Commentary

Biological Psychiatry

The Bayesian Savant

Karl Friston

Computational psychiatry promises a fresh and formal approach to mental health, and autism has become its so-called poster child. Key concepts from computational neuroscience are now finding their way into discussions about the pathophysiology and psychopathology of autism spectrum disorder (ASD) (1–4). This is exemplified beautifully by Sevgi *et al.* (5), who report that "higher autistic traits in healthy subjects are related to lower scores in a learning task that requires social cue integration." Careful Bayesian modeling of this learning suggests that trait-related differences are not explained by a failure to process social stimuli per se, but rather by the extent to which participants afford precision to—or attend—social cues. So why is it important? For people unfamiliar with things like the Bayesian brain and precision, we start with a brief review of the ideas that motivated Sevgi *et al.* (5).

The Bayesian Brain and Autism

The story starts with a compelling heuristic (1) suggesting that the problem in ASD is a failure to integrate sensory evidence with prior beliefs about the causes of sensations. To talk about psychopathology in these terms required a theoretical framework that can accommodate beliefs, namely, the Bayesian brain. In this setting, the brain becomes a statistical organ that generates hypotheses and predictions that are tested against sensory evidence. This perspective can be traced back to Helmholtz and the notion of unconscious inference and how these inferences induce beliefs and behavior.

Predictive Coding

Modern versions of Helmholtz's notion usually appeal to predictive coding. Predictive coding describes how the brain processes sensory information as optimizing explanations for its sensations. In this scheme, neuronal representations in higher levels of cortical hierarchies generate predictions of representations in lower levels. These top-down predictions are compared with representations at lower levels to form prediction errors (associated with the activity of superficial pyramidal cells). The ensuing mismatch is passed back up the hierarchy to update higher representations (associated with the activity of deep pyramidal cells). This recursive exchange of signals suppresses prediction error at every level to provide deep hierarchical explanations for sensory input at the lowest level. Computationally, neuronal activity is thought to encode beliefs about states of the world that cause sensations (e.g., my visual sensations are caused by a dog). The simplest encoding corresponds to the expected value of a (hidden) cause or expectation. These causes are referred to as "hidden" because they have to be inferred from their sensory consequences. In short, predictive coding represents a biologically plausible scheme for updating beliefs about the world using sensory samples (Figure 1).

How Precise Are Predictions?

Predictive coding provides a compelling explanation for several aspects of functional anatomy and perception. However, simply predicting the content of our sensations is only half the story; we also have to predict the confidence or precision that should be ascribed to sensory information. This represents a subtle but important problem for the brain. Heuristically, one can regard ascending prediction errors as broadcasting newsworthy information that has yet to be explained by descending predictions. However, the brain also has to select the channels it attends to by adjusting the volume of competing channels. Neurophysiologically, this corresponds to adjusting the gain of prediction errors that compete to update expectations. The boosting or precision weighting of prediction errors is thought to be mediated by neuromodulatory mechanisms or synaptic gain control. This has been associated with attentional gain control in sensory processing and also has been discussed in terms of affordance and action selection. Crucially, the delicate balance of precision—over hierarchical levels—has a profound effect on inference and may hold the key to understanding false inference in autism (3).

Precision and Autism

So how does this help us understand autism? At its simplest, the explanation rests on an imbalance between sensory and prior precision, where prior precision refers to the precision of prediction errors (and subsequent representations) at high levels of the hierarchy. This can be construed either as an inability to ignore sensory information or as holding imprecise prior beliefs, therefore precluding deeply structured explanations for the sensorium. This explains the loss of central coherence and a pathologic tendency to engage with the sensory world (6). But what causes this state of affairs?

More detailed developmental accounts call on a number of concepts in predictive coding, such as active inference, sensory attenuation, and agency. Active inference explains action in terms of minimizing (proprioceptive and interoceptive) prediction errors, not through adjusting representations but by engaging (motor and autonomic) reflexes. In brief, reflexes fulfill top-down predictions about the consequences of action. This applies to both motor control (through minimizing proprioceptive prediction errors) and autonomic function (through minimizing interoceptive prediction errors).

Sensory attenuation refers to the attenuation of sensory precision that is necessary to suspend attention to sensory evidence that contradicts the predicted consequences of an

SEE CORRESPONDING ARTICLE ON PAGE 112

http://dx.doi.org/10.1016/j.biopsych.2016.05.006
ISSN: 0006-3223

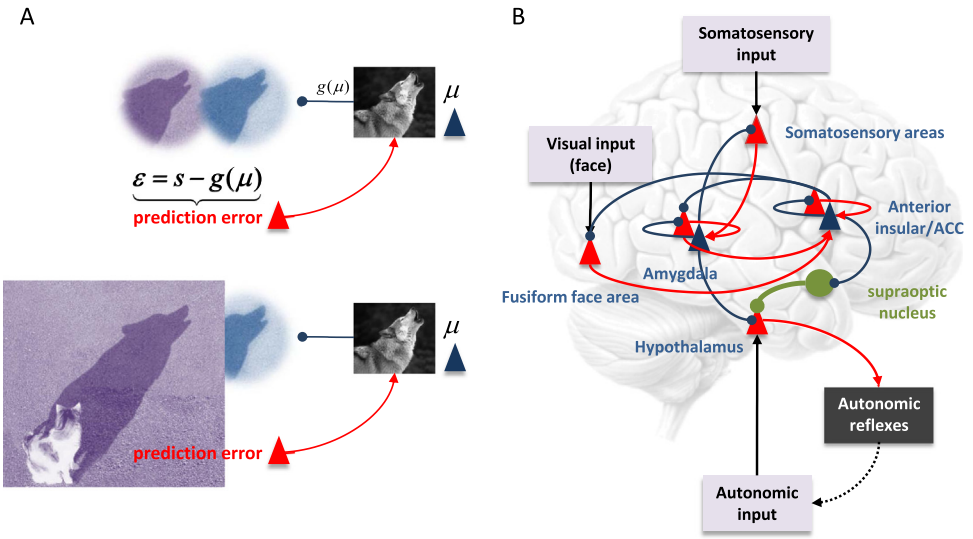

Figure 1. **(A)** Predictive coding and perceptual inference. Predictive coding deals with the problem of inferring the causes of (generally sparse and ambiguous) sensory inputs. This is illustrated in the upper panel with a shadow that can be regarded as a sensory impression. A plausible explanation for this sensory input could be a howling dog. Predictive coding assumes that the brain has a model that generates predictions of sensory input given a hypothesis or expectation about how that input was caused. Here, the expectation is denoted by μ (e.g., a dog), and the sensory prediction generated by the model is summarized with $g(\mu)$. The prediction error is the difference between the input and the prediction. This prediction error then is used to update or revise the expectation, until errors are minimized. At this point, the expectation provides the best explanation or inference for the causes of sensations. Note that this inference does not have to be veridical. In the lower panel, the actual cause was a cat; however, the beholder may never know the true causes; provided that we minimize our prediction errors consistently, our model of the world will be sufficient to infer plausible causes in the outside world that are hidden behind a veil of sensations. **(B)** Oxytocin and the failure of sensory attenuation. This schematic describes (simplified) neural architectures underlying the predictive coding of visual, somatosensory, and autonomic signals. The anatomic designations should not be taken too seriously; they are just used to illustrate how predictive coding can be mapped onto neuronal systems. Red triangles correspond to neuronal populations (superficial pyramidal cells) encoding prediction error, whereas blue triangles represent populations (deep pyramidal cells) encoding expectations. These populations provide descending predictions to prediction error populations in lower hierarchical levels (blue connections). The prediction error populations then reciprocate ascending prediction errors to adjust the expectations (red connections). Arrows denote excitatory connections, whereas circles denote inhibitory effects (mediated by inhibitory interneurons). These recurrent connections mediate innate (epigenetically specified) reflexes, such as the suckling reflex, that elicit autonomic (e.g., vasovagal) reflexes in response to appropriate somatosensory input. These reflexes depend on high-level representations predicting both somatosensory input and interoceptive consequences. The representations are activated by somatosensory prediction errors and send interoceptive predictions to the hypothalamic area to elicit interoceptive prediction errors that are resolved in the periphery by autonomic reflexes. Oxytocin (in green) is shown to project to the hypothalamic area to modulate the gain or precision of interoceptive prediction error units. One hypothesis for autism rests on a failure to attenuate the precision of autonomic prediction errors, thereby precluding expectations about visual and somatosensory information (e.g., a mother's face or affiliative touch) that is not accompanied by autonomic input. ACC, anterior cingulate cortex.

intended action. Furthermore, the attenuation of prediction errors that elicit reflexes enables exteroceptive predictions to be repurposed to infer the intentional and interoceptive states of others. In the absence of sensory attenuation, this repurposing results in echopraxia or interoceptive (emotional) contagion. That is, sensory attenuation is crucial for voluntary and involuntary action, and action observation. If the basic problem in autism is unduly precise sensory precision (i.e., a failure of sensory attenuation), what would this look like developmentally?

Imagine a neuromodulatory deficit (e.g., mediated by subtle changes in the synaptic effects of oxytocin) that precluded the attenuation of interoceptive prediction errors. Not only would this render autistic infants unduly sensitive to interoceptive cues (i.e., autonomic hypersensitivity), but also it would have profound implications for a sense of agency and the distinction between self and other (i.e., theory of mind). This follows from the inability to disengage interoceptive inference during affiliative interactions with [m]others. That is, the autistic infant would be unable to ignore the absence of interoceptive signals induced by maternal nurturing (e.g., breastfeeding) during purely affiliative and prosocial interactions with [m]others. In short, the autistic infant would never realize that the nurturing and prosocial [m]other were the same hidden cause or person (7).

One can see how this fundamental failure to learn the causal structure of a prosocial world could lead to impoverished and imprecise models of interpersonal interactions, and the causes of bodily sensations. In this light, the findings of Sevgi *et al.* (5) speak to the specificity of false inference in ASD, namely, an inability to elaborate precise predictions in an interpersonal setting. Furthermore, their results speak to a failure to contextualize or attend to social cues (via a failure to predict sensory precision). This account raises many interesting questions about the roles of interoception in the development of social cognition and the relationship between alexithymia and autism (8).

Aberrant Precision and Other Theories

The predictive coding account of autism is not the only computational option. Last year, a group of computational neuroscientists met to consider three dominant paradigms (see Acknowledgments and Disclosures). In addition to aberrant precision, we considered the pruning hypothesis (9) and the low-noise hypothesis (10).

The pruning hypothesis accounts for developmental phenotypes within ASD (early-onset, late-onset, and regressive-recovering phenotypes). It posits an initial formation of exuberant neuronal connections that is followed by a period of synaptic

pruning. This process has been modeled in supervised neural networks (that learned the past tense of English). The basic idea is that pruning is too severe in ASD, leading to behavioral deficits, followed by some recovery as the system self-organizes. Alternatively, Davis and Plaisted-Grant (10) compare accounts of ASD based on opposing assumptions about high and low levels of endogenous neuronal noise. They argue that low levels explain some of the psychophysical characteristics of ASD, such as enhanced perceptual discrimination. Crucially, these performance enhancements come at a cost: this follows from the fact that a degree of imprecision (endogenous noise) is necessary to preclude perseveration.

The exciting thing about all three theories is that they rest on precision (as a computational construct) and synaptic gain or efficacy (as a physiological construct). For example, in predictive coding, synaptic pruning depends on the precision encoded by synaptic gain and is construed as a form of Bayesian model selection. Low prior precision therefore renders synaptic connections or associations (at higher hierarchical levels) more vulnerable to pruning. The low endogenous noise hypothesis is exactly congruent with a high sensory precision. This is easy to demonstrate by formulating gain in terms of the sensitivity of neuronal firing rates to depolarization at the level of neuronal populations (using something called the Fokker Planck equation). This means that low sensory noise corresponds to high sensory precision. Interestingly, fundamental statistical imperatives (e.g., Occam's razor) speak to the optimal attenuation of precision to ensure parsimonious and accurate explanations of sensory data. These points of contact illustrate the discourse that is enabled by a formal approach, and computationally informed studies of the sort offered by Sevgi *et al.* (5).

Acknowledgments and Disclosures

I thank the following people, whose correspondence and ideas formed the basis of this commentary: Geoff Bird, David Burr, Greg Davis, Chris Frith, Teodora Gliga, Mark Johnson, Rebecca Lawson, Jay McClelland, Andrew Oliver, Liz Pellicano, Lisa Quattrocki Knight, and Michael Thomas.

KF is supported by the Wellcome Trust (Principal Research Fellowship 088130/Z/09/Z). He reported no biomedical financial interests or potential conflicts of interest.

Article Information

From the Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, London, United Kingdom.

Address correspondence to Karl Friston, Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, London WC1N 3BG, United Kingdom; E-mail: k.friston@ucl.ac.uk.

Received May 10, 2016; accepted May 10, 2016.

References

1. Pellicano E, Burr D (2012): When the world becomes too real: a Bayesian explanation of autistic perception. Trends Cogn Sci 16: 504–510.
2. Van de Cruys S, Evers K, Van der Hallen R, Van Eylen L, Boets B, de-Wit L, *et al*. (2014): Precise minds in uncertain worlds: predictive coding in autism. Psychol Rev 121(4):649–675.
3. Lawson RP, Rees G, Friston K (2014): An aberrant precision account of autism. Front Hum Neurosci 8:302.
4. Palmer CJ, Seth AK, Hohwy J (2015): The felt presence of other minds: Predictive processing, counterfactual predictions, and mentalising in autism. Conscious Cogn 36:376–389.
5. Sevgi M, Diaconescu AO, Tittgemeyer M, Schilbach L (2016): Social Bayes: Using Bayesian modeling to study autistic trait–related differences in social cognition. Biol Psychiatry 80:112–119.
6. Happé F, Frith U (2006): The weak coherence account: detail focused cognitive style in autism spectrum disorders. J Autism Dev Disord 36:5–25.
7. Quattrocki E, Friston K (2014): Autism, oxytocin and interoception. Neurosci Biobehav Rev 47:410–430.
8. Brewer R, Happé F, Cook R, Bird G (2015): Commentary on "Autism, oxytocin and interoception": Alexithymia, not autism spectrum disorders, is the consequence of interoceptive failure. Neurosci Biobehav Rev 56:348–353.
9. Thomas MSC, Davis R, Karmiloff-Smith A, Knowland VCP, Charman T (2016): The over-pruning hypothesis of autism. Dev Sci 19(2):284–305.
10. Davis G, Plaisted-Grant K (2015): Low endogenous neural noise in autism. Autism 19(3):351–362.