Trends in
Cognitive Sciences

CellPress

Opinion

# A theory of autism bridging across levels of description

Jean-Paul Noel [1,3,*,@] and Dora E. Angelaki [1,2,4]

**Autism impacts a wide range of behaviors and neural functions. As such, theories of autism spectrum disorder (ASD) are numerous and span different levels of description, from neurocognitive to molecular. We propose how existent behavioral, computational, algorithmic, and neural accounts of ASD may relate to one another. Specifically, we argue that ASD may be cast as a disorder of causal inference (computational level). This computation relies on marginalization, which is thought to be subserved by divisive normalization (algorithmic level). In turn, divisive normalization may be impaired by excitatory-to-inhibitory imbalances (neural implementation level). We also discuss ASD within similar frameworks, those of predictive coding and circular inference. Together, we hope to motivate work unifying the different accounts of ASD.**

## Highlights

Autism is a pervasive condition broadly afflicting perceptual, cognitive, social, and motor function.

There are a large number of theories of autism spectrum disorder (ASD), and these span the gamut in terms of levels of description: behavioral, algorithmic, and neural instantiation.

Here we attempt to close the gap between different theories (causal inference, marginalization, divisive normalization, excitatory-to-inhibitory ratio balance, and predictive coding) and levels of description.

## In need of a new kind of theory of ASD

ASD is a heterogeneous neurodevelopmental condition of unknown etiology. It impacts a wide range of functions, from social and communicative faculties [1] to motor behaviors [2] and sensory processing [3]. Similarly, a wide range of genetic [4] and environmental [5] factors have been linked to the disorder. As such, there are a great number of theories of ASD, and these are as diverse as the condition itself. Some emphasize psychological descriptions such as **weak central coherence** (see Glossary) [6] and theory of mind [7], or focus on brainwide [8] or microcircuit [9] connectivity patterns, and still others highlight particular neurotransmitters [10] or genetic mutations [11–13]. Many of these theories are well supported by empirical results; yet, we do not understand the root cause, or causes, of ASD.

We argue that a new kind of theory is needed, one that bridges across levels of description. Computational models are well suited to this task because they allow complex phenomena to be distilled into a set of precisely defined equations and parameters. Furthermore, the latter may then be searched for in neural activity. In turn, in recent years, we have seen a flurry of formal proposals [14–20], most commonly within the framework of the **Bayesian brain** and addressing mechanisms of perception. A mathematical formulation, however, is only the first step. Next, we need to specify how neurons, synapses, proteins, and genes may support this computation, both in neurotypical brains and in those of individuals with ASD. Here, we sketch a working hypothesis, starting from computation and bridging to algorithm and neural instantiation. Although many of the details are still missing, this highlights the need for a cross-disciplinary framework guiding new, theoretically motivated experiments.

## Behavior and computation: causal inference

ASD phenotypes are varied and impact a wide range of functions. In perception, for instance, anomalies have been reported across vision [21], audition [22], touch [23], and olfaction [24], among other sensory modalities. Likewise, reports have highlighted differences in both low-level encoding [25] and the higher-level interpretation of sensory evidence, or semantic

[1]Center for Neural Science, New York University, New York, NY, USA
[2]Tandon School of Engineering, New York University, New York, NY, USA
[3]https://www.jeanpaulnoel.com/
[4]https://angelakilabnyu.org/

*Correspondence:
jpn5@nyu.edu (J.-P. Noel).
@Twitter: @JP__NOEL

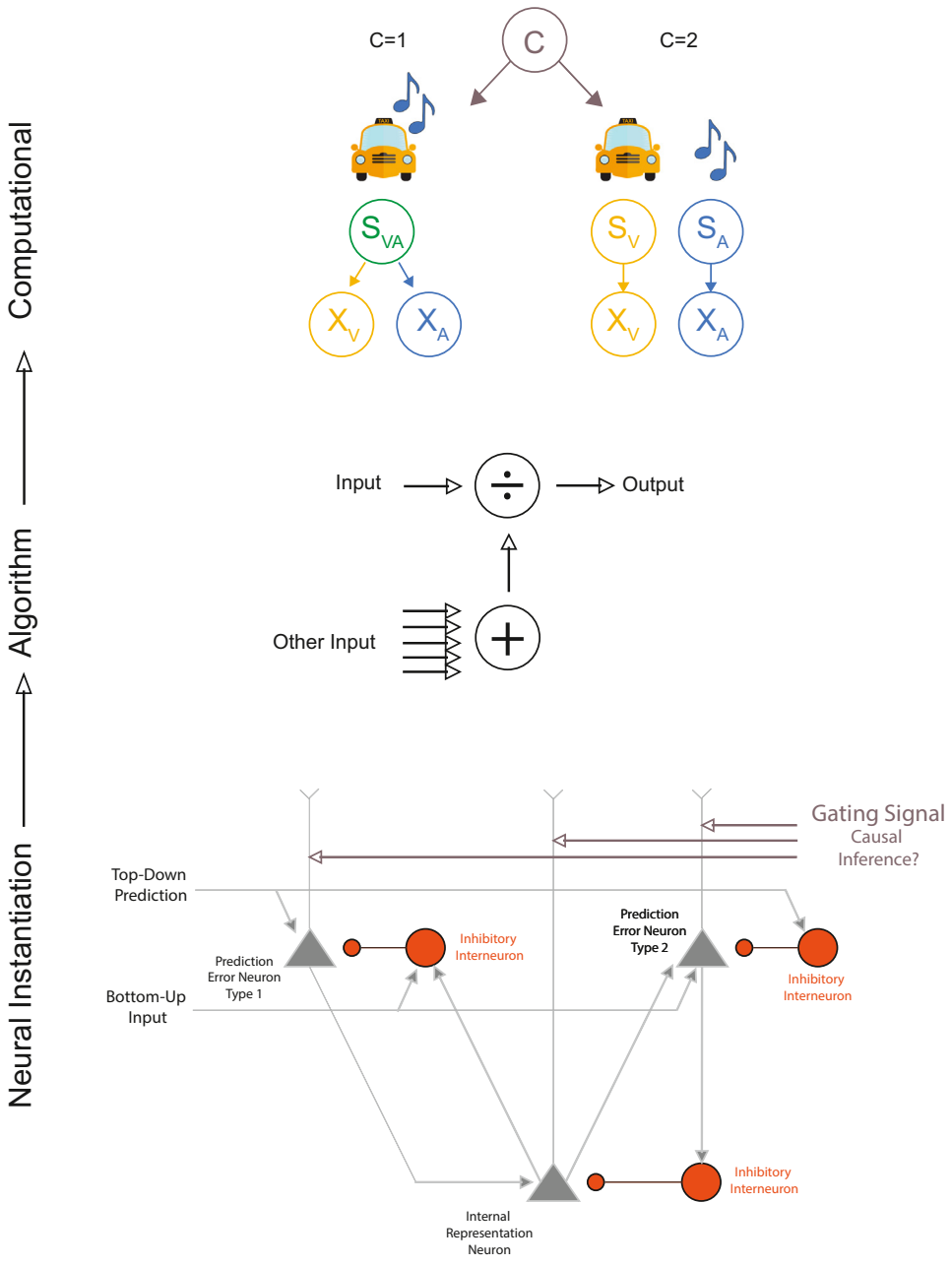


Figure 1. Bridging across computational, algorithmic, and neural accounts of autism spectrum disorder (ASD). Top: causal inference. A generative model for the example where sounds and sights pertain to a single source (C = 1) or multiple sources (C = 2). Left panel: Visual ($X_v$, yellow) and auditory ($X_A$, blue) signals pertain to a single source (i.e., green = yellow + blue). Right panel: Visual ($X_v$, yellow) and auditory ($X_A$, blue) signals pertain to separate sources. Observers first infer the likely causal structure (see Equation I in Box 1) and then weigh these interpretations in perceiving (see Equations II–IV in Box 1). Middle: divisive normalization. Responses of single neurons (output) are divided by the sum of activity from a normalization pool. Bottom: predictive coding. In cortical circuits, there are likely at least two flavors of prediction error neurons: those indicating the unexpected presence (type 1) or absence (type 2) of a signal. This implies a critical role for inhibitory interneurons (orange) in the schematized circuit. Furthermore, a gating or modulatory mechanism dictates when errors might indicate a true change in the environment or not. We propose that this mechanism may be driven by causal inference, attributing deviations from an expectation to either 'noise' (expected fluctuations) or a true change in the environment (unexpected fluctuations), and instantiated via divisive normalization.

## Glossary

**Bayesian brain:** the assumption that the brain can be modeled as performing Bayesian computations and thus inverting generative models via Bayes' rule and representing information in probabilistic terms.
**Divisive normalization:** neural operation wherein responses are the ratio between an excitatory driven and a normalization signal.
**Dynamic range:** within the context of spiking activity, the range (defined over the state space of the stimuli) over which a change in stimuli will lead to a change in neural activity.
**Fano factor:** variance-to-mean ratio of spike counts, usually within a particular time window. Used to measure the variability of spike trains.
**Gain control:** mechanism that allows a change in gain that multiplies or divides the amplitude or strength of (neural) responses to a given input.
**Localization thresholds:** minimal detectable spatial separation between cues, or between a cue and a reference point (e.g., measured in degrees).
**Magnetic resonance spectroscopy:** noninvasive technique allowing the detection and quantification of tissue metabolites.
**Marginalization:** mathematical term referring to summing over all possible values of one variable to determine the 'marginal' contribution of another.
**Mixture priors:** a prior distribution that is a mixture of two or more probability distributions.
**p-common:** the prior probability that there is a single cause versus two causes.
**Probabilistic population codes:** encoding-decoding framework wherein neural populations represent probability distributions over the stimulus.
**Saturation:** within the context of spiking activity, the level at which neurons cannot increase their firing rate.
**Surround suppression:** the phenomenon where the relative firing of a neuron decreases when a stimulus is enlarged (usually beyond the size of the neuron's receptive field).
**Weak central coherence:** a theory suggesting that individuals with ASD fail to see the big picture.

understanding [26]. It follows that if there is a single computation that has gone awry in ASD, this ought to be a canonical one – a modular computation that can be widely applied across behavioral challenges, brain regions, and even animal species. In our opinion, a strong candidate for such a computation is causal inference [27,28].

Causal inference is the ubiquitous and flexible process of attributing observations to (hidden) cause(s) [28]. Some refer to this process as 'structure learning', the process of deducing which hidden environmental structure may have generated observed signals. In perception, for example, we do not have direct access to environmental objects and events or to their causal relationship (the 'generative model' producing environmental signals). Instead, our biological sensors process only a limited set of available inputs. These are stochastic and may be ambiguous. Thus, in building internal models of the external environment, we must first ascribe observations to their likely underlying environmental cause or causes. That is, we must deduce the set of causal relationships within a hypothesized generative model linking observations and hidden (or latent) causes (Figure 1). For instance, an auditory and a visual cue may relate to a single object (e.g., a heard and seen cab) or could indicate two separate ones (e.g., a seen cab and another heard-yet-unseen cab; Figure 1A, see caption and Box 1 for mathematical formalism). Thus, to

## Box 1. Mathematical formalism of causal inference

Let us take the example of inferring the location of an object (or objects) in the environment, given two cues: an auditory ($a$) and visual ($v$) one. In probabilistic terms, to infer the causal structure given sensory observations, we apply Bayes' rule:

$$p(C|X_a, X_V) = \frac{p(X_a, X_V|C)p(C)}{p(X_a, X_V)} \quad [\text{I}]$$

where $p(C)$ is the prior probability of assuming that signals belong to either a single or multiple sources (i.e., cabs), and $p(X_a, X_V)$ is chosen such that $p(C = 1|X_a, X_V)$ and $p(C = 2|X_a, X_V)$ sum to 1 (in this case, where the graphical model assumes $C = 1$ and $C = 2$ represent the totality of all possibilities; see Figure 1A in main text). $p(C = 1)$ is typically referred to as 'p-common', the prior probability of combining cues. Then, assuming a particular cost function (i.e., what type of error we would like to minimize), we may estimate the visual source as a weighted sum of the estimates under different possible causal structures:

$$\hat{S}_v = p(C = 1|X_a, X_V)\hat{S}_{v,c=1} + \left(1 - p(C = 1|X_a, X_V)\right)\hat{S}_{v,c=2} \quad [\text{II}]$$

where $\hat{S}_{v,c=1}$ is the estimate, given we assume auditory and visual cues pertain to the same cab. The optimal (i.e., reliability-weighted [50]) solution with Gaussian likelihoods is

$$\hat{S}_{v,c=1} = \frac{\frac{x_v}{\sigma_v^2} + \frac{x_a}{\sigma_a^2} + \frac{\mu_v}{\sigma_p^2}}{\frac{1}{\sigma_v^2} + \frac{1}{\sigma_a^2} + \frac{1}{\sigma_p^2}} \quad [\text{III}]$$

$\hat{S}_{v,c=2}$ is the visual estimate independent of auditory signals

$$\hat{S}_{v,c=2} = \frac{\frac{x_v}{\sigma_v^2} + \frac{\mu_v}{\sigma_p^2}}{\frac{1}{\sigma_v^2} + \frac{1}{\sigma_p^2}} \quad [\text{IV}]$$

The details specifying the rest of parameters (e.g., $\mu_v$, $\sigma_p^2$, $\sigma_v^2$, $\sigma_a^2$) and analytical forms for the full equation are described elsewhere [27] and depend on the generative model of the specific problem at hand (cf. Figure 2 in [28] for depiction of a number of different generative models). It is important to note, however, that causal inference is the general operation of weighting different worldviews according to their inferred probabilities. Namely, a critical step in this computation is summing the different interpretations and the parameters within these interpretations (Equation II). A component ingredient of causal inference is the optimal combination of cues (Equation III, [50]), provided that these were inferred to relate to the same source.

behave adaptively, we first have to form a hypothesis specifying the probability associated with potential world states. Then, these hypotheses are weighted according to their relative evidence in generating a percept, taking into account all possible worlds. This is a process of **marginalization**, where we remove a variable (in this case, the hypothesized world states) by summing over all possible values this variable could take. (Note that this process of marginalization may be implicit, such as in the use of **mixture priors** [29,30].) Importantly, this process of causal inference is true not only in perception but also in cognition and action. For example, in reasoning, we may attempt to deduce the causal structure between a set of factors (e.g., age, sex, contact with a virus) and an outcome (e.g., disease), or, during motor behaviors, we may attempt to classify sensory input as consequences of our actions or as emanating from the external environment.

In perception, the telltale sign of causal inference is a stereotyped pattern of behavior wherein independent sources may bias perception of one another when they are weakly (dis)similar, but not when they are very distinct [27,28]. For example, when trying to localize a cab, if auditory and visual signals originate from vastly distinct locations, these ought to be treated independently and thus not bias the localization of one another. However, if a visible cab and a honking-yet-unseen cab are near one another, we may deduce an incorrect set of causal relationships (i.e., internal model) in attributing both visual and auditory signals to a single cab. In this scenario, we would misuse both auditory and visual cues to estimate the location of our (e.g., visual) source of interest, yielding perceptual biases. This pattern of behavior, with (attractive) biases being expressed exclusively when distinct cues have relatively similar features, has been observed across a wide array of tasks, including but not limited to spatial localization [27,31–33], orientation judgments [34], rate detection [35], speech intelligibility [36], weight [37] and body [38–40] perception, motor learning [41], heading estimation [42], and spatial navigation [43].

The suggestion that causal inference is impaired in ASD was supported first by tangential evidence and more recently by a larger-scale, multiexperiment approach combined with formal model fitting. The early hints came from studies in binocular rivalry demonstrating that individuals with ASD altered between ambiguous percepts at a slower rate than their neurotypical counterparts [44]. That is, the set of causal relationships hypothesized to link observed photons and their underlying generative images is less flexible in individuals with ASD (see [45] for a related interpretation of bistable perception based on circular inference). Also, there is a vast literature demonstrating that individuals with ASD will claim that audiovisual cues were presented synchronously at larger asynchronies than control subjects do ([46]; see [47] for a meta-analysis of 53 such studies). Leveraging speech stimuli, researchers [48] applied a causal inference model [36] to simultaneous judgments from neurotypical controls and individuals with ASD. Findings suggested that individuals with ASD had a greater *a priori* probability of combining audio and visual cues, regardless of whether these signals ought or ought not to be integrated. This is typically referred to as a heightened '**p-common**' (i.e., an *a priori* probability of combining cues).

More recently, a large-scale study (five psychophysical studies including over 90 adolescents) was conducted to directly test whether individuals with ASD showed impairments in causal inference [49]. The researchers measured (i) unisensory auditory and visual **localization thresholds**, (ii) Bayes optimal cues combination not requiring the deduction of causal relationships [50], and (iii) explicit causal inference as well as implicit causal inference across (iv) audiovisual and (v) visual-visual pairings. That is, this study tested low-level sensory encoding, multisensory integration (i.e., optimal cue combination), and causal inference within a single cohort and stimulus set. All data were used to create a single model fit per participant, which demonstrated no impairment in sensory processing or cue combination in ASD but did show a larger (implicit) 'p-common' in these individuals relative to their neurotypical counterparts. Namely, even at very large audiovisual

or visual-visual disparities (the former in spatial location, the latter in velocity of motion), individuals with ASD behaved as if always (maladaptively) combining cues. Together, these results explain why individuals with ASD have been widely reported to show anomalous multisensory behavior [46,47,51]. It is not due to an impairment in the process of multisensory integration itself, which is in line with statistical optimality [50] in both controls and individuals with ASD [49,52], but because of a more general anomaly in establishing and updating an internal model specifying when cues ought or ought not to be combined. The observation that causal inference is impaired in ASD but optimal cue combination [50] is not [49,52] provides an important clue to a potential algorithmic deficit that may exist in ASD.

## Algorithm: marginalization and divisive normalization

Bayesian cue combination [50] specifies how to combine signals in a statistically optimal manner, provided that these belong to the same source (see Box 1). That is, it is a subcomponent of the larger causal inference scheme. This process of cue combination is not impaired in ASD [49,52]. From the standpoint of mathematical operations, a major differentiator between cue combination (intact in ASD) and causal inference (impaired in ASD) is the presence of marginalization in the latter but not the former. Interestingly, theoretical work [53] showed that biologically plausible neural networks (e.g., those associated with constant **Fano factors** as in linear **probabilistic population codes** [54]) perform marginalization via a neural computation known as **divisive normalization** [55]. The implication is that causal inference may rely on divisive normalization, and thus this is a neural operation that may have gone awry in ASD.

Divisive normalization is a critical operation allowing neurons to adapt their **dynamic range** according to context and thus combat **saturation**. This is accomplished by computing a ratio between the response of an individual neuron and the summed activity of a pool of neurons (Figure 1B, [55,56]). Importantly, this form of **gain control** effectively modulates synaptic weights and is considered a canonical neural computation [56] in that it has been described across primary [55] and extrastriate [57] visual cortex, auditory cortex [58], the superior colliculus [59], and the antenna lobe of fruit flies [60], among others. It is thought to be involved in contrast- and pattern-invariant visual representations [56,61], as well as in concentration-invariant odorant encoding in the olfactory system [60]. Namely, just as causal inference is a canonical behavioral computation, divisive normalization is a canonical neural computation.

In line with the argument that marginalization through divisive normalization may have gone awry in ASD, *in silico* simulations [62] showed that alterations in divisive normalization could account for diverse phenomena present in the condition, such as alterations in visuospatial suppression [63,64] or greater tunnel vision by **surround suppression** [65,66] (see [21] and [67], respectively, for a classic and recent review on sensory perception in ASD, including the highlighting of numerous conflicting results). This modeling work [62] argued that differences in divisive normalization could also putatively account for enhanced local (vs. global [68]) and simple (vs. complex [69]) processing in ASD, as well as for deficits in social cognition [70], altered visual search [71], and increased variance [72,73] and size of receptive fields [74] in ASD. However, subsequent reports examining cross-orientation suppression [75] and adaptation to gaze direction [76], processes thought to rely on divisive normalization, showed no difference between controls and individuals with ASD.

To summarize, divisive normalization is a canonical neural computation [54] that may subserve causal inference ([53]; see also [77,78]) and accounts for a variety of the peculiarities present in ASD [62], but not all [75,76]. The likely resolution to the latter observation is that 'vanilla' divisive normalization models are imperfect. Fittingly, recent work [79] examined whether a standard divisive normalization model could account for neural responses in V1 as macaques viewed natural

images (as opposed to the better-studied yet less ecologically relevant Gabors). The authors reported that variability in V1 spiking activity was not well explained by existing descriptions of surround suppression through divisive normalization. Instead, they were better predicted by a model wherein divisive normalization was gated by a probabilistic inference regarding the homogeneity of images. In short, according to this model [79], divisive normalization should be or is only applied when neurons and their normalization pools encode a single object. Put differently, in the augmented divisive normalization model [79], causal inference gates divisive normalization (see [75] for an earlier account arguing that this augmented divisive normalization model may better explain disparities between ASD and neurotypical perception than the standard model [62]).

## Neural instantiation: excitatory-to-inhibitory ratio balance

So far, we have argued that anomalies in perception-driven behavior in ASD may emanate from abnormalities in causal inference [48,49] and that this computation is subserved algorithmically by a process of marginalization. Furthermore, we have suggested that marginalization relies on divisive normalization [53,55,56]. Consistent with this, we have seen that idiosyncrasies in divisive normalization may partially account for phenotypes observed in ASD [62] and that authors have argued for a strong relationship between divisive normalization and causal inference (either the latter being subserved by the former [53,77,78] or the latter gating the former [79]). In this section, we briefly discuss neural instantiation.

There is a strong and previously discussed (e.g., [62]) relationship between divisive normalization and the well-known hypothesis that ASD symptomatology arises from a heightened excitatory-to-inhibitory ratio (E/I balance, [80]). This hypothesis explains the frequent comorbidity between autism and seizures [81] and has been supported by observations indicating increased glutamatergic and decreased GABAergic activity, as well as a decrease in the number of γ-aminobutyric acid (GABA) receptors [82] in ASD. Similarly, many of the genes associated with ASD – and, as a consequence, many of the developed mouse models of the disorder [83,84] – relate to glutamate or GABA signaling. Studies leveraging **magnetic resonance spectroscopy** have also indicated disruptions in GABA concentration in visual [85], auditory [85], and somatosensory [86] cortices of humans with ASD. In fact, GABA concentrations seemingly correlate with ASD symptomatology severity [87], even though there is not always a categorical difference in GABA concentration between neurotypical and atypical groups [87], putatively due to homeostatic compensatory mechanisms [84].

Relating back to divisive normalization, one could, of course, conjecture that alterations in the genetic and neurochemical makeup of the brains of individuals with ASD may impact the machinery necessary for divisive normalization. The inverse relationship may also be true, with the inability for gain modulation through divisive normalization leading to, for example, increased seizures [81] and alterations in the amplitude of neural responses [73] and the size of receptive fields [74]. We are agnostic regarding whether molecular (e.g., availability of GABA) or circuit-level (e.g., gain modulation via divisive normalization) mechanisms are 'primary' in ASD, with the condition potentially being expressed due to several different biological root causes (i.e., an equifinal entity). Instead, we aim to highlight that a computational deficit in causal inference [27,28,48,49] – which implies (nonlinear) inference over putative generative structures in addition to (linear) combination of Bayesian likelihoods and priors within a given structure – naturally suggests impairments in divisive normalization [62] and E/I balance [72].

## Multilevel models: predictive coding and circular inference

The framework we have described has the advantage of neatly dissociating between computational, algorithmic, and biological levels of analysis. As such, it may be useful in making predictions

at each level of analysis (e.g., computational prediction of the current account: no deficit in ASD in combining priors and likelihoods within static generative structures, but in adaptively updating inferred generative structures [88]; biological prediction: changing interpretations of a given sensory milieu should be accompanied by changing functional connectivity within normalization pools, and maladaptively in ASD). However, the distinction between levels of analysis may be more useful to researchers than a reality of brain function (e.g., there is no computational level in the brain). In turn, it is important to highlight largely compatible models that apply across levels of description, such as predictive coding [89] and notions of circular inference [90] (see Box 2 for further comparison between our account and predictive coding theories of ASD).

ASD has been described as a disorder of prediction [15,16,91], and, just as causal inference and divisive normalization may, respectively, be canonical computations at a behavioral and algorithmic level, predictive coding has been described as a basic computational and neural motif [92]. In short, in predictive coding, perception is as much 'top-down' as it is 'bottom-up'. Instead of neural codes becoming increasingly complex as one ascends the neural hierarchy (i.e., a representational framework), in predictive coding, neural nodes compare their inputs to their expectations as inherited from 'top-down' signals. Then, the differences between these, the prediction errors, are sent downstream and become the input to the next node in the neuraxis. Within this framework, the inhibition of neural responses that occurs as stimuli extend into surround fields ([93], algorithmically explained by divisive normalization [62] and anomalous in ASD [63]) is the consequence of stimuli at a given location acting as a prediction for stimuli in neighboring regions [92]. Thus, within this framework, differences in surround suppression in ASD [63] are due to deficits in making adaptive predictions.

To instantiate the framework of predictive coding, neural networks necessitate a microcircuit to maintain an internal model giving rise to predictions, as well as a comparator microcircuit computing the prediction error between bottom-up inputs and top-down predictions. In cortical nodes where neurons do not have high intrinsic firing rates (in contrast to subcortical areas), neurons

## Box 2. Bayesian brain hypotheses of ASD

The initial 'Bayesian brain' account of ASD stated that individuals with the condition underused their expectations or Bayesian priors [14]. This account has been challenged by empirical work (e.g., [17,52,100]; see [88] for a review) but did lead to a more subtle formulation: that individuals with ASD have 'high, inflexible precision of prediction errors (HIPPEA)' [15]. This latter account has received support [16,101] but has also made predictions that have not been borne out in empirical findings. For instance, a biological implementation of HIPPEA would involve the inappropriate cancellation of prediction errors leading to circular inference [90,99] wherein bottom-up signals reverberate and are misinterpreted as predictions. A recent study [102], however, did not find increased circular inference in ASD (although this work awaits replication due to a limited range in ASD or ASD-like phenotypes and the fact that it was not conducted in a controlled environment). Similarly, HIPPEA [15] states that 'optimal [multisensory] integration will not take place, because all cues, even redundant or very uncertain ones, will be weighed equally'. This prediction has been falsified across visual-vestibular [52], audiovisual [49], and visual-visual [49] pairings, where individuals with ASD combine cues in a statistically optimal manner, akin to their neurotypical counterparts. This is true even when cue reliability changes on a trial-by-trial basis [49] and thus demands appropriate reliability-based weighting.

We consider the causal inference account proposed here to be an evolution of rather than a departure from HIPPEA [15]. Indeed, these models have more in common than they have in difference (see, e.g., [103]). As HIPPEA, we consider the defining characteristic of ASD pathology to be the inability to flexibly adjust to context. However, we consider this to be true only when the adjustment of inferred causal structures ('what observation relates to what latent cause') is needed. This is seemingly congruent with the observation that updating of contextual (i.e., experiment-specific) but not structural (i.e., true in the external world) priors is impaired in ASD [88]. In particular, the causal inference account would not predict a heightened weight of prediction errors, as during bottom-up circular inference [102], or impaired optimal cue combination.

It is true that two studies fitting formal causal inference models to ASD data [48,49] demonstrated a heightened 'p-common'. However, we consider a heightened 'p-common' not as the defining characteristic, but simply as the inflexible adjustment of inferred causal structures, as demonstrated by studies in binocular rivalry [44,104–109].

may be able to signal the presence of an unexpected input via an increase in firing rate, but they would not be able to signal the absence of an expected input (putatively requiring a decrease in firing rate). Thus, it is likely that two types of prediction error neurons exist [92], either increasing their firing according to a difference between input and prediction (type 1; input – prediction) or the difference between prediction and input (type 2; prediction – input). This could be accomplished by a circuit akin to that depicted in Figure 1C, which emphasizes the critical need for inhibitory interneurons (i.e., E/I balance). This critical need is similarly emphasized in quantitative simulations [90] demonstrating that the absence of proper inhibitory loops subtracting predictions and prediction errors results in pathological circular belief propagation wherein predictions (priors) or prediction errors (likelihoods) reverberate throughout the circuit and end up exerting undue influence (i.e., as in the heightened precision account of ASD [15,93]; see Box 2 for further discussion contrasting causal inference and the heightened precision of prediction errors accounts of ASD).

A second critical element of the predictive coding circuitry ought to be a gating or modulatory signal ([16,92,94], Figure 1C), given that not all errors (or surprises) are created equally. Some should be expected, given contextual cues (as when stimuli are noisy), whereas others should not. Only 'unexpected uncertainty' should result in a changed internal model, because these likely reflect a change in the external environment [95,96]. Just as dopaminergic signals are thought to adjust learning rates in reward prediction errors [97], some [16,94] have suggested that neuromodulators (particularly acetylcholine and norepinephrine) fine-tune the precision, and thus weight, attributed to predictions and prediction errors. It is precisely this context-dependent modulation that is hypothesized to be aberrant in ASD [15,16,93] and which potentially leads to the slow updating of internal models, including their associated expectations, or priors [16,18,19,98]. Here, we speculate that divisive normalization may be a circuit-level mechanism to regulate the precision of prediction errors and that this mechanism may be selectively involved in computations leading to an inference over generative structure(s), or structure learning [79]. Nonetheless, broadly, our account based on causal inference is largely compatible with the inflexible precision of prediction errors account [15,16,93] (see Box 2), which may render the process of structure learning an optimization rather than an inference problem. In our opinion, however, the current account more naturally links to existent theories based on divisive normalization [62] and E/I balance [80].

## Concluding remarks

In summary, we have sketched a roadmap from behavior and causal inference to marginalization, divisive normalization, and E/I imbalance. We also discussed predictive coding and its similarities and dissimilarities with the current account (see Box 2). Prior work has suggested that individuals with ASD show impairments in each of these elements (causal inference [48,49], divisive normalization [56,62], predictive coding [15,16], context-dependent uncertainty partitioning [16,95], and E/I balance [80]). Similarly, prior work has made the association between some of these elements (see [15,62,75,90]). Our contribution here is in highlighting the relationship between each of these individual elements and in arguing that these may not be different theories or accounts of the disorder, but all part of a unified one bridging across levels of description. Furthermore, unlike prior work, we emphasize deficits in the flexible updating of inferred generative structures – the process of causal inference [27,28,48,49] – as key in ASD pathology. In our opinion, this view most naturally relates to impairments in marginalization, divisive normalization, and E/I imbalances. But it is certainly not incompatible or contradictory with previous predictive coding [15] and circular inference [90,99] accounts of ASD and psychopathology more generally. Of course, much work remains to be done (see Outstanding questions) from both basic science and clinical perspectives (see Box 3 for speculation on how our framework may relate to ASD symptomatology). Although direct evidence for certain aspects of our proposal is limited at present, the strength of the current framework is that it naturally ties in previous, more established accounts

## Outstanding questions

Would individuals with ASD show anomalies in causal inference within naturalistic contexts? Causal inference has been well studied in laboratory tasks requiring binary choices and discretizing action and perception. However, we do not know how this translates to more natural contexts with continuous time and active sensing.

How is causal inference computed by neurons and neural ensembles? There seems to be a strong association between causal inference and divisive normalization, but the exact relationship is unknown. The relationship currently rests on the observation that algorithmically causal inference necessitates marginalization, which can be implemented via divisive normalization.

Can 'augmented' divisive normalization models account for recent empirical observations within the cognitive neuroscience study of ASD? How does divisive normalization relate to hypo- and hypersensitivity to sensory stimuli (see Box 3)?

What animal models and what behavioral protocols would be most effective in querying predictive coding in ASD from a systems neuroscience perspective? Dissection of the neural circuit supporting predictive coding is occurring at the level of systems neuroscience, whereas theories of predictive coding in ASD exist at the levels of psychology and cognitive neuroscience.

Can we assume that cost functions are similar across individuals with ASD and their neurotypical counterparts? Most reports fitting Bayesian models to the study of ASD assume a given cost function (e.g., maximizing reward rate), but these are subjective and difficult to establish. For instance, the subjective cost associated with performing actions may be different across ASD and neurotypical individuals.

Would manipulations of E/I balance and divisive normalization selectively impair causal inference without impairing optimal cue combination?

### Box 3. Relationship to ASD symptomatology

ASD is clinically characterized by sociocommunicative impairments, restrictive and repetitive behaviors, and perceptual anomalies including hyper- and hyposensitivity to sensory stimulation [1]. How does an account based on links between causal inference, marginalization, divisive normalization, and E/I balance relate to this symptomatology?

We can only speculate, given that few studies have formally tested causal inference within an ASD population, and none has meticulously characterized the symptomatology of the tested cohort. Given this caveat, we consider the key behavioral deficit in ASD to be the inability to update internal models linking observations to hidden causes – for example, flexibly deciding whether multisensory cues relate to the same objects and thus ought to be integrated or not. This includes the inflexible updating of expectations [15,18,19]. Accordingly, anomalies in perception are due to a mismatch between the true and inferred generative structures yielding sensory signals. Hyper- and hyposensitivity may also emanate from deficits in contextualization, where a given stimulus may appear disproportionality intense/dull, depending on context. In this line, some [67] have recently argued that disturbances in divisive normalization and E/I balance may explain the striking finding that perception in ASD does not adhere to Weber's law [110] (i.e., is not properly contextualized).

Repetitive behaviors, insistence on sameness, and stereotypy could result from attempts to align incoming sensory evidence with the inflexible expectations of what that sensory input ought to be. In fact, recent work demonstrates that individuals with ASD will take overt compensatory measures [49] when faced with ambiguous generative structures.

As others have before us (e.g., [15,21]), we consider that sociocommunicative impairments may be scaffolded on perceptual deficits and/or be particularly complex scenarios where latent structures are intricate and continuously changing, thus readily highlighting the computational deficits present in ASD.

(particularly the association between causal inference and divisive normalization). We hope this big-picture perspective will help design future theory-motivated experiments and ultimately impact positively the lives of those with ASD.

## Acknowledgments

We thank Edoardo Balzani for commenting on an earlier draft of the manuscript. We also thank the reviewers for constructive feedback and greatly improving the manuscript. The work was supported by National Institutes of Health grants U19NS118246 (to D.E.A.) and K99NS128075 (to J-P.N.).

## Declaration of interests

No interests are declared.

## References

1. American Psychiatric Association (2013) *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*, American Psychiatric Association
2. Zampella, C.J. *et al.* (2021) Motor skill differences in autism spectrum disorder: a clinically focused review. *Curr. Psychiatry Rep.* 23, 64
3. Robertson, C.E. and Baron-Cohen, S. (2017) Sensory perception in autism. *Nat. Rev. Neurosci.* 18, 671–684
4. Ronemus, M. *et al.* (2014) The role of de novo mutations in the genetics of autism spectrum disorders. *Nat. Rev. Genet.* 15, 133–141
5. Modabbernia, A. *et al.* (2017) Environmental risk factors for autism: an evidence-based review of systematic reviews and meta-analyses. *Mol. Autism* 8, 13
6. Happé, F. and Frith, U. (2006) The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *J. Autism Dev. Disord.* 36, 5–25
7. Baron-Cohen, S. *et al.* (1985) Does the autistic child have a 'theory of mind'? *Cognition* 21, 37–46
8. Vasa, R.A. *et al.* (2016) The disrupted connectivity hypothesis of autism spectrum disorders: time for the next phase in research. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 1, 245–252
9. Shepherd, G.M. and Katz, D.M. (2011) Synaptic microcircuit dysfunction in genetic models of neurodevelopmental disorders: focus on Mecp2 and Met. *Curr. Opin. Neurobiol.* 21, 827–833
10. Marotta, R. *et al.* (2020) The neurochemistry of autism. *Brain Sci.* 10, 163
11. Peça, J. *et al.* (2011) Shank3 mutant mice display autistic-like behaviours and striatal dysfunction. *Nature* 472, 437–442
12. Peñagarikano, O. *et al.* (2011) Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. *Cell* 147, 235–246
13. Dutch-Belgian Fragile X Consortium (1994) Fmr1 knockout mice: a model to study fragile X mental retardation. *Cell* 78, 23–33
14. Pellicano, E. and Burr, D. (2012) When the world becomes 'too real': a Bayesian explanation of autistic perception. *Trends Cogn. Sci.* 16, 504–510
15. Van de Cruys, S. *et al.* (2014) Precise minds in uncertain worlds: predictive coding in autism. *Psychol. Rev.* 121, 649–675
16. Lawson, R.P. *et al.* (2017) Adults with autism overestimate the volatility of the sensory environment. *Nat. Neurosci.* 20, 1293–1299
17. Karvelis, P. *et al.* (2018) Autistic traits, but not schizotypy, predict increased weighting of sensory information in Bayesian visual integration. *eLife* 7, e34115
18. Lieder, I. *et al.* (2019) Perceptual bias reveals slow-updating in autism and fast-forgetting in dyslexia. *Nat. Neurosci.* 22, 256–264
19. Vishne, G. *et al.* (2021) Slow update of internal representations impedes synchronization in autism. *Nat. Commun.* 12, 5439
20. Schneebeli, M. *et al.* (2022) Disentangling Bayesian brain theories of autism spectrum disorder. *medRxiv* Published online February 8, 2022. https://doi.org/10.1101/2022.02.07.22270242
21. Simmons, D.R. *et al.* (2009) Vision in autism spectrum disorders. *Vis. Res.* 49, 2705–2739
22. Kujala, T. *et al.* (2013) The neural basis of aberrant speech and audition in autism spectrum disorders. *Neurosci. Biobehav. Rev.* 37, 697–704

23. Cascio, C.J. (2010) Somatosensory processing in neurodevelopmental disorders. *J. Neurodev. Disord.* 2, 62–69
24. Tonacci, A. *et al.* (2017) Olfaction in autism spectrum disorders: a systematic review. *Child Neuropsychol.* 23, 1–25
25. Noel, J.-P. *et al.* (2021) Individuals with autism spectrum disorder have altered visual encoding capacity. *PLoS Biol.* 19, e3001215
26. Golarai, G. *et al.* (2006) Autism and the development of face processing. *Clin. Neurosci. Res.* 6, 145–160
27. Körding, K.P. *et al.* (2007) Causal inference in multisensory perception. *PLoS One* 2, e943
28. Shams, L. and Beierholm, U. (2022) Bayesian causal inference: a unifying neuroscience theory. *Neurosci. Biobehav. Rev.* 137, 104619
29. Shams, L. and Beierholm, U.R. (2010) Causal inference in perception. *Trends Cogn. Sci.* 14, 425–432
30. Noel, J.P. *et al.* (2018) Peri-personal space as a prior in coupling visual and proprioceptive signals. *Sci. Rep.* 8, 15819
31. Aller, M. and Noppeney, U. (2019) To integrate or not to integrate: temporal dynamics of hierarchical Bayesian causal inference. *PLoS Biol.* 17, e3000210
32. Rohe, T. and Noppeney, U. (2015) Cortical hierarchies perform Bayesian causal inference in multisensory perception. *PLoS Biol.* 13, e1002073
33. Ferrari, A. and Noppeney, U. (2021) Attention controls multisensory perception via two distinct mechanisms at different levels of the cortical hierarchy. *PLoS Biol.* 19, e3001465
34. van den Berg, R. *et al.* (2012) Optimal inference of sameness. *Proc. Natl. Acad. Sci. U. S. A.* 109, 3178–3183
35. Cao, Y. *et al.* (2019) Causal inference in the multisensory brain. *Neuron.* 102, 1076–1087.e8
36. Magnotti, J.F. *et al.* (2013) Causal inference of asynchronous audiovisual speech. *Front. Psychol.* 4, 798
37. Peters, M.A. *et al.* (2016) The size-weight illusion is not anti-Bayesian after all: a unifying Bayesian account. *PeerJ.* 4, e2124
38. Samad, M. *et al.* (2015) Perception of body ownership is driven by Bayesian sensory inference. *PLoS One* 10, e0117178
39. Fang, W. *et al.* (2019) Statistical inference of body representation in the macaque brain. *Proc. Natl. Acad. Sci. U. S. A.* 116, 20151–20157
40. Qi, G. *et al.* (2022) Neural dynamics of causal inference in the macaque frontoparietal circuit. *eLife* 11, e76145
41. Kording, K. and Tenenbaum, J. (2006) *Causal Inference in Sensorimotor Integration*, NeurIPS
42. Dokka, K. *et al.* (2019) Causal inference accounts for heading perception in the presence of object motion. *Proc. Natl. Acad. Sci. U. S. A.* 116, 9060–9065
43. Noel, J.P. *et al.* (2023) Causal inference during closed-loop navigation: parsing of self- and object-motion. *bioRxiv* Published online January 30, 2023. https://doi.org/10.1101/2023.01.27.525974
44. Spiegel, A. *et al.* (2019) Slower binocular rivalry in the autistic brain. *Curr. Biol.* 29, 2948–2953.e3
45. Leptourgos, P. *et al.* (2020) A functional theory of bistable perception based on dynamical circular inference. *PLoS Comput. Biol.* 16, e1008480
46. Stevenson, R.A. *et al.* (2014) Multisensory temporal integration in autism spectrum disorders. *J. Neurosci.* 34, 691–697
47. Feldman, J.I. *et al.* (2018) Audiovisual multisensory integration in individuals with autism spectrum disorder: a systematic review and meta-analysis. *Neurosci. Biobehav. Rev.* 95, 220–234
48. Noel, J.P. *et al.* (2018) Atypical audiovisual temporal function in autism and schizophrenia: similar phenotype, different cause. *Eur. J. Neurosci.* 47, 1230–1241
49. Noel, J.P. *et al.* (2022) Aberrant causal inference and presence of a compensatory mechanism in autism spectrum disorder. *eLife* 11, e71866
50. Ernst, M.O. and Banks, M.S. (2002) Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415, 429–433
51. Baum, S.H. *et al.* (2015) Behavioral, perceptual, and neural alterations in sensory and multisensory function in autism spectrum disorder. *Prog. Neurobiol.* 134, 140–160
52. Zaidel, A. *et al.* (2015) Self-motion perception in autism is compromised by visual noise but integrated optimally across multiple senses. *Proc. Natl. Acad. Sci. U. S. A.* 112, 6461–6466
53. Beck, J.M. *et al.* (2011) Marginalization in neural circuits with divisive normalization. *J. Neurosci.* 31, 15310–153109
54. Ma, W.J. *et al.* (2006) Bayesian inference with probabilistic population codes. *Nat. Neurosci.* 9, 1432–1438
55. Carandini, M. and Heeger, D.J. (1994) Summation and division by neurons in primate visual cortex. *Science.* 264, 1333–1336
56. Carandini, M. and Heeger, D.J. (2011) Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* 13, 51–62 Erratum in: *Nat. Rev. Neurosci.* (2013) 14,152
57. Ohshiro, T. *et al.* (2017) A neural signature of divisive normalization at the level of multisensory integration in primate cortex. *Neuron* 95, 399–411.e8
58. Rabinowitz, N.C. *et al.* (2011) Contrast gain control in auditory cortex. *Neuron* 70, 1178–1191
59. Vokoun, C.R. *et al.* (2014) Response normalization in the superficial layers of the superior colliculus as a possible mechanism for saccadic averaging. *J. Neurosci.* 34, 7976–7987
60. Olsen, S.R. *et al.* (2010) Divisive normalization in olfactory population codes. *Neuron.* 66, 287–299
61. Aqil, M. *et al.* (2021) Divisive normalization unifies disparate response signatures throughout the human visual hierarchy. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2108713118
62. Rosenberg, A. *et al.* (2015) A computational perspective on autism. *Proc. Natl. Acad. Sci. U. S. A.* 112, 9158–9165
63. Foss-Feig, J.H. *et al.* (2013) A substantial and unexpected enhancement of motion perception in autism. *J. Neurosci.* 33, 8243–8249 Erratum in: J. Neurosci. (2013) 33, 10583
64. Flevaris, A.V. and Murray, S.O. (2015) Orientation-specific surround suppression in the primary visual cortex varies as a function of autistic tendency. *Front. Hum. Neurosci.* 8, 1017
65. Mann, T.A. and Walker, P. (2003) Autism and a deficit in broadening the spread of visual attention. *J. Child Psychol. Psychiatry* 44, 274–284
66. Robertson, C.E. *et al.* (2013) Tunnel vision: sharper gradient of spatial attention in autism. *J. Neurosci.* 33, 6776–6781
67. Hadad, B.S. and Yashar, A. (2022) Sensory perception in autism: what can we learn? *Annu. Rev. Vis. Sci.* 8, 239–264
68. Dakin, S. and Frith, U. (2005) Vagaries of visual perception in autism. *Neuron* 48, 497–507
69. Bertone, A. *et al.* (2005) Enhanced and diminished visuo-spatial information processing in autism depends on stimulus complexity. *Brain* 128, 2430–2441
70. Mundy, P. (2018) A review of joint attention and social-cognitive brain systems in typical development and autism spectrum disorder. *Eur. J. Neurosci.* 47, 497–514
71. Keehn, B. and Joseph, R.M. (2016) Slowed search in the context of unimpaired grouping in autism: evidence from multiple conjunction search. *Autism Res.* 9, 333–339
72. Noel, J.P. *et al.* (2020) Increased variability but intact integration during visual navigation in autism spectrum disorder. *Proc. Natl. Acad. Sci. U. S. A.* 117, 11158–11166
73. Dinstein, I. *et al.* (2012) Unreliable evoked responses in autism. *Neuron.* 75, 981–991
74. Schwarzkopf, D.S. *et al.* (2014) Larger extrastriate population receptive fields in autism spectrum disorders. *J. Neurosci.* 34, 2713–2724
75. Van de Cruys, S. *et al.* (2018) Intact perceptual bias in autism contradicts the decreased normalization model. *Sci. Rep.* 8, 12559
76. Palmer, C.J. *et al.* (2018) Autistic adults show preserved normalisation of sensory responses in gaze processing. *Cortex* 103, 13–23
77. Cuppini, C. *et al.* (2017) A biologically inspired neurocomputational model for audiovisual integration and causal inference. *Eur. J. Neurosci.* 46, 2481–2498
78. Zhang, W. *et al.* (2019) A normative theory for causal inference and Bayes factor computation in neural circuits. *NeurIPS* 32, 3804–3813
79. Coen-Cagli, R. *et al.* (2015) Flexible gating of contextual influences in natural vision. *Nat. Neurosci.* 18, 1648–1655
80. Rubenstein, J.L. and Merzenich, M.M. (2003) Model of autism: increased ratio of excitation/inhibition in key neural systems. *Genes Brain Behav.* 2, 255–267
81. Canitano, R. (2007) Epilepsy in autism spectrum disorders. *Eur. Child Adolesc. Psychiatry* 16, 61–66

82. Yizhar, O. *et al.* (2011) Neocortical excitation/inhibition balance in information processing and social dysfunction. *Nature* 477, 171–178
83. Kazdoba, T.M. *et al.* (2016) Translational mouse models of autism: advancing toward pharmacological therapeutics. *Curr. Top. Behav. Neurosci.* 28, 1–52
84. Nelson, S.B. and Valakh, V. (2015) Excitatory/inhibitory balance and circuit homeostasis in autism spectrum disorders. *Neuron* 87, 684–698
85. Robertson, C.E. *et al.* (2016) Reduced GABAergic action in the autistic brain. *Curr. Biol.* 26, 80–85
86. Puts, N.A.J. *et al.* (2017) Reduced GABA and altered somatosensory function in children with autism spectrum disorder. *Autism Res.* 10, 608–619
87. Brix, M.K. *et al.* (2015) Brain MR spectroscopy in autism spectrum disorder-the GABA excitatory/inhibitory imbalance theory revisited. *Front. Hum. Neurosci.* 9, 365
88. Angeletos Chrysaitis, N. and Seriès, P. (2023) 10 years of Bayesian theories of autism: a comprehensive review. *Neurosci. Biobehav. Rev.* 145, 105022
89. Rao, R.P. and Ballard, D.H. (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87
90. Jardri, R. and Denève, S. (2013) Circular inferences in schizophrenia. *Brain* 136, 3227–3241
91. Sinha, P. *et al.* (2014) Autism as a disorder of prediction. *Proc. Natl. Acad. Sci. U. S. A.* 111, 15220–15225
92. Keller, G.B. and Mrsic-Flogel, T.D. (2018) Predictive processing: a canonical cortical computation. *Neuron.* 100, 424–435
93. Van de Cruys, S. *et al.* (2017) Disentangling signal and noise in autism spectrum disorder. *Brain Cogn.* 112, 78–83
94. Friston, K.J. (2010) The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138
95. Bervoets, J. *et al.* (2021) Autism and intolerance of uncertainty: an ill-fitting pair. *Trends Cogn. Sci.* 25, 1009–1010
96. Pulcu, E. and Browning, M. (2019) The misestimation of uncertainty in affective disorders. *Trends Cogn. Sci.* 23, 865–875
97. Schultz, W. (2016) Dopamine reward prediction error coding. *Dialogues Clin. Neurosci.* 18, 23–32
98. Vine, A.T. *et al.* (2022) Testing predictive coding theories of autism spectrum disorder using models of active inference. *PsyArXiv* Published online December 12, 2022. http://doi.org/10.31234/osf.io/zyfkn
99. Jardri, R. *et al.* (2017) Experimental evidence for circular inference in schizophrenia. *Nat. Commun.* 8, 14218
100. Smith, D. *et al.* (2017) The integration of occlusion and disparity information for judging depth in autism spectrum disorder. *J. Autism Dev. Disord.* 47, 3112–3124
101. Goris, J. *et al.* (2018) Sensory prediction errors are less modulated by global context in autism spectrum disorder. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 3, 667–674
102. Angeletos Chrysaitis, N. *et al.* (2021) No increased circular inference in adults with high levels of autistic traits or autism. *PLoS Comput. Biol.* 17, e1009006
103. Gershman, S.J. (2019) What does the free energy principle tell us about the brain? *arXiv* 1901.07945
104. Karaminis, T. *et al.* (2017) Binocular rivalry in children on the autism spectrum. *Autism Res.* 10, 1096–1106
105. Skerswetat, J. *et al.* (2022) Visual consciousness dynamics in adults with and without autism. *Sci. Rep.* 12, 4376
106. Dunn, S. and Jones, M. (2020) Binocular rivalry dynamics associated with high levels of self-reported autistic traits suggest an imbalance of cortical excitation and inhibition. *Behav. Brain Res.* 388, 112603
107. Mentch, J. *et al.* (2019) GABAergic inhibition gates perceptual awareness during binocular rivalry. *J. Neurosci.* 39, 8398–8407
108. Freyberg, J. *et al.* (2015) Reduced perceptual exclusivity during object and grating rivalry in autism. *J. Vis.* 15, 11
109. Robertson, C.E. *et al.* (2013) Slower rate of binocular rivalry in autism. *J. Neurosci.* 33, 16983–169891
110. Hadad, B.S. and Schwartz, S. (2019) Perception in autism does not adhere to Weber's law. *eLife* 8, e42223