 for Scn2a$^{+/-}$ animals. This may explain why ASD individuals tend to focus more on immediate sensory inputs and associated outcomes (Lawson et al., 2017; Baron-Cohen et al. 2009; Bertone et al., 2005), as they both belong to lower-level HGF inference processes (Mathys, 2011). If this is indeed the case, resulting sensory overload and/or oversensitivity to unpredictability in ASD (Gomot et al., 2011) may be alleviated by adjusting cortical feedback to bias neural computation towards higher-level HGF inferences, hence redirecting the focus from immediate details to broader contexts.

Our results largely align with the Bayesian brain theory (Friston, 2009, 2012). It describes the brain as recursive circuits that receive feedforward and feedback inputs within a hierarchical structure, continuously updating beliefs and computing the posterior. The output is then sent upstream as the likelihood for higher-order brain structures, and downstream as a prior belief for lower-order regions. In our case, the RGC-SC-V1(L5) loop can be considered as a lower-level circuit in this recursive system, with a higher-level circuit consisting of the visual thalamus-V1(L5)-associative cortex loop (**Fig. 7**). We suggest that anomalies in Scn2a$^{+/-}$ reside in V1 L5 neurons that compute an incomplete posterior and send it to SC as a weak prior (Tang and Higley, 2020); and such dysfunction in V1 likely arises from further upstream, such as the anterior cingulate cortex whose activity has been suggested to be sensitive to the environmental volatility (Behrens et al., 2007). While we focus on a specific brain region and a specific computation involved, our findings may extend as a more generic computational pattern of ASD within complex, recursive feedforward-feedback interactions of sensory and associative systems. In future studies, it would be interesting to investigate other brain areas in a similar Bayesian framework, and also test different ASD mouse models for better understanding neurophysiological underpinnings of ASD.

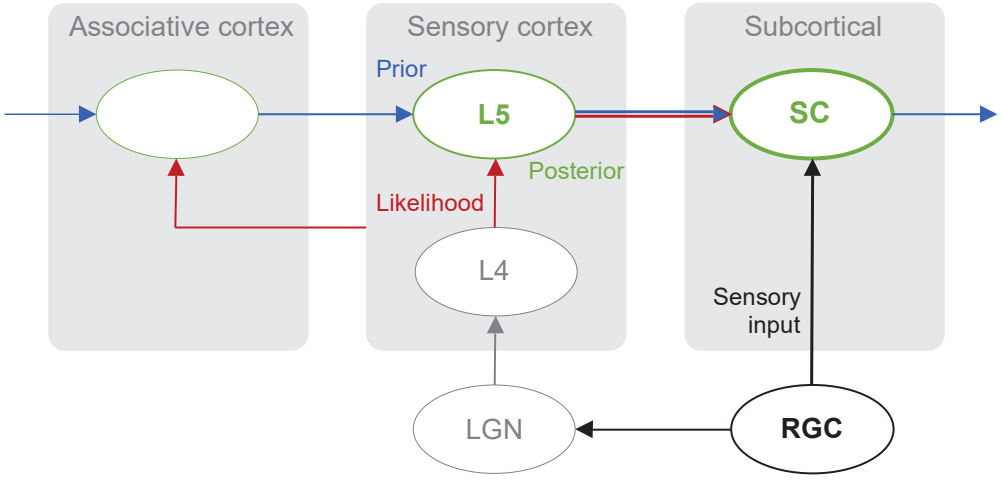


Figure 7: Circuit model for hierarchical interactions in the sensory system.
A lower-level subcortical circuit of the RGC-SC-V1(L5) loop (right) interacts with a sensory cortical circuit of the lateral geniculate nucleus (LGN)-V1-associative cortex loop (middle), which in turn interacts with a higher-level cortical circuit (left). Arrows represent axonal projections, and color indicates information from the Bayesian perspective (red, likelihood; blue, prior; green, posterior).

Methods

No statistical method was used to predetermine the sample size. We performed all animal experiments under license 233/2017-PR from the Italian Ministry of Health, following the protocol approved by the Institutional Animal Care and Use Committee at European Molecular Biology Laboratory. We conducted data analyses using Python and Matlab (Mathworks), and reported all summary statistics as mean ± standard error of the mean (s.e.m.) unless specified otherwise. We considered *p*- and *q*-values significant at $\alpha < 0.05$ *, 0.01 **, 0.001 ***, unless stated otherwise.

Animals

We used C57BL/6J mice (*Mus musculus*, both sex), aged approximately 5-10 weeks at the time of the initial surgery. We housed the mice under a 12-hour light/12-hour dark cycle and gave them *ad libitum* access to water and food. After the surgical procedures, we grouped the animals based on their operation day. We conducted the imaging experiments when the mice were between 8-16 weeks old.

567 We acquired SCN2A haploinsufficient mice as frozen sperm (Scn2a^{-}) from Riken
568 Bioresource Research Center (RBRC10243), and revitalized the line in a C57BL/6J background.
569 The homozygous mutants (Scn2a$^{-/-}$) were lethal. We used the heterozygous mice (Scn2a$^{+/-}$) as a
570 model of autism spectrum disorder (ASD; Sandlers 2012; Ben-Shalom et al., 2017), and
571 considered their homozygous littermates (Scn2a$^{+/+}$) as wild-type (WT).

572 Intracranial viral injections

573 We locally injected pseudotyped adeno-associated viruses (AAVs) into the mouse superior
574 colliculus (SC). These AAVs consisted of AAV2 *rep* and AAV9, AAV5, or AAV2-retro *cap* genes,
575 or a hybrid of AAV1 and AAV2 *cap* genes; and were used to express genetically-encoded calcium
576 indicators (GECIs; jGCaMP7f, jGCaMP8m, or jRGECO1a; **Fig. 2A**) or Cre-recombinase under
577 the pan-neuronal human synapsin (hSyn) promoter. We also injected pseudotyped AAVs,
578 composed of AAV2 *rep* and AAV8 *cap*, into the primary visual cortex (V1) to express of the
579 inhibitory designer receptor exclusively activated by designer drugs (DREADD), hM4Di-mCherry,
580 under the hSyn promoter (Dana et al., 2019; **Fig. 4A**).

581 We performed the intracranial viral injection simultaneously with the cranial implantation,
582 as described below. We created a craniotomy over the right SC or V1 and inserted an injection
583 pipette (with a tip diameter of approximately 30 μm; WPI 1B120F-3 borosilicate glass capillary
584 pulled with Zeitz DMZ puller) filled with a virus solution (ranging from approximately $5x10^{12}$ to
585 $4x10^{14}$ vg/mL in phosphate-buffered saline, PBS) through the dura. For SC, we positioned the
586 pipette using coordinates relative to Bregma, approximately [-4]-[-4.5] mm anterior-posterior (AP),
587 0.5-0.7 mm medial-lateral (ML), and slowly advanced it to a depth of about 1.25 mm. For V1, we
588 used the following coordinates: -3.5 mm AP, 1.8-2 mm ML, and 0.6 mm dorsal-ventral (DV). We
589 injected the virus solution at a rate of 1-2 nL/s for SC (0.4-0.6 μL in total) and 0.5-1 nL/s for V1
590 (0.25-0.4 μL in total) using a microinjection pump (either Neurostar NanoW or WPI NanoLiter

591 2010). After completing the injection, we waited at least 10 minutes before slowly withdrawing the
592 pipette, and then proceeded with the cranial window implantation procedure.

593 Intravitreal viral injections

594 We performed Intravitreal injections of AAV2, pseudotyped with a hybrid of AAV1 and AAV2
595 capsids, to deliver hSyn-axon-GCaMP6s expression cassette to the mouse retinal ganglion cells
596 (RGCs; Molotkov, et al., 2023; **Fig. 3A**). We exposed the scleral surface on the left eye and made
597 a small piercing with a sterile 28-30G needle in between the sclera and the cornea. We then
598 inserted an injection pipette (~50 μm tip diameter with 30-40° bevel) prefilled with a virus solution
599 (~1.5×10^{14} vg/mL in phosphate-buffered saline with 0.001% Pluronic F68 and 0.001% FastGreen)
600 into the vitreous chamber, approximately 1 mm deep. After a good sealing of the pipette was
601 formed, we injected 1.2 μL of the virus solution at a rate of 10 nL/s, using a microinjection pump.
602 We slowly withdrew the pipette at least 5 minutes after the completion of the injection, and covered
603 the treated eye with the eye ointment. We carried on the rest of the procedure and the head
604 implant as described in the previous paragraph on intracranial injections.

605 Cranial implantations

606 We performed the cranial window implantation over the mouse SC as described previously
607 (Molotkov et al., 2023). Before the surgery, we prepared a cranial window assembly by activating
608 the surface of a circular glass coverslip (5 mm diameter, 0.13-0.15 mm thickness; Assistent Karl
609 Hecht) with a laboratory corona treater (BD-20ACV Electro-Technic Products). We then fused the
610 activated glass coverslip to a cylindrical silicone plug (1.5 mm diameter, 0.75-1.00 mm height;
611 Kwik-Sil, WPI) by baking it at 70-80 °C for 24 hours.

612 During the implantation, we anesthetized the mice (induction: 4% isoflurane in oxygen;
613 maintenance: 1.5-2.0%) and positioned them in a stereotaxic apparatus (Stoelting 51625). We
614 used a heated plate (Supertech Physiological Temperature Controller) to maintain body

615 temperature and prevent hypothermia, and applied eye ointment (VitA-POS, Ursapharm) to
616 protect the eyes. We disinfected and removed the scalp (Betadine 10%, Meda Pharma), then
617 scratched and cleaned the skull surface to ensure the cement adhered well.

618 Using a high-speed surgical drill (OmniDrill35, WPI) with a 0.4 mm ball-tip carbide bur
619 (Meisinger), we made a craniotomy of approximately 3.0 mm AP by 2.5 mm ML over the right SC.
620 To prevent bleeding, we applied haemostatic sponges (Cutanplast, Mascia Brunelli) soaked in
621 sterile cortex buffer (NaCl 125 mM, KCl 5 mM, Glucose 10 mM, HEPES 10 mM, $CaCl_2$ 2 mM,
622 $MgSO_4$ 2 mM, pH 7.4) to the craniotomy site. We carefully positioned the cranial window
623 assembly, pushing the transversal sinus and posterior cortex about 0.5 mm forward, with the
624 silicone plug covering the medial-caudal region of the right SC (see, e.g., **Fig. 2A**). We fixed and
625 sealed the implant with tissue adhesive (Vetbond, 3M). We then cemented a custom-made
626 titanium headplate (0.8 mm thick) to the skull using a blend of acrylic cement powder (Paladur,
627 Kulzer) and cyanoacrylate adhesive (Loctite 401, Henkel).

628 After the surgery, the animal was allowed to recover from anesthesia in a warmed chamber
629 and then returned to its home cage. For postoperative care, we administered intraperitoneal
630 injections of 5 mg/kg Rimadyl (Zoetis) and 5 mg/kg Baytril (Bayer) to the mice daily for 3-5 days.
631 We then waited an extra 20-30 days to ensure the cranial window fully recovered before starting
632 *in vivo* two-photon imaging sessions.

633 Habituation

634 One week before the experiment began, we habituated the mice to both the experimenter and the
635 head-fixation procedure in a dark setting. We positioned the animals on a stable platform (non-
636 rotating disc; about 20 cm diameter) and let them stay still. We mounted a plexiglass tunnel over
637 the rear part of their bodies to increase comfort. We conducted these habituation sessions twice

638 a day, each lasting for 2 hours, over five to ten days. This process accustomed the mice to the
639 experimental setup and aimed to reduce any stress during the actual data collection.

640 Experimental setup

641 Two to seven days after training, we placed the animal on a stable platform with its head fixed
642 under a two-photon microscope, in front of a spherical projection screen. We used a non-rotating
643 disc (about 20 cm diameter) as the platform to minimize the effect of motor activities on neuronal
644 firing during the recordings (Schröder et al., 2020). The platform, however, allowed for vertical
645 excursions, which was compensated by a spring and a counter-weight adjustable for each animal
646 and monitored by a Hall sensor. This helped minimize head movements by preventing the mouse
647 from exerting excessive force on the head-holder with its paws. We excluded from the analysis
648 any trials if the average vertical excursion of the platform during the trial exceeded 0.25 mm. We
649 used an infrared (IR) camera (Imaging Source DMK23UX174) with IR illumination to monitor the
650 movement of the animal's left eye during recordings (**Fig. 2A**; 30 frames/s; 640-by-480 pixels).

651 Chemogenetics and histology

652 We had a subset of mice with hM4Di expressed in V1 (7 WT and 7 Scn2a$^{+/-}$; **Fig. 4;**
653 **Supplementary Fig. 2**). These animals underwent the experiment twice, with at least a two-week
654 interval. For both experiment repetitions, mice received intraperitoneal injections of either a saline
655 solution (0.9% NaCl) containing 0.2 mg/ml clozapine N-oxide (CNO), leading to a final dose of 3
656 mg/kg (administered in a volume of 0.1-0.3 ml, adjusted for mouse weight), or the saline alone.
657 We alternated the solution allocation: mice receiving saline alone in the first iteration were given
658 the CNO solution in the second, and vice versa. Starting 45-70 minutes post-injection, the mice
659 underwent the experiment for approximately one hour.

660 Once all the recording sessions were done, we histologically examined the expression of
661 hM4Di-mCherry. We anesthetized the mice terminally with a ketamine/xylazine overdose, and

662 perfused them intracardially with 4% paraformaldehyde (PFA) in PBS. We then harvested the
663 brains, incubated them for a day in 4% PFA, and for at least another day in PBS. We sliced the
664 fixed brains into coronal sections (100 μm thickness) with a cryostat (Leica CM3050S). We then
665 mounted the slices on a glass coverslip using 4′,6-diamidino-2-phenylindole (DAPI)-mount. We
666 performed fluorescent imaging with a slide scanner (VS200 automated slide scanner, Olympus)
667 with a 10X dry objective (UPLAPO 10x, NA: 0.4, X-line) across three channels: DAPI, GCaMP,
668 and mCherry (**Fig. 2B**). Only those animals that displayed clear mCherry expression (hence
669 hM4Di expression) in V1 layer 5 (L5) somata (e.g., **Fig. 4B**) were included in the analysis (**Fig.**
670 **4C,D; Supplementary Fig. 2**; "V1-block" groups).

671 Visual stimulation

672 We presented visual stimuli to the mice as described previously (Boissonnet et al., 2023). We
673 used a custom gamma-corrected digital light processing device to project images onto a spherical
674 screen placed approximately 20 cm from the left eye of the animal. The projected images covered
675 the left monocular visual field within ±22° in elevation and ±36.5° in azimuth (**Fig. 2A**). The images
676 were displayed at a resolution of 1280-by-720 pixels and a frame rate of 60 Hz.

677 One day before the experiment, mice underwent a training process. We first exposed to
678 them a gray screen for at least 10 minutes to adapt to the average luminance in a head-fixed
679 condition under the microscope. Subsequently, we presented the following visual stimuli:

680 1. Still gratings: A random sequence of black and white square wave gratings in four
681 orientations (0°, 45°, 90°, 135°) with a spatial frequency of 0.1 cycles/°. Each stimulus
682 lasted for 1 second, followed by a 0.5-second interstimulus interval (ISI). This sequence
683 was repeated 46 times in 5 minutes.
684 2. Cue-outcome stable 100%: Horizontal and vertical bars (abovementioned still gratings
685 with 0° and 90° orientations, respectively) served as cue stimuli. Horizontal bars were

followed by an omitted stimulus (prolonged ISI with a gray screen for 2 seconds), while vertical bars were followed by one out of four possible aversive stimuli (2 s): Sweep-left (a black disk 8° in diameter, appearing at 10° from the top edge and moving horizontally from the left edge to the right one at 42 °/s), Sweep-right (moving from right to left), Sweep-left-loom (a Sweep-left for the first second to bring the disk at the upper center of the screen, followed by the disk increasing in size for the next second at a speed of 57°/s while moving its center downward until reaching the horizontal center and a final size of 64° in diameter), and Sweep-right-loom (Sweep-right for the first half, followed by the loom for the second half). To avoid pupillary light reflex, the average luminance of the screen was kept constant at any time. For the sweep and loom stimuli, the background luminance was adjusted for the relative area of the black disk at every frame. We presented three blocks of 72 cue-outcome trials (approximately 5 min), where the stimulus sequence of each block was pseudo-randomized with different seeds.

3. Cue-outcome stable 83%: The cue-outcome probabilities were changed so that horizontal-omitted and vertical-aversive trials occurred with a probability of 83.4%, while horizontal-aversive and vertical-omitted trials occurred with a probability of 16.6%. We presented three blocks of 72 cue-outcome trials (approximately 5 min), where the stimulus sequence of each block was pseudo-randomized with different seeds.

On the day of the experiment, mice were again habituated to the average luminance for at least 10 minutes. We then presented the following stimuli:

1. Still gratings.
2. Cue-outcome stable 83%, but with six different types of aversive stimuli. Besides the four aversive stimuli described above, we included: Slow-loom (a black 3° disk appeared in the vertical center at the top edge and increased its size at a speed of 28.5°/s while moving its center downward until reaching the horizontal center and a final size of 64° in diameter)

711 and Fast-loom (a black 3° disk appeared in the vertical center at the top edge and stayed
712 still for 0.5 seconds; it then increased its size at a speed of 57°/s while moving its center
713 downward until reaching the horizontal center and a final size of 64° in diameter; it then
714 stayed still again for 0.5 seconds). In total, 216 trials were presented in three blocks (72
715 trials each; **Fig. 1B and Supplementary Fig. 1A**). Animals typically moved a lot at the
716 beginning and stayed still after a while. Thus, the first block was excluded from the analysis
717 to avoid any movement artifact. The second and third ones were used for the analysis as
718 early-stable (ES) and late-stable (LS) epochs, respectively (**Fig. 1C**).

719 3. Cue-outcome volatile: The cue-outcome stimulus sequences were presented with three
720 different contingencies: a given cue-outcome (horizontal-omitted and vertical-aversive)
721 probability of 16.6% for the first 24 trials; 50% for the next 24 trials; and 83.4% for the last
722 24 trials (**Fig. 1C**). These data (72 trials in total) were referred to as the volatile (VL) epoch
723 in the analysis. Collectively, the cue-outcome stimuli #2 and #3 constituted an implicit
724 visual learning task.

725 4. Random water-wave stimuli (3 minutes): These stimuli were used to generate binary
726 masks for signal source extraction in calcium image analysis (see below "Calcium image
727 analysis" section).

728 5. Randomly flickering black-and-white checkerboard stimuli (10 minutes): These stimuli
729 were used for receptive field mapping. The rectangular fields had a width of 3.7° and a
730 height of 2.9°. Each field was independently modulated by white noise at a frequency of 4
731 Hz.

732 In the cue-outcome stimuli (#2 and #3), the trials were evenly balanced within each analysis epoch
733 (72 trials each), with 36 horizontal and 36 vertical cues presented before 36 omitted and 36
734 aversive outcomes. Each of the six different aversive stimuli was presented six times within the
735 36 aversive trials. The sequences of stimuli in each epoch were pseudorandomized with different

736 seeds, but we used the same set of seeds for all mice. This ensures that each mouse was
737 presented with a new random sequence of stimuli in each epoch, but all mice were presented with
738 the same sets of sequences and cue-outcome contingencies (**Fig. 1B**).

739 For the subset of mice subject to the chemogenetics experiments (see "Chemogenetics
740 and histology" section above), we repeated the stimulus paradigm on the same experimental day
741 for CNO or saline treatment.

742 *In vivo* two-photon imaging

743 During the experiment, we kept the animal with its head fixed inside the rig for a maximum duration
744 of 2 hours (typically 2 to 5 imaging sessions per animal). We performed two-photon calcium
745 imaging using a galvo-resonant scanner (Scientifica HyperScope) and SciScan image acquisition
746 software. The microscope setup included a mode-locked Ti:sapphire tunable laser (InSight DS+,
747 Spectra-Physics) and a plan fluorite objective (CFI75 LWD 16X W, Nikon).

748 In each imaging session, we conducted single-plane time-lapse recordings with a field of
749 view approximately ranging from 0.65-by-0.65 mm to 0.43-by-0.43 mm at a depth of 115-130 μm
750 from the surface of SC. We used the excitation wavelength of 920 nm for GCaMPs, and 1,040 nm
751 for jRGECO1a. The average laser power under the objective ranged between 40-80 mW. We
752 bandpass-filtered the emitted fluorescence signal (BP 527/70 or BP 650/100 after beam-splitter
753 FF580-FDi01, Semrock) and detected it using a non-descanned gallium arsenide phosphide
754 photomultiplier tube (Hamamatsu GaAsP PMT). We acquired each frame with either 512-by-512
755 pixels (16-bit depth) at a rate of 30.9 Hz, or 1,024-by-1,024 pixels at a rate of 15.5 Hz (**Figs. 2B**
756 **and 3A**).

757 Calcium image analysis

758 To simplify manual alignment across different recordings, we first cropped the recorded images
759 from 512-by-512 pixels to 480-by-480 pixels (1.3-0.85 μm/pixel), or from 1,024-by-1,024 to 960-

760 by-960 (0.65-0.43 μm/pixel) using ImageJ software. We then applied a series of rigid and non-
761 rigid motion corrections in CaImAn (version 8 or 9; Pnevmatikakis et al., 2019).

762 To generate masks for cellular detection in CaImAn, we used either a segment of the
763 recording containing 3,000 frames from the random water-wave stimulus presentation period or
764 the entire recording frames. We divided the images into 6-by-6 patches for the constrained non-
765 negative matrix factorization (CNMF) deconvolution, resulting in a total of 36 patches. We set the
766 expected number of neurons per patch (*params.K*) to 5 for SC cell bodies, and set the size of the
767 neurons (*params.gSig*) to 5-by-5 pixels (*half size*) for 512-by-512 acquisitions, or 10-by-10 pixels
768 for 1,024-by-1,024 acquisitions. For RGC boutons, we set 8 for *params.K* and 2.5-by-2.5 pixels
769 for *params.gSig*. We manually selected putative cells that had a uniformly-filled round shape of
770 approximately 10-20 μm in size, and putative boutons to be below 2-3 μm. We then converted
771 these selected components into binary spatial masks and performed two iterations of masked
772 CNMF in CaImAn on the entire time-lapse recordings. We used the resulting spatial components
773 (*estimates.A*; e.g., **Figs. 2C,F and 3A; Supplementary Fig. 2A**) and deconvolved neural
774 activities, represented as spike estimations based on the df/f calcium transients (*estimates.S*; e.g.,
775 **Supplementary Fig. 1**), for subsequent analyses.

776 Analysis of neuronal activity contextual modulation

777 First, the estimated spiking activity of each putative neuronal or axonal component in each
778 analysis epoch was extracted from the entire time lapse and converted into z-scores (Matlab:
779 zscore): $Z(t)$. To compare the aversive stimulus response across analysis epochs, we then
780 calculated the average response across trials, $z(t) = \langle Z(t) \rangle_{\text{trials}}$, for each cell for each epoch (e.g.,
781 **Fig. 2C,F**), excluding those trials with excessive vertical platform excursion (>0.25 mm on
782 average). We then selected responsive cells for subsequent analyses that had $z(t)$ exceeding a
783 threshold in at least one analysis epoch. We varied this threshold value (8 recordings with 1, e.g.,

SFig. 1B; 22 with 1.5; and 11 with 2) to maintain a balanced representation of responsive cells across recordings (ranging from 22 to 99 cells; 41 recordings in total).

For each responsive neuron, we took an average of $z(t)$ over time during the aversive stimulus presentation period (2 s) for each analysis epoch (e.g., **Fig. 2D,G**): $\overline{z} = \langle z(t) \rangle_t$. We then subtracted the average of $\overline{z}$ across the three epochs for normalization: $\Delta z = \overline{z} - \langle \overline{z} \rangle_{\text{epochs}}$, and used this quantity Δz for the population analysis (e.g., **Fig. 2E,H**). Using this normalized activity of the responsive cells, Δz, we then performed a 3-way analysis of variance (ANOVA; Matlab: anovan) with a full interaction model for comparisons across animals. The three variable categories are: Mouse (41 mice), Experimental Group (WT, Scn2a$^{+/-}$, WT V1-block, Scn2a$^{+/-}$ V1-block), and Epochs (ES, LS, VL) with the "Mouse" nested within the "Experimental Group." For pairwise comparisons across epochs within each experimental group, we used the Tukey-Kramer tests (Matlab: multcompare; e.g., **Fig. 2E,H**).

Pupillometry

Pupil detection and eye-tracking were performed with a region-based convolutional neural network (R-CNN; He et al. 2017). The network was trained by a set of 260 mouse eye video frames with manually segmented pupil outlines (Boissonnet et al, 2023). The detected pupil was parameterized by a 2-dimensional Gaussian curve fit to the pixel map. From these parameters, we calculated the area of the pupil, $A(t)$, for each frame (6×10^3 pixel^2 = 1 mm^2; e.g., **Fig. 5A,B**). Mann-Whitney *U*-test was used to compare the median pupil area during the implicit visual task between WT and Scn2a$^{+/-}$ mice (**Fig. 5C**).

To analyze pupil size dynamics across analysis epochs, we first calculated the proportional change in pupil size upon each trial of the aversive stimulus presentation, and then took the average over trials for each epoch: $a(t) = \langle A(t)/A_0 \rangle_{\text{trials}}$, where A_0 is the average pupil size over the 10 frames before the stimulus onset for each trial (**Fig. 5D-G**). We then took the average over

the stimulation presentation period (2 s) for each analysis epoch: $\overline{a} = \langle a(t) \rangle_t$, and subtracted the mean across epochs for normalization: $\Delta a = \overline{a} - \langle \overline{a} \rangle_{\text{epochs}}$ (**Fig. 5H-K**). For pairwise comparisons of Δa across epochs at the population level, Wilcoxon signed-rank tests (Matlab: signrank) were used with false discovery rate (FDR) correction for multiple comparisons (Matlab: mafdr). We also performed 2-way ANOVA (Matlab: anovan) to make a comparison across "Experimental Group" = {WT, Scn2a$^{+/-}$, WT V1-block, Scn2a$^{+/-}$ V1-block} and "Epoch" = {ES, LS, VL} (**Fig. 5H-K**).

Hierarchical Gaussian filter model

To simulate how an ideal Bayesian observer would behave in response to the implicit visual learning task, we implemented the hierarchical Gaussian filter (HGF) model (Mathys, 2011; **Fig. 6A**) and calculated its parameter trajectories based on the 216 cue-outcome contingencies (**Fig. 6B**). The stimuli presented during the experiments followed a pseudo-random sequence that was identical for all animals as described in the "visual stimuli" section (**Fig. 1B,C**). Thus, the sequence of contingencies remained the same across the subjects. For each trial k, we assigned a binary value to the contingency: 0 for an expected outcome (vertical-aversive or horizontal-omitted) and 1 for an unexpected outcome (vertical-omitted or horizontal-aversive). These values were represented as a binary vector u that served as the sensory input for the task. From u, the posterior expectations at three levels (μ_i with $i = 1,2,3$) were calculated as follows.

- At level 1, the posterior is equal to the sensory input: $\mu_1 = u$.
- At level 2, the posterior expectation about $u = 1$ from the previous trial $(k-1)$ is: $\mu_2 = \mu_2^{(k-1)} + \sigma_2^k \delta_1^k$, where $\delta_1^k = \mu_1^k - s\left(\mu_2^{(k-1)}\right)$ is the prediction error for level 1, with $s(\cdot)$ being a sigmoid function; and $\sigma_2^k = \frac{1}{1/\hat{\sigma}_2^k + \hat{\sigma}_1^k}$ is the standard deviation or variance at level 2.

- At level 3, the posterior expectation regarding environmental volatility is: $\mu_3 = \mu_3^{(k-1)} + \sigma_3^{k^{\frac{k}{2}}} w_2^k \delta_2^k$, where the weight is given as $w_2^k = \frac{\exp\left(\kappa\mu_3^{(k-1)}+\omega_2\right)}{\sigma_2^{(k-1)}\exp\left(\kappa\mu_3^{(k-1)}+\omega_2\right)}$ and the prediction error as $\delta_2^k = \frac{\sigma_2^k+\left(\mu_2^k-\mu_2^{(k-1)}\right)^2}{\sigma_2^{(k-1)}+\exp\left(k\mu_3^{(k-1)}+\omega_2\right)} - 1$.

From these posterior expectations μ_i, we derived the prior expectations at each level $\hat{\mu}_i$ that represent the prior prediction about the probability of having $u = 1$ before the sensory input is presented.

- At level 1, the prior prediction is the sigmoid of the posterior expectation at level 2: $\hat{\mu}_1 = s(\mu_2) = \frac{1}{1+\exp(-\mu_2)}$.
- At level 2, the prior prediction about the probability that $\mu_1 = 1$ is: $\hat{\mu}_2 = \mu_2^{(k-1)}$.
- At level 3, the prior prediction about environment volatility, or probability in change of μ_2, is: $\hat{\mu}_3 = \mu_3^{(k-1)}$.

From here, the variance of the belief at every level $\hat{\sigma}_i$, or the inverse of the predicted precision $\frac{1}{\hat{\pi}_i}$, is given as follows:

- At level 1, the belief variance or imprecision of prediction that $u = 1$ is: $\hat{\sigma}_1^k = \hat{\mu}_1(1-\hat{\mu}_1)$.
- At level 2, the belief variance about the probability that $\mu_1 = 1$ is: $\hat{\sigma}_2^k = \sigma_2^{(k-1)} + \exp\left(\kappa_2\mu_3^{(k-1)} + \omega_2\right)$.
- At level 3, the belief variance about environment volatility, or probability in change of μ_2, is: $\hat{\sigma}_3^k = \sigma_3^{(k-1)} + \exp(\omega_3)$.

To calculate the precision-weighted prediction errors ϵ_i, we first calculated prediction errors at each level:

- level 1: $\delta_1^k = \mu_1^k - \hat{\mu}_1^k$,
- level 2: $\delta_2^k = \left(\left(\sigma_2^k + \left(\mu_2^k - \hat{\mu}_2^k\right)^2\right)\hat{\pi}_2^k\right) - 1$, and
- level 3: $\delta_3^k = \left(\left(\sigma_3^k + \left(\mu_3^k - \hat{\mu}_3^k\right)^2\right)\hat{\pi}_3^k\right) - 1$, where $\pi_i = 1/\sigma_i$.

Because precision weights on prediction errors are $\psi_2 = \sigma_2$ at level 2; and $\psi_3 = \pi_3/\hat{\pi}_2$ at level 3, we have: $\epsilon_2 = \psi_2\delta_1$ and $\epsilon_3 = \psi_3\delta_2$.

The uncertainties unc_i are calculated as outcome uncertainty.

- At level 1, this corresponds to the predicted variance: $unc_1^k = \hat{\sigma}_1^k$.
- At level 2, the informational uncertainty is: $unc_2^k = s\left(\mu_2^k\right)\left(1 - s\left(\mu_2^k\right)\right)\sigma_2^k$.
- At level 3, the environmental uncertainty is: $unc_3^k = \exp\left(\kappa_2\mu_3^{(k-1)} + \omega_2\right)$.

Finally, the learning rates α_i are calculated as follows:

- At level 1, $\alpha_1 = \left(s\left(\mu_2^k\right) - s\left(\mu_2^{(k-1)}\right)\right)/\left(u - s(\mu_2)\right)$.
- At level 2, learning rate is equivalent to the precision weight, or variance at level 2: $\alpha_2 = \psi_2 = \sigma_2$.
- At level 3, $\alpha_3 = \frac{\exp\left(\kappa_2\mu_3^{(k-1)} + \omega_2\right)}{2\,\psi_3}$.

In sum, the HGF model consisted of a set of these 13 parameters in five classes across three hierarchical levels **(Fig. 6B)**:

1. Learning rates: $\alpha_1, \alpha_2, \alpha_3$
2. Prior predictions: $\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3$
3. Variances of Belief: $\hat{\sigma}_2, \hat{\sigma}_3$
4. Uncertainties: $unc_1 (= \hat{\sigma}_1), unc_2, unc_3$
5. Precision-weighted prediction errors: ϵ_2, ϵ_3

After generating the model parameter trajectories from the cue-outcome contingency vector u, we eliminated those trials with non-aversive outcomes to make a fair comparison with the neuronal response trajectories that included the responses to aversive outcomes only (see below; **Fig. 6C**).

Neuronal trajectory estimation

Because we used six different types of aversive stimuli, and because the cue-outcome stimulus patterns were presented in blocks, the observed neuronal responses had a certain systematic bias that needed to be corrected before comparing them to the HGF model parameter trajectories. To this end, we normalized the neuronal responses in two steps: 1) stimulus-specific normalization and 2) non-stationarity elimination.

First, for each neuron, we calculated the average response in z-score for each aversive stimulus trial: $\overline{z}_k = \langle Z(t) \rangle_t$, and subtracted the average per stimulus type across three epochs (i.e., 18 trials in total per type): $\Delta z_{k|\text{type}} = \overline{z}_{k|\text{type}} - \langle \overline{z}_{k|\text{type}} \rangle_{\text{trials}|\text{type}}$. The resulting concatenated response Δz_k, termed as "neuronal trajectory" here, represented trial-to-trial differences of the neuronal activity, independent of the stimulus type.

These neuronal trajectories typically showed a slow decay over trials in each analysis epoch, likely due to photobleaching. To eliminate such non-stationary components, we next fitted a power function, $f(k) = a \cdot k^b + c$ (Matlab: fit "power2"), to the population average of all the neuronal trajectories in the dataset, averaged across epochs. We then subtracted a scaled decay curve, $d \cdot f(k)$, from the neuronal trajectory of each neuron. The same scalar, d, was used across epochs to avoid altering the context-dependent neuronal activity. The residual trajectory ($\Delta z_k^{\text{res}} = \Delta z_k - d \cdot f(k)$; **Fig.6C**) had a Pearson's correlation with the decay within ±0.05, and was used for the subsequent correlation analyses with the HGF model parameter trajectories.

Correlation analysis between neuronal trajectories and HGF trajectories

We used Pearson's correlation (Matlab: corr) to examine the relationship between the trajectories of neurons and the HGF model parameters. Specifically, we correlated each of the 13 HGF parameters (**Fig. 6B**) with every individual neuronal trajectory. The associated *p*-values were then converted to *q*-values with Storey's FDR methods (Matlab: mafdr) to determine the statistical significance. We then counted the fraction of SC neurons significantly correlated with each HGF parameter for each of the four experimental groups (WT, Scn2a$^{+/-}$, WT V1-block, Scn2a$^{+/-}$ V1-block; **Fig. 6D**, left panels), which we considered as the neuronal representation of the HGF model. A bias of the correlation polarity was tested with a shuffling test (10,000 repetitions, with FDR correction). The same procedure was applied to RGC axonal responses (**Fig. 6E,F**).

To compare the neural representations of HGF between the experimental groups, we used a cosine similarity index: $cos(\theta) = R_i \cdot R_j^{\mathrm{T}} / \|R_i\| \cdot \|R_j\|$, where R_i is a 1-by-13 vector for experimental group i with each element given by the difference between the fractions of the positively and negatively correlated cells with each HGF parameter (**Fig. 6G**).

At the population level, we also calculated Pearson's correlation, ρ, between the population average of the neuronal trajectories $\langle \Delta z_k^{\mathrm{res}} \rangle_{\mathrm{cells}}$ (**Fig. 6C**) and each of the HGF parameter trajectories (**Fig. 6B**). We then took the average of the absolute correlation values over the parameters at each level, $\langle |\rho| \rangle_{\mathrm{params|level}}$, to compare how each experimental group represents the HGF parameters across different levels (**Fig. 6D**, right panels).

References

Ahmadlou, M., Tafreshiha, A., Heimel, J.A. (2017). Visual Cortex Limits Pop-Out in the Superior Colliculus of Awake Mice. *Cereb. Cortex* **27**, 5772–5783.

914 Armbruster, B.N., Li, X., Pausch, M.H., Herlitze, S., Roth, B.L. (2007). Evolving the lock to fit the
915 key to create a family of G protein-coupled receptors potently activated by an inert ligand. *Proc.*
916 *Natl. Acad. Sci. U.S.A.* **104**, 5163–5168.

917 Baron-Cohen, S., Ashwin, E., Ashwin, C., Tavassoli, T., Chakrabarti, B. (2009). Talent in autism:
918 hyper-systemizing, hyper-attention to detail and sensory hypersensitivity. *Phil. Trans. R. Soc. B*
919 **364**, 1377–1383.

920 Basso, M.A., May, P.J. (2017). Circuits for Action and Cognition: A View from the Superior
921 Colliculus. *Annu. Rev. Vis. Sci.* **3**, 197–226.

922 Benavidez, N.L., Bienkowski, M.S., Zhu, M., Garcia, L.H., Fayzullina, M., Gao, L., Bowman, I.,
923 Gou, L., Khanjani, N., Cotter, K.R., Korobkova, L., Becerra, M., Cao, C., Song, M.Y., Zhang, B.,
924 Yamashita, S., Tugangui, A.J., Zingg, B., Rose, K., Lo, D., Foster, N.N., Boesen, T., Mun, H.-S.,
925 Aquino, S., Wickersham, I.R., Ascoli, G.A., Hintiryan, H., Dong, H.-W. (2021). Organization of the
926 inputs and outputs of the mouse superior colliculus. *Nat. Commun.* **12**, 4004.

927 Ben-Shalom, R., Keeshen, C.M., Berrios, K.N., An, J.Y., Sanders, S.J., Bender, K.J. (2017).
928 Opposing Effects on Na V 1.2 Function Underlie Differences Between SCN2A Variants Observed
929 in Individuals With Autism Spectrum Disorder or Infantile Seizures. *Biol. Psychiatry* **82**, 224–232.

930 Behrens, T.E.J., Woolrich, M.W., Walton, M.E., Rushworth, M.F.S. (2007). Learning the value of
931 information in an uncertain world. *Nat. Neurosci.* **10**, 1214–1221.

932 Bertone, A., Mottron, L., Jelenic, P., Faubert, J. (2005). Enhanced and diminished visuo-spatial
933 information processing in autism depends on stimulus complexity. *Brain* **128**, 2430–2441.

934 Boissonnet, T., Tripodi, M., Asari, H. (2023). Awake responses suggest inefficient dense coding
935 in the mouse retina. *eLife* **12**, e78005.

936 Branco, T., Redgrave, P. (2020). The Neural Basis of Escape Behavior in Vertebrates. *Annu. Rev.*
937 *Neurosci.* **43**, 417–439.

938 Cang, J., Savier, E., Barchini, J., Liu, X. (2018). Visual Function, Organization, and Development
939 of the Mouse Superior Colliculus. *Annu. Rev. Vis. Sci.* **4**, 239–262.

940 Chrysaitis, A.N., Seriès, P. (2023). 10 years of Bayesian theories of autism: A comprehensive
941 review. *Neurosci. Biobehav. Rev.* **145**, 105022.

Dana, H., Sun, Y., Mohar, B., Hulse, B.K., Kerlin, A.M., Hasseman, J.P., Tsegaye, G., Tsang, A., Wong, A., Patel, R., Macklin, J.J., Chen, Y., Konnerth, A., Jayaraman, V., Looger, L.L., Schreiter, E.R., Svoboda, K., Kim, D.S. (2019). High-performance calcium sensors for imaging activity in neuronal populations and microcompartments. *Nat. Methods* **16**, 649–657.

Dean, P., Redgrave, P., Westby, G.W.M. (1989). Event or emergency? Two response systems in the mammalian superior colliculus. *Trends Neurosci.* **12**, 137–147.

De Franceschi, G., Vivattanasarn, T., Saleem, A.B., Solomon, S.G. (2016). Vision Guides Selection of Freeze or Flight Defense Strategies in Mice. *Curr. Biol.* **26**, 2150–2154.

De Franceschi, Solomon, S.G. (2018) Visual response properties of neurons in the superficial layers of the superior colliculus of awake mouse. *J. Physiol.* **596**, 6307–6332.

Dobbins, I.G. (2021). Pupil dilation signals recognition salience. *Psychon. Bull. Rev.* **28**, 565–573.

Dombeck, D.A., Khabbaz, A.N., Collman, F., Adelman, T.L., Tank, D.W. (2007). Imaging large-scale neural activity with cellular resolution in awake, mobile mice. *Neuron* **56**, 43–57.

Evans, D.A., Stempel, A.V., Vale, R., Ruehle, S., Lefler, Y., Branco, T. (2018). A synaptic threshold mechanism for computing escape decisions. *Nature* **558**, 590–594.

Feinberg, E.H., Meister, M. (2015). Orientation columns in the mouse superior colliculus. *Nature* **519**, 229–232.

Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends Cogn. Sci.* **13**, 293–301.

Friston, K. (2012). Prediction, perception and agency. *Int. J. Psychophysiol.* **83**, 248–252.

Ganea, D.A., Bexter, A., Günther, M., Gardères, P.-M., Kampa, B.M., Haiss, F. (2020). Pupillary Dilations of Mice Performing a Vibrotactile Discrimination Task Reflect Task Engagement and Response Confidence. *Front. Behav. Neurosci.* **14**, 159.

Gomot, M., Blanc, R., Clery, H., Roux, S., Barthelemy, C., Bruneau, N. (2011). Candidate Electrophysiological Endophenotypes of Hyper-Reactivity to Change in Autism. *J. Autism Dev. Disord.* **41**, 705–714.

Haker, H., Schneebeli, M., Stephan, K.E. (2016). Can Bayesian Theories of Autism Spectrum Disorder Help Improve Clinical Practice? *Front Psychiatry* **7**, 107.

He, K., Gkioxari, G., Dollár, P., Girshick, R. (2018). Mask R-CNN. *arXiv*, 1703.06870.

Hoffmann, K.P., Straschill, M. (1971). Influences of cortico-tectal and intertectal connections on visual responses in the cat's superior colliculus. *Exp. Brain. Res.* **12**, 120–131.

Inayat, S., Barchini, J., Chen, H., Feng, L., Liu, X., Cang, J. (2015). Neurons in the Most Superficial Lamina of the Mouse Superior Colliculus Are Highly Selective for Stimulus Direction. *J. Neurosci.* **35**, 7992–8003.

Johnson, B.P., Lum, J.A.G., Rinehart, N.J., Fielding, J. (2016). Ocular motor disturbances in autism spectrum disorders: Systematic review and comprehensive meta-analysis. *Neurosci. Biobehav. Rev.* **69**, 260–279.

Joshi, S., Li, Y., Kalwani, R.M., Gold, J.I. (2016). Relationships between Pupil Diameter and Neuronal Activity in the Locus Coeruleus, Colliculi, and Cingulate Cortex. *Neuron* **89**, 221–234.

Jure, R. (2022). The "Primitive Brain Dysfunction" Theory of Autism: The Superior Colliculus Role. *Front. Integr. Neurosci.* **16**, 797391.

Kato, R., Takaura, K., Ikeda, T., Yoshida, M., Isa, T. (2011). Contribution of the retino-tectal pathway to visually guided saccades after lesion of the primary visual cortex in monkeys. *Eur. J. Neurosci.* **33**, 1952–1960.

Larkum, M. (2013). A cellular mechanism for cortical associations: an organizing principle for the cerebral cortex. *Trends Neurosci.* **36**, 141–151.

Lawson, R.P., Mathys, C., Rees, G. (2017). Adults with autism overestimate the volatility of the sensory environment. *Nat Neurosci.* **20**, 1293–1299.

LeDoux, J.E. (2012). Evolution of human emotion: a view through fear. *Prog. Brain. Res.* **195**, 431–442.

Li, Y.-T., Turan, Z., Meister, M. (2020). Functional Architecture of Motion Direction in the Mouse Superior Colliculus. *Curr. Biol.* **30**, 3304–3315.e4.

Loewenfeld, I.E. (1958). Mechanisms of reflex dilatation of the pupil; historical review and experimental analysis. *Doc. Ophthalmol.* **12**, 185–448.

Lomber, S.G. (2002). Learning to see the trees before the forest: Reversible deactivation of the superior colliculus during learning of local and global visual features. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 4049–4054.

Lomber, S.G., Payne, B.R., Cornwell, P. (2001). Role of the superior colliculus in analyses of space: superficial and intermediate layer contributions to visual orienting, auditory orienting, and visuospatial discriminations during unilateral and bilateral deactivations. *J. Comp. Neurol.* **441**, 44–57.

Marco, E.J., Hinkley, L.B.N., Hill, S.S., Nagarajan, S.S. (2011). Sensory Processing in Autism: A Review of Neurophysiologic Finding. *Pediatr. Res.* **69**, 48R-54R.

Mathys, C. (2011). A Bayesian foundation for individual learning under uncertainty. *Front. Hum. Neurosci.* **5**, 39.

Mathys, C.D., Lomakina, E.I., Daunizeau, J., Iglesias, S., Brodersen, K.H., Friston, K.J., Stephan, K.E. (2014). Uncertainty in perception and the Hierarchical Gaussian Filter. *Front. Hum. Neurosci.* **8**, 825.

May, P.J. (2006). The mammalian superior colliculus: laminar structure and connections. *Prog. Brain Res.* **151**, 321–378.

Molotkov, D., Ferrarese, L., Boissonnet, T., Asari, H. (2023). Topographic axonal projection at single-cell precision supports local retinotopy in the mouse superior colliculus. *bioRxiv* 2022.03.25.485790.

Ogasawara, K., McHaffie, J.G., Stein, B.E. (1984). Two visual corticotectal systems in cat. *J. Neurophysiol.* **52**, 1226–1245.

Pellicano, E., Burr, D. (2012). When the world becomes 'too real': a Bayesian explanation of autistic perception. *Trends Cogn. Sci.* **16**, 504–510.

Planells-Cases, R., Caprini, M., Zhang, J., Rockenstein, E.M., Rivera, R.R., Murre, C., Masliah, E., Montal, M. (2000). Neuronal death and perinatal lethality in voltage-gated sodium channel alpha(II)-deficient mice. *Biophys. J.* **78**, 2878–2891.

1022 Pnevmatikakis, E.A., Soudry, D., Gao, Y., Machado, T.A., Merel, J., Pfau, D., Reardon, T., Mu,
1023 Y., Lacefield, C., Yang, W., Ahrens, M., Bruno, R., Jessell, T.M., Peterka, D.S., Yuste, R.,
1024 Paninski, L. (2016). Simultaneous Denoising, Deconvolution, and Demixing of Calcium Imaging
1025 Data. *Neuron* **89**, 285–299.

1026 Ruediger, S., Scanziani, M., (2020). Learning speed and detection sensitivity controlled by distinct
1027 cortico-fugal neurons in visual cortex *eLife* **9**, e59247.

1028 Sahibzada, N., Dean, P., Redgrave, P. (1986). Movements resembling orientation or avoidance
1029 elicited by electrical stimulation of the superior colliculus in rats. *J. Neurosci.* **6**, 723–733.

1030 Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J., Willsey, A.J., Ercan-
1031 Sencicek, A.G., DiLullo, N.M., Parikshak, N.N., Stein, J.L., Walker, M.F., Ober, G.T., Teran, N.A.,
1032 Song, Y., El-Fishawy, P., Murtha, R.C., Choi, M., Overton, J.D., Bjornson, R.D., Carriero, N.J.,
1033 Meyer, K.A., Bilguvar, K., Mane, S.M., Šestan, N., Lifton, R.P., Günel, M., Roeder, K., Geschwind,
1034 D.H., Devlin, B., State, M.W. (2012). De novo mutations revealed by whole-exome sequencing
1035 are strongly associated with autism. *Nature* **485**, 237–241.

1036 Schröder, S., Steinmetz, N.A., Krumin, M., Pachitariu, M., Rizzi, M., Lagnado, L., Harris, K.D.,
1037 Carandini, M. (2020). Arousal Modulates Retinal Output. *Neuron* **107**, 487-495.e9.

1038 Shang, C., Chen, Z., Liu, A., Li, Y., Zhang, J., Qu, B., Yan, F., Zhang, Y., Liu, W., Liu, Z., Guo, X.,
1039 Li, D., Wang, Y., Cao, P. (2018). Divergent midbrain circuits orchestrate escape and freezing
1040 responses to looming stimuli in mice. *Nat. Commun.* **9**, 1232.

1041 Spratt, P.W.E., Ben-Shalom, R., Keeshen, C.M., Burke, K.J., Clarkson, R.L., Sanders, S.J.,
1042 Bender, K.J. (2019). The Autism-Associated Gene Scn2a Contributes to Dendritic Excitability and
1043 Synaptic Function in the Prefrontal Cortex. *Neuron* **103**, 673–685.e5.

1044 Sridharan, D., Steinmetz, N.A., Moore, T., Knudsen, E.I. (2017). Does the Superior Colliculus
1045 Control Perceptual Sensitivity or Choice Bias during Attention? Evidence from a Multialternative
1046 Decision Framework. *J. Neurosci.* **37**, 480–511.

1047 Suzuki, Y., Minami, T., Nakauchi, S. (2018). Association between pupil dilation and implicit
1048 processing prior to object recognition via insight. *Sci. Rep.* **8**, 6874.

1049 Tang, L., Higley, M.J. (2020). Layer 5 Circuits in V1 Differentially Control Visuomotor Behavior.
1050 *Neuron* **105**, 346-354.e5.

1051 Tinbergen, N. (1951). The study of instinct, The study of instinct. Clarendon Press/Oxford
1052 University Press, New York, NY, US.

1053 Wang, L., Sarnaik, R., Rangarajan, K., Liu, X., Cang, J. (2010). Visual Receptive Field Properties
1054 of Neurons in the Superficial Superior Colliculus of the Mouse. *J. Neurosci.* **30**, 16573–16584.

1055 Wang, C.-A., Boehnke, S.E., White, B.J., Munoz, D.P. (2012). Microstimulation of the Monkey
1056 Superior Colliculus Induces Pupil Dilation Without Evoking Saccades. *J. Neurosci.* **32**, 3629–
1057 3636.

1058 Wang, Q., Burkhalter, A. (2013). Stream-Related Preferences of Inputs to the Superior Colliculus
1059 from Areas of Dorsal and Ventral Streams of Mouse Visual Cortex. *J. Neurosci.* **33**, 1696–1705.

1060 Wei, P., Liu, N., Zhang, Z., Liu, X., Tang, Y., He, X., Wu, B., Zhou, Z., Liu, Y., Li, J., Zhang, Y.,
1061 Zhou, X., Xu, L., Chen, L., Bi, G., Hu, X., Xu, F., Wang, L. (2015). Processing of visually evoked
1062 innate fear by a non-canonical thalamic pathway. *Nat. Commun.* **6**, 6756.

1063 Yilmaz, M., Meister, M. (2013). Rapid Innate Defensive Responses of Mice to Looming Visual
1064 Stimuli. *Curr. Biol.* **23**, 2011–2015.

1065 Zhao, X., Liu, M., Cang, J. (2014). Visual Cortex Modulates the Magnitude but Not the Selectivity
1066 of Looming-Evoked Responses in the Superior Colliculus of Awake Mice. *Neuron* **84**, 202–213.

1067 Zingg, B., Chou, X., Zhang, Z., Mesik, L., Liang, F., Tao, H.W., Zhang, L.I. (2017). AAV-Mediated
1068 Anterograde Transsynaptic Tagging: Mapping Corticocollicular Input-Defined Neural Pathways
1069 for Defense Behaviors. *Neuron* **93**, 33–47.