

1 RESEARCH

2 **Bridging physiological and perceptual views of autism by means of**
3 **sampling-based Bayesian inference**

4 **Rodrigo Echeveste¹, Enzo Ferrante^{1,*}, Diego H. Milone^{1,*},**
5 **and Inés Samengo^{2,*}**

6 ¹Research Institute for Signals, Systems and Computational Intelligence sinc(i) (FICH-UNL/CONICET), 3000 Santa Fe, Argentina.

7 ²Medical Physics Department and Balseiro Institute (CNEA-UNCUYO/CONICET), 8400 Bariloche, Argentina.

8 *These authors contributed equally to this work

9 **Keywords:** Autism, Neural Circuits, Inhibitory Dysfunction, Hypopriors, Sampling-based Inference

ABSTRACT

10 Theories for autism spectrum disorder (ASD) have been formulated at different levels: ranging from
11 physiological observations to perceptual and behavioral descriptions. Understanding the physiological
12 underpinnings of perceptual traits in ASD remains a significant challenge in the field. Here we show
13 how a recurrent neural circuit model which was optimized to perform sampling-based inference and
14 displays characteristic features of cortical dynamics can help bridge this gap. The model was able to
15 establish a mechanistic link between two descriptive levels for ASD: a physiological level, in terms of
16 inhibitory dysfunction, neural variability and oscillations, and a perceptual level, in terms of
17 hypopriors in Bayesian computations. We took two parallel paths: inducing hypopriors in the
18 probabilistic model, and an inhibitory dysfunction in the network model, which lead to consistent
19 results in terms of the represented posteriors, providing support for the view that both descriptions
20 might constitute two sides of the same coin.

AUTHOR SUMMARY

21 Two different views of autism, one regarding altered probabilistic computations, and one regarding
22 inhibitory dysfunction, are brought together by means of a recurrent neural network model trained to
23 perform sampling-based inference in a visual setting. Moreover, the model captures a variety of
24 experimental observations regarding differences in neural variability and oscillations in subjects with
25 autism. By linking neural connectivity, dynamics and function, this work contributes to the
26 understanding of the physiological underpinnings of perceptual traits in autism spectrum disorder.

INTRODUCTION

27 Autism spectrum disorder (ASD) refers to a complex neurodevelopmental condition involving
28 persistent challenges in social interaction and communicative skills, and restricted/repetitive behaviors
29 (Association, 2013). While some recent studies suggest that ASD could be detected during the first year
30 of life in some children, early signs seem to be non-specific, with group differences more robustly
31 found after children's first birthday (see Ozonoff, Heung, Byrd, Hansen, and Hertz-Picciotto (2008) for a
32 review).

33 Almost two decades ago, John Rubenstein and Michael Merzenich suggested that many of the
34 symptoms related to ASD might reflect an abnormal ratio between excitation and inhibition leading to
35 hyper-excitability of cortical circuits in ASD subjects (Rubenstein & Merzenich, 2003). Since then, a
36 variety of studies have linked reduced inhibitory signaling in the brain with ASD symptoms, either
37 observing how behavior typically associated with ASD emerges in animals when inhibitory pathways
38 are altered, or measuring gamma-aminobutyric acid (GABA) concentration or GABA receptors in
39 several brain regions (see Cellot and Cherubini (2014) for a detailed review). Further support for this
40 view comes from the fact that ASD patients suffer from epilepsy with a prevalence up to 25 times that
41 of the neurotypical population (Bolton et al., 2011).

42 Establishing a direct link between ASD and impaired inhibition in specific circuits in humans has not
43 been easy. Indeed, two recent in-vivo studies in humans have shown puzzling results (Horder et al.,
44 2018; Robertson, Ratai, & Kanwisher, 2016). In these studies inhibition was assessed both behaviorally

(in visual tasks where inhibition is widely believed to play a key role in neurotypical behavior) and by measuring either GABA concentration (Robertson et al., 2016) or number of GABA receptors (Horder et al., 2018) in the brains of ASD and control subjects. Interestingly, while ASD subjects showed a marked deficit in binocular rivalry, characteristic of a disruption in inhibitory signaling, GABA concentrations in the visual cortex were normal (Robertson et al., 2016). However, while GABA concentration was predictive of rivalry dynamics in controls, the same was not true within the ASD population, evidencing a disruption of inhibitory action. Similarly, while ASD subjects show an altered performance in the paradoxical motion perception task (a proxy measure of GABA signaling), GABA receptor availability in the brain of those participants showed no significant difference from controls (Horder et al., 2018). Both studies suggest an impairment in inhibitory signaling which cannot be explained by coarse differences in GABA concentration or receptor availability at the level of brain areas, and which might affect specific circuits instead. To complicate matters further, there is evidence for not only inhibitory but also excitatory dysfunction in ASD, and it has been hypothesized that homeostatic principles might be the reason behind this seemingly contradictory result (Nelson & Valakh, 2015). The idea being that if, for instance inhibition is reduced, excitatory synapses might be then adjusted to try to compensate for the overall change in neural activity that reduction would ensue. Computational modeling of local cortical circuits expressed in terms of excitation and inhibition might therefore provide a fruitful avenue of research to guide future experiments.

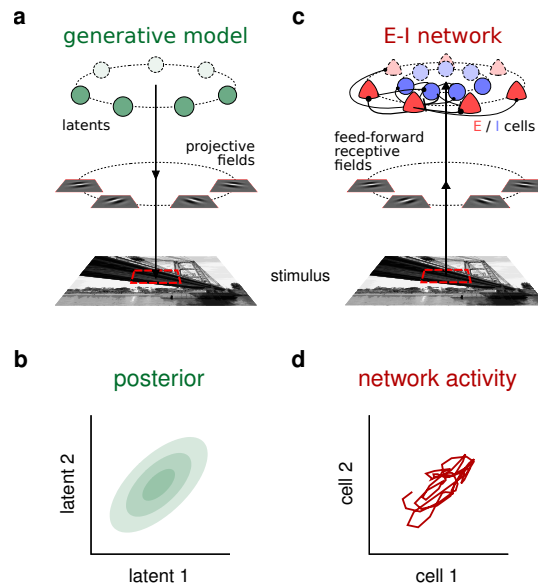
From the point of view of perception in ASD, a variety of theories have been put forward over the last two decades. Highly influential descriptive theories include: the weak central coherence theory (Happé & Frith, 2006) and the enhanced perceptual functioning theory (Mottron, Dawson, Soulières, Hubert, & Burack, 2006). Here we will focus on computational accounts of perception in ASD, and in particular on a Bayesian view of perception (Palmer, Lawson, & Hohwy, 2017). We will later also make connections to another influential computational theory formulated in terms of predictive coding (Van Boxtel & Lu, 2013; Van de Cruys et al., 2014).

Within the Bayesian framework, inference about the external world proceeds by multiplicatively combining pre-existent knowledge (expressed in terms of a *prior* probability distribution) and current sensory evidence (represented in terms of a *likelihood* function), to form a *posterior* distribution which encapsulates our belief about the state of the world after having observed a given stimulus (Knill &

74 Richards, 1996). Rather than expressing that belief as a single point estimate of what is most probable,
75 the posterior distribution provides a richer description, naturally incorporating the associated
76 uncertainty which remains after the observation. A growing body of evidence indicates that, at least in
77 some settings, the brain is able to operate with probability distributions in this way to perform
78 approximate Bayesian inference (see Fiser, Berkes, Orbán, and Lengyel (2010), for a review). In recent
79 years it has been proposed that in ASD subjects these forms of Bayesian computations are carried out
80 abnormally: overweighting sensory evidence with respect to prior information (Palmer et al., 2017;
81 Pellicano & Burr, 2012). Concretely, the authors in Pellicano and Burr (2012) proposed that this is a
82 consequence of chronically attenuated priors (termed *hypopriors*), characterized by broader
83 distributions (i.e. higher uncertainty).

84 The related theoretical framework of predictive coding proposes that the cortex is organized
85 following a circuit motif where feedback connections from higher- to lower-order sensory areas signal
86 predictions of lower-level responses, while feedforward connections signal errors between predictions
87 and actually observed lower-level responses (Rao & Ballard, 1999). Proponents of predictive coding
88 theories have rightfully pointed out that Bayesian theories by themselves (without specifying a
89 concrete implementation) do not offer a mechanistic explanation for ASD perception (Van Boxtel & Lu,
90 2013), which is key to understand how physiological observations may be linked to perceptual and
91 behavioral traits in ASD subjects. As has been observed by Aitchison and Lengyel (2017), Bayesian
92 inference and predictive coding are not necessarily mutually exclusive: predictive coding can be seen
93 as a computational motif which can implement several computational goals (one of which is Bayesian
94 inference), while Bayesian inference can be seen as a computational objective which can have several
95 implementations (one of which is predictive coding). Moreover, as noted in the aforementioned review,
96 telling apart the use of a Bayesian predictive coding scheme from a direct variable code in an empirical
97 setting is no trivial matter. Strong transient overshoots at stimulus onset, for instance, which are a
98 typical signature of predictive coding, can also emerge in direct variable coding schemes (Aitchison &
99 Lengyel, 2016; Echeveste, Aitchison, Hennequin, & Lengyel, 2020). Indeed, while weighting predictive
100 errors more strongly by increasing synaptic gains in the motif could explain sensory hypersensitivity
101 in ASD subjects (Palmer et al., 2017), a competing explanation can be provided within a direct variable
102 coding scheme, as we show in the present study. We note however that while predictive coding

103 schemes can incorporate gamma oscillations (Bastos et al., 2012), it is not clear how they would account
 104 for the contrast-dependent frequency modulation of these oscillations (Roberts et al., 2013), or the
 105 stimulus-dependent modulations of neural variability (Churchland et al., 2010; Orbán, Berkes, Fiser, &
 106 Lengyel, 2016).



107 **Figure 1. Sketches of the generative model, and a neural circuit implementing sampling-based probabilistic inference under that model.**

108 **a**, The Gaussian scale mixture (GSM) generative model. Under this model, each image patch is built as a linear combination of local features (projective
 109 fields), whose intensities are drawn from a multivariate Gaussian distribution. This linear combination is then further scaled by a global contrast level
 110 and subject to noise. The features were in this case a set of localized oriented Gabor filters which differed only in their orientations and were uniformly
 111 spread between -90° and 90° . The image serving as stimulus in the figure is for illustration only. Photo Credit: Santa Fe Bridge by Enzo Ferrante
 112 <https://eferrante.github.io/> **b**, 2D projection of the posterior distribution for a given a visual stimulus as computed by the Bayesian ideal
 113 observer under the GSM. **c**, The recurrent E-I neural network receives an image patch as an input, which is filtered by feedforward receptive fields matching
 114 the projective fields of GSM in **a**. Each latent variable in the GSM is represented by the activity of one E cell in the network. **d**, 2D projection of the neural
 115 responses of E cells corresponding the same 2 latent variables shown in **b**. Over time, the network samples from posterior distribution corresponding to the
 116 stimulus it receives.

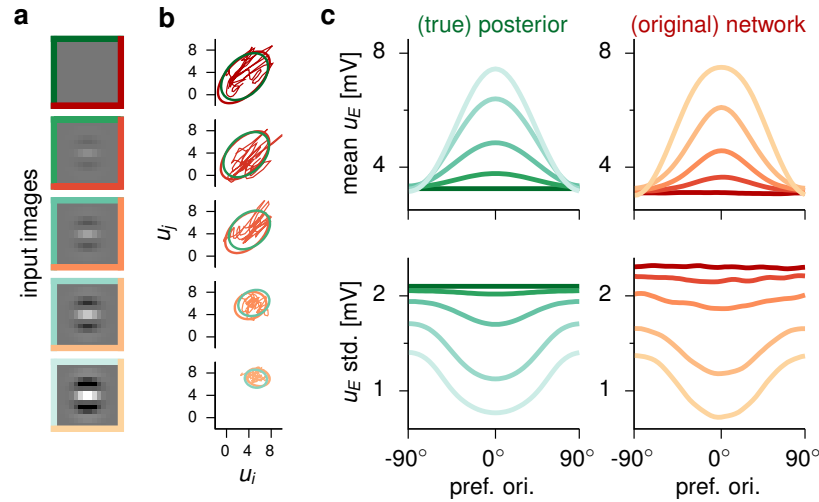
117 A popular implementation choice for probabilistic inference is that of probabilistic population codes
 118 (PPCs) (Ma, Beck, Latham, & Pouget, 2006), where the posterior distribution is encoded in the average
 119 rates of a population of neurons. This framework has been used in the past to link inhibitory deficits

120 and Bayesian computations in an artificial neural network model consisting of two feed-forward layers
 121 followed by a stage of divisive normalization (Rosenberg, Patterson, & Angelaki, 2015). In this work, a
 122 probabilistic version of the model was constructed to capture the “oblique effect”. This term describes
 123 the fact that neurotypical subjects tend to be more sensitive to cardinal than to oblique orientations in a
 124 visual orientation discrimination task (Westheimer & Beard, 1998). Indeed, a modulation of the divisive
 125 normalization factor in this model was shown to account for the observed reduction of the oblique
 126 effect in ASD subjects (Dickinson, Jones, & Milne, 2014). The standard PPC framework requires
 127 constant Fano factors (no variability modulation) (Ma et al., 2006), and furthermore feed-forward
 128 network implementations can only capture mean rate responses, but fail to account for the dynamical
 129 properties of neural responses that arise from recurrent connectivity. It is hence unclear in this
 130 framework how altered neural variability observed in the ASD population (Haigh, Heeger, Dinstein,
 131 Minschew, & Behrmann, 2015; Milne, 2011) and gamma oscillations (van Diessen, Senders, Jansen,
 132 Boersma, & Bruining, 2015) would relate to probabilistic computations in these subjects.

133 Sampling-based theories for probabilistic inference offer an alternative mechanistic implementation
 134 for Bayesian inference. Within this framework, neural circuits represent posterior distributions by
 135 drawing samples over time from those distributions (Berkes, Orbán, Lengyel, & Fiser, 2011; Haefner,
 136 Berkes, & Fiser, 2016). Interestingly, sampling-based models for probabilistic inference have recently
 137 begun to establish direct links between cortical dynamics and perception (Echeveste et al., 2020). A
 138 neural circuit model of a cortical hypercolumn respecting Dale’s principle and performing fast
 139 sampling-based inference in a visual task displayed a suite of features which are typically observed in
 140 cortical recordings across species and experimental conditions. The network showed highly variable
 141 responses with strong inhibition-dominated transients at stimulus onset, and stimulus-dependent
 142 gamma oscillations, as observed in the cortex (Haider, Häusser, & Carandini, 2013; Ray & Maunsell,
 143 2010; Roberts et al., 2013). The model further evidenced stimulus-dependent variability modulations
 144 consistent with experimental findings (Roberts et al., 2013). Divisive normalization of mean responses
 145 (Carandini & Heeger, 2012) was also shown to emerge in this network as a result of its recurrent
 146 dynamics. This is interesting since divisive normalization was precisely the starting point for the
 147 probabilistic model in Rosenberg et al. (2015), and in previous work linking uncertainty and neural
 148 variability via gain modulation (Hénaff, Boundy-Singer, Meding, Ziemba, & Goris, 2020). The

149 computational and dynamical properties of the network make it a viable candidate to test the link
150 between Bayesian computations and several physiological features observed in ASD such as inhibitory
151 dysfunction, as well as differences in neural variability and oscillations.

152 In what follows we will firstly set the basis for this work by recapitulating some of the key findings
153 of Echeveste et al. (2020), relating probabilistic inference, and dynamics in a network model which we
154 will take to describe healthy control subjects. We will then make use of the connection between
155 perception and physiology established by this model and take two parallel routes to explore two
156 different theories for autism: a perceptual theory expressed in terms of hypopriors, and a physiological
157 theory concerning impaired inhibition. The first path will involve modifying the probabilistic model
158 under which perception takes place, and more concretely its prior, and observing the consequences of
159 that choice in terms of the observer's posteriors. The second path will involve inducing an inhibitory
160 deficit in the neural network whose job is to sample from the corresponding posteriors, and analyzing
161 the effect of that modification in the posteriors represented by the network. We will then compare the
162 results of both approaches to determine to what extent these two seemingly unrelated theories are
163 compatible. Finally, we show that the induced inhibitory deficit in the network model produces
164 changes in the variability and dynamics of the network. We will evaluate these changes in the context
165 of empirical observations in ASD subjects and other theoretical accounts for ASD. These include an
166 increase in neural variability, as well as an increase in the power and frequency of gamma oscillations.
167 The network also becomes hypersensitive to intense stimuli, displaying stronger transient responses
168 at stimulus onset.



169 **Figure 2. Inference under the GSM and responses in the original network, here representing healthy neurotypical subjects.** Replotted from
 170 Echeveste et al. (2020). In all panels shades of green correspond to the ideal observer, while red corresponds to network responses, as in Figure 1. Line colors
 171 in **b** and frame colors in **d** indicate different contrast levels, which are the same as stimulus frames in **a**, indicating to which stimulus responses correspond.
 172 **a**, Stimuli (shade of frame color indicates contrast level, split green, blue and red indicates that the same stimuli were used as input to the ideal observer
 173 and to both neural networks). **b**, Covariance ellipses (2 standard deviations) of the ideal observer’s posterior distributions (green) and of the networks’
 174 corresponding response distributions (red). Red trajectories show sample 500 ms-sequences of activities in the networks. As in the sketch of fig. Figure 1, 2D
 175 projections corresponding to two representative latent variables / excitatory cells are shown. These two correspond to projective fields / receptive fields at
 176 preferred orientations 42° and 16° . **c**, Mean (top) and standard deviation (bottom) of latent variable intensities ordered by each latent’s orientation, for each
 177 stimulus in the training set. Left: from the ideal observer’s posterior distribution (green). Right: E cell membrane potentials u_E from the networks’ stationary
 178 distributions (red). **d**, Comparison of correlation matrices. Left: for the ideal observer’s posterior distributions (in green). Right: for the networks’ stationary
 179 response distributions (red). Response moments in **c** and **d** were estimated from $n = 20,000$ independent samples (taken 200 ms apart). Correlations in **d**
 180 are Pearson’s correlations.

RESULTS

181 *Bayesian inference of visual features implemented by a recurrent E-I neural circuit*

182 The starting point for perceptual inference within the Bayesian framework is a probabilistic model that
183 describes one's assumptions about how observed stimuli relate to variables of interest in the outside
184 world. This forward model is usually referred to as a *generative model*, and the role of an ideal Bayesian
185 observer is to invert this probabilistic relationship to obtain posterior distributions over those variables
186 of interest given the observed stimulus. The generative model employed here is a Gaussian
187 Scale-Mixture model (GSM, see Figure 1 a and Methods and Materials), which has been shown to
188 capture the statistics of natural images at the level of small image patches (Wainwright & Simoncelli,
189 2000). Importantly, inference under this model had already been shown to explain features of behavior
190 and stationary response distributions in neural data in visual perception (Coen-Cagli, Kohn, &
191 Schwartz, 2015; Orbán et al., 2016; Schwartz, Sejnowski, & Dayan, 2009). Under this version of the
192 GSM, natural image patches are constructed as linear combinations of Gabor filters of different
193 orientations, which are then scaled by a global contrast variable. The goal of the inference process was
194 to estimate the probability distribution of the intensity with which each Gabor filter (each orientation)
195 participated in the observed image. In turn, in order to model cortical neural dynamics, a common
196 recurrent neural network model is employed: the stabilized supralinear network (SSN, see Figure 1 b
197 and Methods and Materials) (Ahmadian, Rubin, & Miller, 2013; Hennequin, Ahmadian, Rubin, Lengyel,
198 & Miller, 2018). Neurons in the network were arranged around a ring, according to their preferred
199 orientation, under the approximation of visual inference problem being rotationally symmetric (though
200 see Discussion). Moreover, neurons in the network respected Dale's principle, with two separate
201 populations for excitatory (E) and inhibitory (I) cells. The SSN thus formulated was then optimized
202 using current machine learning methods to approximate a Bayesian ideal observer under the GSM:
203 when the network receives an image patch as its input, it produces samples over time with its neural
204 activity so as to represent the corresponding posterior distribution (Figure 1 c–d). Examples of the
205 image patches used to train the network, as well as sample neural trajectories are presented in
206 Figure 2 a–b, respectively. After training, posterior distributions sampled by network responses match
207 those prescribed by the ideal observer (see Figure 2 c, cf. green and red). Once trained, the SSN model
208 thus establishes a mechanistic link between neural dynamics in terms of an E-I circuit and perception

209 formulated as sampling-based probabilistic inference. In what follows we exploit this link to take two
 210 complementary paths: inducing simple perturbations to the GSM to induce hypopriors, and to the SSN
 211 to induce an inhibitory dysfunction.

212 *Perturbing the generative model: the effect of hypopriors*

213 To illustrate and generate intuitions on the effect of hypopriors, we begin by employing a simplified
 214 one-dimensional toy example (Methods and Materials). Let us assume the “true” prior, correctly
 215 describing the statistics of the world concerning a particular inference process, is a zero-mean
 216 Gaussian. Let us further assume for this toy example that the likelihood is also a Gaussian function
 217 whose precision is modulated by a contrast variable which expresses the degree of reliability of the
 218 sensory stimulus. If we vary the stimulus contrast we can compute a posterior distribution for each
 219 stimulus under this true prior (Figure 3 a – b, in green). If, however, we were to employ a hypoprior,
 220 that is a prior with a higher variance, we would obtain posterior distributions which overweight
 221 sensory evidence, in the sense that they more closely resemble the likelihood function (both in mean
 222 and variance) than they should. This in turn results in a higher posterior mean and in higher
 223 uncertainty about the estimate (Figure 3 b, cf. green and blue lines).

224 Let us now turn to the GSM. Also in this case, a global contrast variable regulates the reliability of
 225 the stimulus. However, in contrast to the 1D toy example presented before, inference in this case takes
 226 place in a higher dimensional space. We again modify the prior distribution to induce a hypoprior. We
 227 do so in the simplest possible way, by scaling the prior co-variance matrix by a constant factor larger
 228 than 1.0 (Methods and Materials). In Figure 3 c we compare the posterior distributions calculated under
 229 the true prior (in green) with those computed under the hypoprior (in blue). As expected, we again find
 230 that hypopriors result in overweighting of sensory stimuli, with higher posterior means and higher
 231 uncertainty about the estimates (Figure 3 d, cf. green and blue lines), consistently with the postulates of
 232 Pellicano and Burr (2012).

247 *Perturbing the network: the effect of inhibitory deficits*

248 We now turn our attention to the network model. In what follows we will refer to the original SSN
 249 presented in Figure 2, as the *neurotypical* (NT) network. As previously stated, the NT-network was

250 constructed in terms of separate excitatory and inhibitory populations. Here we target inhibitory
 251 connections by scaling down their efficacy by a global constant value (Methods and Materials). In order
 252 to ensure that baseline activity levels are not affected, and following the ideas of Nelson and Valakh

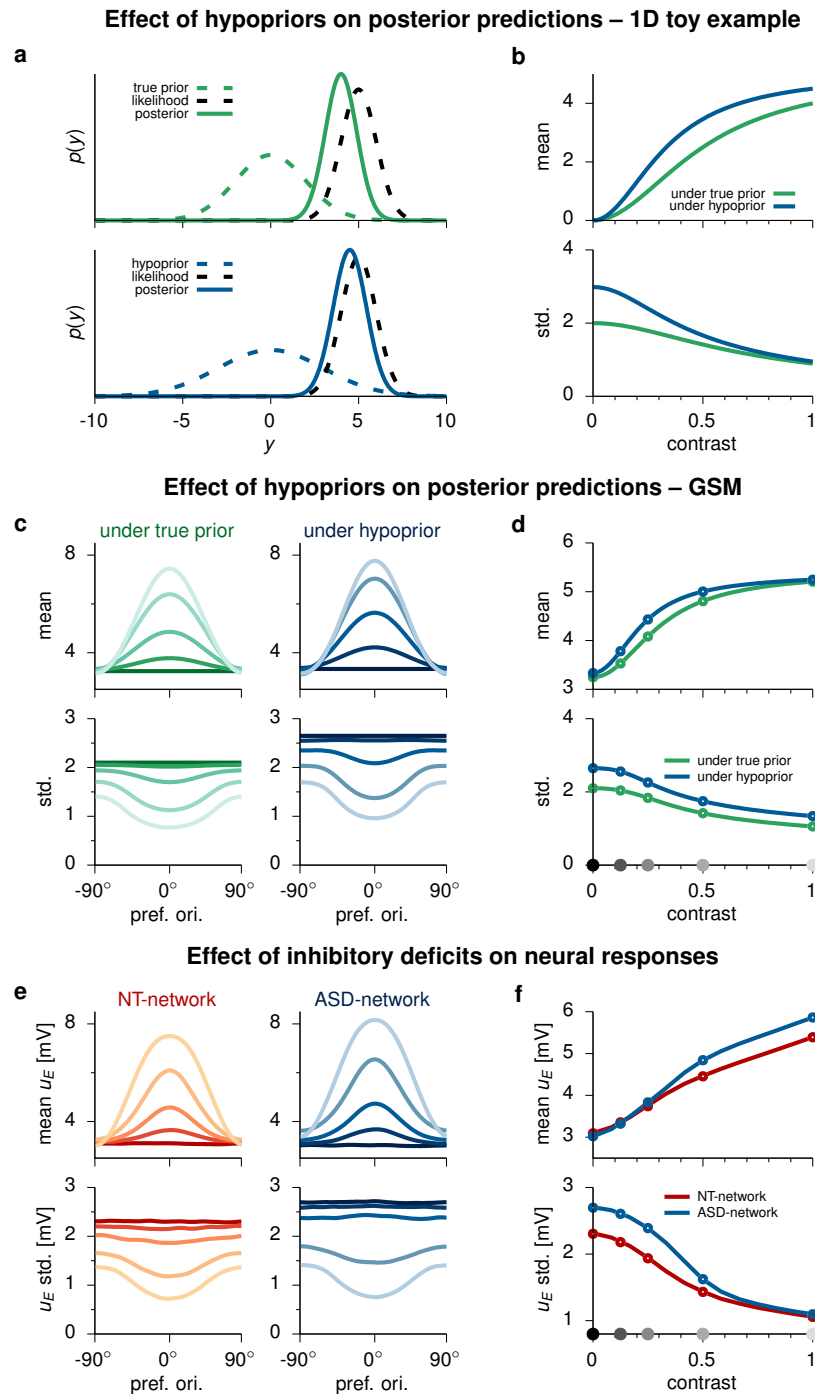


Figure 3. Hypopriors and impaired inhibition. (Continues on next page)

Figure 3. Hypopriors and impaired inhibition. **a–b** : Effect of hypopriors on posterior predictions for a 1D toy example. Priors, likelihoods and posteriors are all Gaussian. A contrast variable regulating the likelihood precision plays the role of the perceptual reliability of stimuli. Two example inference cases are presented: under the true (well-calibrated) prior (dashed, green) and under a wider hypoprior (dashed, blue). **a** The prior (dashed, color) and likelihood (dashed, black) are multiplicatively combined according to Bayes’ rule to form the posterior (continuous, color). **b** Posterior mean (top plot) and standard deviation (bottom plot) under the true prior (green) and the hypoprior (blue), as a function of contrast (likelihood precision). **c–d** : Effect of hypopriors on posterior predictions for the full multivariate GSM model. **c** Mean (top plots) and standard deviation (bottom plots) of latent variable intensities ordered by each latent’s orientation, for each stimulus in Figure 2. Left: for the well calibrated ideal observer’s posterior distribution (green). Right: under a hypoprior (blue). **d** Posterior mean (Top) and standard deviation (Bottom), averaged across all latent variables, under the true prior (green) and the hypoprior (blue), as a function of contrast. **e–f** : Effect of impaired inhibition on network responses. **e**, Mean (top) and standard deviation (bottom) of latent variable intensities ordered by each latent’s orientation, for each stimulus in the training set. **e** cell membrane potentials u_E from the stationary response distributions for the NT-network (Left, red), and for the ASD-network (Right, blue). **f** Mean (Top) and standard deviation (Bottom) of neural responses, averaged across all cells, for the NT-network (red) and the ASD-network (blue), as a function of contrast. Circles, and gray dots on x-axis of panels **d** and **f** indicate training contrast levels.

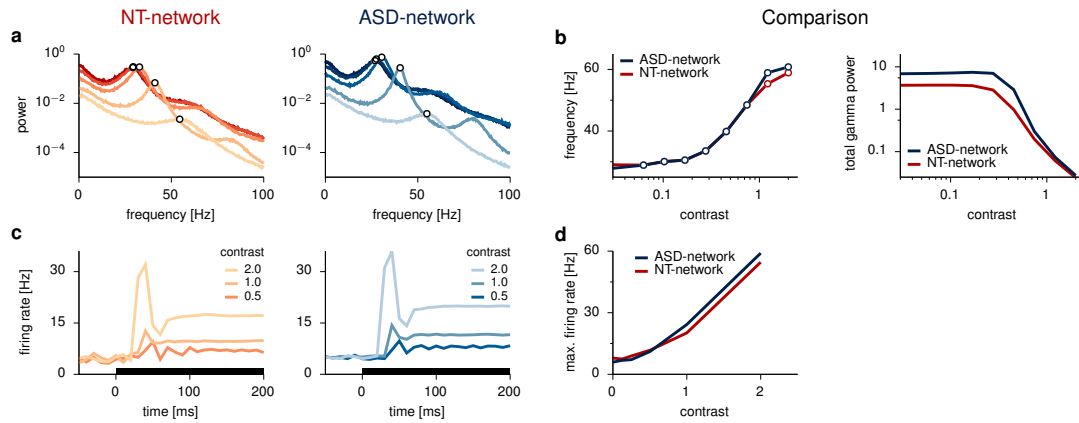
(2015), we also scaled excitatory connections globally in a homeostatic fashion (see Supplementary Fig. 1 and Methods and Materials). We will henceforth refer to the network where inhibitory deficits have been induced as the ASD-network. As we did for the generative model, we then compared the mean and standard deviation of the posterior distributions encoded by both networks in terms of their response samples (Figure 3 e – f). Notably, we observed that ASD-network representations of the posteriors also seemed to overweight current sensory information. Indeed, posterior means were higher in the ASD- than in the NT-network (Figure 3 f top panel, cf. red and blue lines). In passing, we note that because of the original approximate inference scheme, the scaling of the mean and standard deviation with contrast between the original network and the posterior are similar but not identical. In particular, while mean responses in the generative model saturate at high contrasts, they only decelerate in the network model, without actually saturating. Indeed, responses in this type of network models do not saturate. They either continue to grow or ‘bounce back’ and begin to decrease (Ahmadian et al., 2013). Similarly, a slightly higher standard deviation is observed in the network with respect to the posterior at low contrast, which stems from an underestimation of the variance of neural responses under the Gaussian approximation during training of the network (Echeveste et al., 2020).

268 Higher uncertainty about the estimates was also found in the network (Figure 3 f bottom panel, cf.
269 red and blue lines), just as it happened for the generative model under hypopriors (compare Figure 3
270 panels d and f). Interestingly, we have reached the same qualitative traits by two very different
271 approaches and following two theories expressed at widely different levels: one perceptual, one
272 physiological.

273 It is important to note that sampling-based implementations of Bayesian inference establish a direct
274 link between uncertainty and neural variability, since the width of the posterior distribution is directly
275 related to the amount of variability. Indeed we observe that weaker inhibition leads to higher
276 variability in the neural responses of the ASD-network compared to the NT-network (Figure 3 f, bottom
277 panel, cf. red and blue lines), as had been suggested in Rubenstein and Merzenich (2003), where the
278 point had been made that a disruption of E-I balance leading to a hyperexcitable cortex would lead to
279 increased cortical ‘noise’. Indeed, higher neural variability has been experimentally reported in ASD
280 subjects both in EEG (Milne, 2011) and fMRI (Haigh et al., 2015) studies.

281 An advantage of employing a neural network model such as the SSN, which shows characteristic
282 features of cortical dynamics, such as gamma oscillations and transient overshoots (including their
283 contrast dependence), is that we can also explore the predictions the model makes for these features,
284 now for the ASD-network.

294 Firstly, we look at gamma oscillations. To that end we computed the power spectrum from the local
295 field potential (LFP), from which we extracted the peak gamma frequency for different contrast levels
296 for both networks (Figure 4 a). We note that the overall frequency modulation is very similar in both
297 networks, with slightly higher peak gamma frequency in the ASD-network for high contrast stimuli
298 (cf. Figure 4 b, left panel, red and blue). Previous work has reported higher peak gamma frequency in
299 ASD subjects solving a visual task, which was interpreted as a sign of “increased neural inhibition”
300 (Dickinson, Bruyns-Haylett, Smith, Jones, & Milne, 2016). At first glance, this might seem at odds with
301 the starting point for our work where we have weakened inhibitory synapses. It is worth noting
302 however that total inputs (both E and I) result in a balanced recurrent network from a dynamic
303 equilibrium, which may result in higher inhibitory currents, despite weaker inhibitory synapses. This is
304 precisely the case here (see Supplementary Fig. 1 d). Indeed, it has been known for decades that
305 balanced networks are prone to so-called “paradoxical effects” (Tsodyks, Skaggs, Sejnowski, &



285 **Figure 4. Transient responses and oscillations.** **a**, LFP power as a function of frequency for stimuli of different contrast levels (same stimuli and
 286 colors as in Figure 3) in the NT-network (left), and in the ASD-network (right). Both networks present strong gamma oscillations (see peaks in the gamma
 287 band, indicated by empty circles). **b**, Comparison of oscillatory behavior in both networks. On the left, the peak gamma frequency is presented as a function
 288 of stimulus contrast for both networks. Very minimal differences are observed. On the right, the total power within the gamma band is presented as a
 289 function of contrast for both networks. A higher gamma power is observed for the ASD network at all contrasts, with strong differences at low contrasts.
 290 **c**, Across-trial average transient responses for stimuli of different contrast levels in the neurotypical network (left) and in the ASD network (right). Both
 291 networks present strong stimulus dependent transient overshoots. **d**, Comparison of overshoot sizes. The maximal firing rate is presented as a function
 292 of stimulus contrast for both networks. We observe that the ASD network presents stronger peak responses at higher contrasts, over-reacting to intense
 293 stimuli. NT-network results reproduced from Echeveste et al. (2020).

306 McNaughton, 1997), whereby direct external inhibitory inputs to I cells, can actually lead to increased I
 307 rates. This also hints at why seemingly contradictory results are often found regarding inhibition in
 308 ASD depending on what exactly is chosen as a measure of inhibition.

309 Interestingly however, gamma power is higher for the ASD-network (see sharper gamma peaks in
 310 the spectra of Figure 4 a, and in Figure 4 b, right plot, blue vs red). An insight into the functional
 311 interpretation of this effect can be obtained from analyzing neural responses at zero contrast,
 312 representing what is usually termed spontaneous activity in the literature. In sampling based models,
 313 such as this one, spontaneous activity is postulated to encode this prior distribution (Berkes et al.,
 314 2011). Indeed, when the stimulus is completely uninformative, as is the case at zero contrast, the
 315 posterior matches the prior. The model hence predicts higher gamma power in spontaneous activity,

316 which is in line with previous reports of higher gamma band power in resting state activity of ASD
317 subjects (van Diessen et al., 2015).

318 We finally turn our attention to transient responses. We compared the ASD- and NT-networks in
319 terms of their trial-averaged firing rates around stimulus onset (Figure 4 c). The model predicts higher
320 maximal firing rates (and not only mean rates) for the ASD network than for the NT network at
321 intermediate and high contrasts (cf. Figure 4 d, red and blue), indicating that the ASD-network has
322 become hypersensitive to intense stimuli. We note that theories of perception expressed in terms of
323 predictive coding usually interpret peak rates as a measure of surprise, novelty or unexpectedness (Rao
324 & Ballard, 1999), and indeed a predictive coding account of ASD perceptual traits, including abnormal
325 sensory sensitivity, has been postulated by several authors in the past (Van Boxtel & Lu, 2013; Van de
326 Cruys et al., 2014). Results from the ASD network, which we here interpret from a Bayesian inference
327 perspective, are then not inconsistent with a predictive coding view of perceptual differences in the
328 ASD population.

DISCUSSION

329 Neural neural network models are increasingly being used as a tool to study how differences in neural
330 architectures may be linked to symptoms in different disorders (Lanillos et al., 2020). In this work we
331 have employed a neural network model of a V1 cortical hypercolumn trained to perform
332 sampling-based probabilistic inference in a visual task to build a mechanistic bridge between
333 descriptions of ASD formulated at two very different levels: a physiological level (in terms of inhibitory
334 dysfunction (Rubenstein & Merzenich, 2003), neural variability (Haigh et al., 2015; Milne, 2011), and
335 gamma oscillations (van Diessen et al., 2015)), and a perceptual level (in terms of hypopriors in Bayesian
336 computations (Pellicano & Burr, 2012)). In what follows we describe merits of this work, limitations
337 and open questions.

Merits

339 We have taken two parallel paths: in one perturbing the probabilistic generative model in order to
340 induce hypopriors, and in the other perturbing the neural network model to induce an inhibitory
341 dysfunction. We observed that both approaches lead to consistent results in terms of the represented

342 posterior distributions, providing support for the possibility that both views of ASD might actually
343 constitute two sides of the same coin.

344 Employing a neural network model such as the SSN, which not only performs inference in a
345 perceptual task but also displays characteristic features of cortical dynamics while doing so (Echeveste
346 et al., 2020), allowed us to make further connections between characteristic differences in these
347 dynamics and inhibitory dysfunction in ASD subjects. Stimulus-dependent variability modulations in
348 the network, and concretely the direct link between neural variability and uncertainty established by
349 sampling-based implementations of inference, predicted higher variability in neural responses in the
350 ASD- vs the NT-network. Indeed increased neural variability has been reported in ASD subjects both in
351 EEG (Milne, 2011) and fMRI (Haigh et al., 2015) studies. Moreover, transient overshoots, usually
352 interpreted in predictive coding theories to represent novelty, surprise or unexpectedness (Rao &
353 Ballard, 1999), are present in the network, with higher responses for strong stimuli in the ASD-network
354 vs the NT-network, indicating an oversensitivity to intense stimuli, a feature often reported in children
355 with ASD (Kern et al., 2006).

356 Furthermore, oscillations in the ASD-network displayed higher gamma-band oscillatory power,
357 consistent with observations in resting-state EEG recordings of ASD subjects (van Diessen et al., 2015).
358 Peak gamma frequencies were also higher in the ASD network for high-contrast stimuli, a fact which
359 has indeed been observed in EEG recordings from subjects performing an orientation discrimination
360 task (Dickinson et al., 2016), and which had been attributed to increased inhibition. We confirmed that,
361 despite having decreased the efficacy of inhibitory synapses in our network, mean inhibitory inputs
362 were indeed actually larger for high-contrast stimuli. This observation is in line with the known fact
363 that balanced E-I networks are prone to “paradoxical effects” regarding inhibition (Tsodyks et al., 1997),
364 where average rates result from a dynamic balance of excitation and inhibition, and might explain
365 apparent contradictions between studies reporting increased/decreased inhibition (Cellot & Cherubini,
366 2014; Dickinson et al., 2016). These results also highlight the importance of neural network simulations
367 to assist in the interpretation of physiological observations regarding the role of inhibition in cortical
368 recordings.

369 *Limitations and open questions*

370 Training recurrent neural networks with expansive non-linearities beyond mean responses is currently
371 a challenging and computationally expensive task. These networks are prone to instabilities and
372 current optimization for second-order moments requires either a large number of trials, or
373 matrix-matrix operations which scale as n^3 in the number of neurons (Hennequin & Lengyel, 2016).
374 Indeed, the choice of the simple generative model played a key role in order to make the training
375 problem tractable with currently available optimization techniques, but imposes some limitations. The
376 GSM produces multivariate Gaussian posteriors (which enabled training the network with currently
377 available second-order moment-matching methods), and was further constructed to be rotationally
378 symmetric (which drastically reduced the number of network parameters to be optimized, as well as
379 the required number of training examples). A model constructed in this way, will however not be able
380 to capture features of human behavior in popular tests of visual perception, such as the “oblique effect”,
381 where neurotypical subjects seem to be more sensitive to cardinal orientations (Westheimer & Beard,
382 1998), an effect which is reduced in ASD subjects (Dickinson et al., 2014). Tackling problems like these
383 in a sampling-based setting will require developing tools to train more flexible networks that can
384 produce richer posterior distributions. It should be noted that these limitations are however of a
385 technical nature, and are not inherent to the sampling-based inference framework.

386 Secondly, the model employed to explain simple, low-level perceptual computations was constructed
387 in terms of a single V1 hypercolumn, and is hence only able to capture local dynamical features, such as
388 locally generated gamma oscillations. Hypothetically, the ideas presented here can be extended to the
389 representation of other circular variables beyond orientation of visual stimuli, such as head direction in
390 rodents Skaggs, Knierim, Kudrimoti, and McNaughton (1995), motor intent in primates Georgopoulos,
391 Taira, and Lukashin (1993), physical space in grid cells McNaughton, Battaglia, Jensen, Moser, and
392 Moser (2006), or oculomotor control Seung (1998). In all these examples, highly specialized brain areas
393 receive assorted inputs that carry a noisy, filtered and distributed representation of a circular variable.
394 The recurrent activity of the network constitutes a mechanistic implementation of an inference process,
395 which could be potentially executed through a sampling-based Bayesian inference strategy, as explored
396 here. If that were the case, the strong reliance of ASD subjects on the likelihood could also be
397 broadened beyond the realm of sensory processing. Extensions of these ideas are also conceivable to
398 other one-dimensional, yet aperiodic, domains, such as sound pitch Aronov, Nevers, and Tank (2017),

399 navigation speed Kropff, Carmichael, Moser, and Moser (2015), or elapsed time Tsao et al. (2018) which,
 400 although still fairly narrow in their semantic content, involve some degree of higher-level processing.
 401 However, as we progress into still higher cognitive functions, the understanding of how
 402 context-dependent modulations of cortical dynamics emerge during complex perceptual tasks will
 403 likely require models where multiple circuits interact (Simon & Wallace, 2016). In this sense,
 404 hierarchical or spatially extended versions of the SSN model employed here may provide adequate
 405 substrates to study inference of higher level perceptual tasks where longer-range aspects of cortical
 406 dynamics, such as gamma synchronization, might emerge.

407 Thirdly, we have focused on one aspect of probabilistic inference: inferring the state of a set of latent
 408 variables under perceptual uncertainty. The study of other aspects of this problem, such as inferring
 409 temporal transitions (Sinha et al., 2014), or causal relationships (Noel, Shivkumar, Dokka, Haefner, &
 410 Angelaki, 2021), and their link to altered inhibition and neural dynamics, will require the use of
 411 different architectures and generative models and constitute worthwhile avenues of future research.

412 *Closing remarks*

413 We have shown how recurrent neural networks optimized for sampling-based inference are viable
 414 candidates to bridge the gap between Bayesian perceptual theories of ASD and their physiological
 415 underpinnings in terms of inhibitory dysfunction, neural variability and oscillations. We believe these
 416 results highlight the potential for the use of the emerging body of function-optimized neural networks
 417 (Echeveste et al., 2020; Hennequin, Vogels, & Gerstner, 2014; Orhan & Ma, 2017; Remington, Narain,
 418 Hosseini, & Jazayeri, 2018; Song, Yang, & Wang, 2016; Yamins et al., 2014) as models to establish
 419 mechanistic links between neural activity and computations in the cortex that go beyond the study of
 420 neurotypical perception.

METHODS

421 In order to link cortical dynamics and probabilistic computations we modified the parameters of the
 422 probabilistic and network models employed in Echeveste et al. (2020). In what follows we describe
 423 those changes and refer the reader to the original paper for a more detailed description of the models
 424 and of the original model parameters.

425 ***The generative model***

426 In this work the Gaussian scale mixture model (GSM, Wainwright and Simoncelli (2000)), is employed
 427 as the generative model of natural images (at the level of small patches) under which inference is
 428 carried out in the primary visual cortex (V1, Coen-Cagli et al. (2015); Orbán et al. (2016)). Under the
 429 GSM an image patch \mathbf{x} is obtained by linearly combining a number of local features (given by the
 430 columns of a matrix \mathbf{A}), which are weighted by a corresponding number of feature coefficients given by
 431 \mathbf{y} , further scaled by a single contrast variable z , and finally corrupted by additive white Gaussian noise.
 432 This forward generative model can then be summarized in terms of the likelihood function given by

$$\mathbf{x}|\mathbf{y}, z \sim \mathcal{N}(z \mathbf{A} \mathbf{y}, \sigma_x^2 \mathbf{I}), \quad (1)$$

433 together with the priors for the feature coefficients and the contrast variable z . Local features were
 434 assumed to be drawn from a multivariate Gaussian:

$$\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}), \quad (2)$$

435 and the contrast was assumed to be drawn from a Gamma prior. To induce hypopriors we modified the
 436 overall scale of the prior covariance matrix \mathbf{C} , by taking $\mathbf{C}_{\text{HP}} = \alpha_{\text{HP}} \mathbf{C}$, with $\alpha_{\text{HP}} = 1.5$. Other values
 437 were explored without qualitative differences (not shown). We note that taking $\alpha_{\text{HP}} > 1$ results in
 438 wider priors, as required for a hypoprior.

439 The 1D toy example model of Figure 3a–b, corresponds to a 1-dimensional GSM with prior variance
 440 $C = 4$, $A = 10$, and $\sigma_x^2 = 100$. As in the full GSM, we took $\alpha_{\text{HP}} = 1.5$.

441 ***Network dynamics and architecture***

442 The circuit model consisted of a nonlinear, stochastic network respecting Dale’s principle, with N_E
 443 excitatory and N_I inhibitory neurons. The evolution of the membrane potential u_i of each neuron i in
 444 this model is described by (Hennequin et al., 2018)

$$\tau_i \frac{du_i}{dt} = -u_i(t) + h_i(t) + \sum_j W_{ij} r_j(t) + \eta_i(t), \quad (3)$$

445 where τ_i represents the membrane time constant for neuron i , h_i its feedforward input, and η_i is the
 446 process noise (capturing both intrinsic and extrinsic forms of neural variability). \mathbf{W} is the matrix of

447 recurrent connections, and hence W_{ij} represents the strength of the synapse connecting neuron j to
 448 neuron i . As previously mentioned, the network is non-linear, with firing rates

$$r_i(t) = k[u_i(t)]^m. \quad (4)$$

449 Here k and m represent the scale and exponent of the firing rate nonlinearity (Ahmadian et al., 2013).
 450 Given the rotational symmetry of the problem, \mathbf{W} itself was parametrized to be rotationally symmetric.
 451 Neurons in the model are arranged in a ring of pairs of E and I cells according to their preferred
 452 orientations (Figure 1c) where W_{ij} was a smoothly decaying function of the tuning difference between
 453 neurons i and j (see Supplementary Fig. 1 a, top and second row). The (stimulus-independent) process
 454 noise covariance was analogously parametrized (see Supplementary Fig. 1 a, third row). Following
 455 canonical models of V1 simple cells (Dayan & Abbott, 2001), feedforward inputs to the network were
 456 computed by applying a linear filter \mathbf{W}^{ff} to the stimulus (the image patch) followed by a nonlinearity
 457 (see Supplementary Fig. 1 a, bottom row).

458 The perturbation here employed to induce an inhibitory deficit has a single free parameter δ_I which
 459 scales the inhibitory columns of \mathbf{W} , $\mathbf{W}_I^{\text{ASD}} = (1 - \delta_I)\mathbf{W}_I^{\text{NT}}$ (see Supplementary Fig. 1 a-b). In order to
 460 maintain the baseline level of activity, a second modification is introduced (simulating homeostatic
 461 adaption of the excitatory connections), scaling the excitatory columns of \mathbf{W} by a factor δ_E :
 462 $\mathbf{W}_E^{\text{ASD}} = (1 - \delta_E)\mathbf{W}_E^{\text{NT}}$. This second factor was found by grid-search minimization of the homeostatic
 463 cost

$$\mathcal{C}_h = |\mu_s^{\text{NT}} - \mu_s^{\text{ASD}}|, \quad (5)$$

464 capturing the change in mean spontaneous activity levels (μ_s) between the original NT- and perturbed
 465 ASD-network. This adaptation procedure returns a single δ_E value for each δ_I value (Supplementary
 466 Fig. 1 c). We note that excitatory changes via this procedure resulted always smaller than inhibitory
 467 ones (cf. to identity line in Supplementary Fig. 1 c, bottom plot). Network results presented throughout
 468 this paper correspond to $\delta_I = 0.1$, for which $\delta_E = 0.076$. Numerical experiments were repeated for
 469 $\delta_I = 0.05$ and $\delta_I = 0.15$ without qualitative differences (not shown).

470 *Numerical simulations*

471 Stationary moments of neural responses to a fixed input (Figure 3e) were computed from 20,000
 472 independent samples (200 ms apart) generated by letting neural activity in the network evolve over
 473 time via Equation 3 (excluding transients). Power spectra in Figure 4 a were obtained from simulated
 474 local field potentials (LFPs), computed as the average (across-cells) membrane potential. Gamma peak
 475 frequencies in Figure 4 b (left) were obtained as the local maximum in the spectrum within the gamma
 476 range (20–80 Hz), while total gamma power in Figure 4 b (right) was computed as the integral of the
 477 spectrum over that same range.

478 Transient responses displayed in Figure 4 c were computed as the mean (across E-cells and trials)
 479 firing rates ($n = 100$), which are then further averaged over a 10-ms sliding window. A random delay
 480 time (sampled from a truncated Gaussian, with a mean of 45 ms and a standard deviation of 5 ms) was
 481 employed for the feedforward input to each pair of E–I cells. These procedures had been put in place to
 482 allow for a comparison to experimental data, and are here kept in order to compare the ASD-network
 483 to replotted results from the original (here NT-) network. Maximal firing rates in Figure 4 d were
 484 obtained as the peak rates from transient firing rate responses.

485 *Code availability.*

486 The (Python) code to create the ASD network is provided in
 487 `bitbucket.org/RSE.1987/inhibitory_dysfunction`. The code for the numerical
 488 experiments can be found at:
 489 `bitbucket.org/RSE.1987/ssn_inference_numerical_experiments`.

ACKNOWLEDGMENTS

490 This work was supported by Argentina’s National Scientific and Technical Research Council
 491 (CONICET), who covered all researchers salaries. We are grateful to Y. Nagai for pointing out this
 492 potential avenue of research after discussing previous work.

TECHNICAL TERMS

493 **Latent variable** A variable of interest to which an observer has no direct access and hence needs to
 494 infer it from an observation of other related variables.

495 **Prior** Probability distribution encapsulating an observer’s knowledge about the latent variables
 496 before observing the stimulus.

497 **Likelihood function** Function describing the conditional probability of an observation for each
 498 state of the latent variables.

499 **Posterior** Conditional probability over the latent variables after observing a given stimulus.

500 **Hypoprior** A chronically attenuated prior, whose uncertainty is higher than implied by the statistics
 501 of stimuli.

502 **GABA** Main inhibitory neurotransmitter.

503 **Gamma Oscillations** Rhythmic patterns of activity with a frequency between 20 and 80Hz.

504 **Transient overshoot** Excursion in neural responses that exceeds mean responses over a brief
 505 period of time after the onset of the stimulus.

506 **Divisive normalization** Process by which the responses of single neurons are divisively modulated
 507 by the responses of other neurons.

508
 509 **REFERENCES**
 510

511 Ahmadian, Y., Rubin, D., & Miller, K. (2013). Analysis of the stabilized supralinear network. *Neural Computation*, 25(8),
 512 1994–2037.

513 Aitchison, L., & Lengyel, M. (2016). The Hamiltonian brain: efficient probabilistic inference with excitatory-inhibitory
 514 neural circuit dynamics. *PLoS computational biology*, 12(12), e1005186.

515 Aitchison, L., & Lengyel, M. (2017). With or without you: predictive coding and bayesian inference in the brain. *Current*
 516 *opinion in neurobiology*, 46, 219–227.

517 Aronov, D., Nevers, R., & Tank, D. W. (2017). Mapping of a non-spatial dimension by the hippocampal-entorhinal circuit.
 518 *Nature*, 543(7647), 719-722.

519 Association, A. P. (2013). *Diagnostic and statistical manual of mental disorders (dsm-5®)*. American Psychiatric Pub.

- 520 Bastos, A., Usrey, W., Adams, R., Mangun, G., Fries, P., & Friston, K. (2012). Canonical microcircuits for predictive coding.
521 *Neuron*, 76(4), 695–711.
- 522 Berkes, P., Orbán, G., Lengyel, M., & Fiser, J. (2011). Spontaneous cortical activity reveals hallmarks of an optimal internal
523 model of the environment. *Science*, 331(6013), 83–87.
- 524 Bolton, P. F., Carcani-Rathwell, I., Hutton, J., Goode, S., Howlin, P., & Rutter, M. (2011). Epilepsy in autism: features and
525 correlates. *The British Journal of Psychiatry*, 198(4), 289–294.
- 526 Carandini, M., & Heeger, D. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, 13(1),
527 51.
- 528 Cellot, G., & Cherubini, E. (2014). Gabaergic signaling as therapeutic target for autism spectrum disorders. *Frontiers in*
529 *pediatrics*, 2, 70.
- 530 Churchland, M. M., Yu, B. M., Cunningham, J. P., Sugrue, L. P., Cohen, M. R., Corrado, G. S., ... Shenoy, K. V. (2010).
531 Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nature Neuroscience*, 13(3), 369.
- 532 Coen-Cagli, R., Kohn, A., & Schwartz, O. (2015). Flexible gating of contextual influences in natural vision. *Nature*
533 *Neuroscience*.
- 534 Dayan, P., & Abbott, L. (2001). *Theoretical neuroscience* (Vol. 806). Cambridge, MA: MIT Press.
- 535 Dickinson, A., Bruyns-Haylett, M., Smith, R., Jones, M., & Milne, E. (2016). Superior orientation discrimination and
536 increased peak gamma frequency in autism spectrum conditions. *Journal of abnormal psychology*, 125(3), 412.
- 537 Dickinson, A., Jones, M., & Milne, E. (2014). Oblique orientation discrimination thresholds are superior in those with a
538 high level of autistic traits. *Journal of autism and developmental disorders*, 44(11), 2844–2850.
- 539 Echeveste, R., Aitchison, L., Hennequin, G., & Lengyel, M. (2020). Cortical-like dynamics in recurrent circuits optimized
540 for sampling-based probabilistic inference. *Nature Neuroscience*, 23(9), 1138–1149.
- 541 Fiser, J., Berkes, P., Orbán, G., & Lengyel, M. (2010). Statistically optimal perception and learning: from behavior to neural
542 representations. *Trends in Cognitive Sciences*, 14(3), 119–130.
- 543 Georgopoulos, A. P., Taira, M., & Lukashin, A. (1993). Cognitive neurophysiology of the motor cortex. *Science*, 260(5104),
544 47–52.

- 545 Haefner, R., Berkes, P., & Fiser, J. (2016). Perceptual decision-making as probabilistic inference by neural sampling.
546 *Neuron*, 90(3), 649–660.
- 547 Haider, B., Häusser, M., & Carandini, M. (2013). Inhibition dominates sensory responses in the awake cortex. *Nature*,
548 493(7430), 97–100.
- 549 Haigh, S. M., Heeger, D. J., Dinstein, I., Minshew, N., & Behrmann, M. (2015). Cortical variability in the sensory-evoked
550 response in autism. *Journal of autism and developmental disorders*, 45(5), 1176–1190.
- 551 Happé, F., & Frith, U. (2006). The weak coherence account: detail-focused cognitive style in autism spectrum disorders.
552 *Journal of autism and developmental disorders*, 36(1), 5–25.
- 553 Hénaff, O. J., Boundy-Singer, Z. M., Meding, K., Ziemba, C. M., & Goris, R. L. (2020). Representation of visual uncertainty
554 through neural gain variability. *Nature communications*, 11(1), 1–12.
- 555 Hennequin, G., Ahmadian, Y., Rubin, D., Lengyel, M., & Miller, K. (2018). The dynamical regime of sensory cortex: stable
556 dynamics around a single stimulus-tuned attractor account for patterns of noise variability. *Neuron*, 98(4), 846–860.
- 557 Hennequin, G., & Lengyel, M. (2016). Characterizing variability in nonlinear recurrent neuronal networks. *arXiv preprint*
558 *arXiv:1610.03110*.
- 559 Hennequin, G., Vogels, T., & Gerstner, W. (2014). Optimal control of transient dynamics in balanced networks supports
560 generation of complex movements. *Neuron*, 82(6), 1394–1406.
- 561 Horder, J., Andersson, M., Mendez, M. A., Singh, N., Ämma Tangen, Lundberg, J., ... Borg, J. (2018). Gaba_A receptor
562 availability is not altered in adults with autism spectrum disorder or in mouse models. *Science Translational Medicine*,
563 10(461).
- 564 Kern, J. K., Trivedi, M. H., Garver, C. R., Grannemann, B. D., Andrews, A. A., Savla, J. S., ... Schroeder, J. L. (2006). The
565 pattern of sensory processing abnormalities in autism. *Autism*, 10(5), 480–494.
- 566 Knill, D., & Richards, W. (1996). *Perception as Bayesian inference*. Cambridge University Press.
- 567 Kropff, E., Carmichael, J. E., Moser, M.-B., & Moser, E. I. (2015). Speed cells in the medial entorhinal cortex. *Nature*,
568 523(7561), 419–424.

- 569 Lanillos, P., Oliva, D., Philippsen, A., Yamashita, Y., Nagai, Y., & Cheng, G. (2020). A review on neural network models of
570 schizophrenia and autism spectrum disorder. *Neural Networks*, 122, 338–363.
- 571 Ma, W., Beck, J., Latham, P., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature*
572 *Neuroscience*, 9(11), 1432–1438.
- 573 McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I., & Moser, M.-B. (2006). Path integration and the neural basis of
574 the “cognitive map”. *Nature Reviews in Neuroscience*, 7(8), 663–678.
- 575 Milne, E. (2011). Increased intra-participant variability in children with autistic spectrum disorders: evidence from
576 single-trial analysis of evoked eeg. *Frontiers in psychology*, 2, 51.
- 577 Mottron, L., Dawson, M., Soulières, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: an
578 update, and eight principles of autistic perception. *Journal of autism and developmental disorders*, 36(1), 27–43.
- 579 Nelson, S. B., & Valakh, V. (2015). Excitatory/inhibitory balance and circuit homeostasis in autism spectrum disorders.
580 *Neuron*, 87(4), 684–698.
- 581 Noel, J.-P., Shivkumar, S., Dokka, K., Haefner, R., & Angelaki, D. (2021). Aberrant causal inference and presence of a
582 compensatory mechanism in autism spectrum disorder. *PsyArXiv*.
- 583 Orbán, G., Berkes, P., Fiser, J., & Lengyel, M. (2016). Neural variability and sampling-based probabilistic representations in
584 the visual cortex. *Neuron*, 92(2), 530–543.
- 585 Orhan, A., & Ma, W. (2017). Efficient probabilistic inference in generic neural networks trained with non-probabilistic
586 feedback. *Nature communications*, 8(1), 138.
- 587 Ozonoff, S., Heung, K., Byrd, R., Hansen, R., & Hertz-Picciotto, I. (2008). The onset of autism: patterns of symptom
588 emergence in the first years of life. *Autism research*, 1(6), 320–328.
- 589 Palmer, C. J., Lawson, R. P., & Hohwy, J. (2017). Bayesian approaches to autism: Towards volatility, action, and behavior.
590 *Psychological bulletin*, 143(5), 521.
- 591 Pellicano, E., & Burr, D. (2012). When the world becomes ‘too real’: a bayesian explanation of autistic perception. *Trends*
592 *in cognitive sciences*, 16(10), 504–510.

- 593 Rao, R., & Ballard, D. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical
594 receptive-field effects. *Nature Neuroscience*, 2(1), 79.
- 595 Ray, S., & Maunsell, J. H. (2010). Differences in gamma frequencies across visual cortex restrict their possible use in
596 computation. *Neuron*, 67(5), 885–896.
- 597 Remington, E., Narain, D., Hosseini, E., & Jazayeri, M. (2018). Flexible sensorimotor computations through rapid
598 reconfiguration of cortical dynamics. *Neuron*, 98(5), 1005–1019.
- 599 Roberts, M., Lowet, E., Brunet, N., Ter Wal, M., Tiesinga, P., Fries, P., & De Weerd, P. (2013). Robust gamma coherence
600 between macaque V1 and V2 by dynamic frequency matching. *Neuron*, 78(3), 523–536.
- 601 Robertson, C. E., Ratai, E.-M., & Kanwisher, N. (2016). Reduced gabaergic action in the autistic brain. *Current Biology*,
602 26(1), 80–85.
- 603 Rosenberg, A., Patterson, J. S., & Angelaki, D. E. (2015). A computational perspective on autism. *Proceedings of the*
604 *National Academy of Sciences*, 112(30), 9158–9165.
- 605 Rubenstein, J., & Merzenich, M. M. (2003). Model of autism: increased ratio of excitation/inhibition in key neural systems.
606 *Genes, Brain and Behavior*, 2(5), 255–267.
- 607 Schwartz, O., Sejnowski, T., & Dayan, P. (2009). Perceptual organization in the tilt illusion. *Journal of Vision*, 9(4), 19–19.
- 608 Seung, S. (1998). Cognitive neurophysiology of the motor cortex. *Neural Networks*, 11(7-8), 1253-1258.
- 609 Simon, D. M., & Wallace, M. T. (2016). Dysfunction of sensory oscillations in autism spectrum disorder. *Neuroscience &*
610 *Biobehavioral Reviews*, 68, 848–861.
- 611 Sinha, P., Kjelgaard, M. M., Gandhi, T. K., Tsourides, K., Cardinaux, A. L., Pantazis, D., ... Held, R. M. (2014). Autism as a
612 disorder of prediction. *Proceedings of the National Academy of Sciences*, 111(42), 15220–15225.
- 613 Skaggs, W. E., Knierim, J. J., Kudrimoti, H. S., & McNaughton, B. L. (1995). A model of the neural basis of the rat's sense of
614 direction. *Advances in Neural Information Processing Systems*, 7, 173-180.
- 615 Song, H., Yang, G., & Wang, X. (2016). Training excitatory-inhibitory recurrent neural networks for cognitive tasks: A
616 simple and flexible framework. *PLoS computational biology*, 12(2), e1004792.

- 617 Tsao, A., Sugar, J., Lu, L., Wang, C., Knierim, J. J., Moser, M. B., & Moser, E. I. (2018). Integrating time from experience in
618 the lateral entorhinal cortex. *Nature*, 561(7721), 57-62.
- 619 Tsodyks, M. V., Skaggs, W. E., Sejnowski, T. J., & McNaughton, B. L. (1997). Paradoxical effects of external modulation of
620 inhibitory interneurons. *Journal of neuroscience*, 17(11), 4382–4388.
- 621 Van Boxtel, J. J., & Lu, H. (2013). A predictive coding perspective on autism spectrum disorders. *Frontiers in psychology*, 4,
622 19.
- 623 Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de Wit, L., & Wagemans, J. (2014). Precise minds in
624 uncertain worlds: predictive coding in autism. *Psychological review*, 121(4), 649.
- 625 van Diessen, E., Senders, J., Jansen, F. E., Boersma, M., & Bruining, H. (2015). Increased power of resting-state gamma
626 oscillations in autism spectrum disorder detected by routine electroencephalography. *European archives of psychiatry
627 and clinical neuroscience*, 265(6), 537–540.
- 628 Wainwright, M., & Simoncelli, E. (2000). Scale mixtures of Gaussians and the statistics of natural images. In *Advances in
629 Neural Information Processing Systems* (pp. 855–861).
- 630 Westheimer, G., & Beard, B. L. (1998). Orientation dependency for foveal line stimuli: detection and intensity
631 discrimination, resolution, orientation discrimination and vernier acuity. *Vision research*, 38(8), 1097–1103.
- 632 Yamins, D., Hong, H., Cadieu, C., Solomon, E., Seibert, D., & DiCarlo, J. (2014). Performance-optimized hierarchical models
633 predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, 111(23), 8619–8624.