The Bayesian savant: a commentary on Sevgi et al

By Karl Friston

***Correspondence**: Karl Friston*
Wellcome Trust Centre for Neuroimaging, Institute of Neurology,
Institute of Neurology, UCL
London. WC1N 3BG UK
Email: k.friston@ucl.ac.uk

Keywords: autism; computational psychiatry; precision; predictive coding; Bayesian; alexithymia; interoception.

***Word Count**: 1509 (excluding references and figure legend)*

Introduction

Computational psychiatry promises a fresh and formal approach to mental health – and Autism has become its 'poster child', generating new questions and debates (Van de Cruys et al 2013). Key concepts from computational neuroscience are now finding their way into discussions about the pathophysiology and psychopathology of autistic spectrum disorder (ASD). This is beautifully exemplified by Sevgi et al (2016), 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 (2016).

The Bayesian brain and autism

The story starts with a compelling heuristic (Pellicano and Burr 2012) 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 or fantasies that are tested against sensory evidence. This perspective can be traced back to Helmholtz and the notion of unconscious inference (Helmholtz, 1866/1962) 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 a hierarchical explanation for sensory input to 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 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. See 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 prediction errors. This represents a subtle but important problem for the brain, whose solution may rest on modulating the gain or excitability of neuronal populations reporting prediction error (Lawson et al 2014). 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 listens 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 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 (Lawson et al., 2014).

Precision and autism

So how does this help 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 either be construed as overly precise sensory information or imprecise prior beliefs, reflecting an incoherent central or deeply structured explanation for the sensorium. This explains the loss of central coherence and a pathological tendency to engage with the sensory world (Happé and Frith, 2006). But how does this state of affairs arise?

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 through minimizing (proprioceptive and interoceptive) prediction errors, not through adjusting representations, but by engaging (motor and autonomic) reflexes. In brief, reflexes enact top-down predictions that fulfill expectations about the active sampling of the environment. 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 top-down predictions of movement. Furthermore, the attenuation of descending prediction errors (that elicit reflexes) enables hierarchical predictions to be 'repurposed' to infer the intentional and interoceptive states of others – without echopraxia or interoceptive (emotional) contagion. In other words, sensory attenuation is crucial for voluntary and involuntary action – and action observation. Recall that the basic problem associated with autism is unduly precise sensory precision (i.e., a failure of sensory attenuation). So 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 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 autonomic reflexes during affiliative interactions with [m]others. In other words, the autistic infant would be unable to suspend autonomic reflexes during prosocial exchanges and never learn that there is a difference between [m]other as part of an 'extended' self (e.g. during breastfeeding) and [m]others as distinct from self (Quattrocki and Friston 2014).

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 (2016) 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. (Brewer et al., 2015).

Aberrant precision and other theories

The predictive coding account of autism is not the only computational game in town. Last year, a group of computational neuroscientists met to consider three dominant paradigms (see acknowledgements): In addition to aberrant precision, we considered the *pruning hypothesis* (Thomas et al., 2015) and the *low-noise hypothesis* (Davis and Plaisted-Grant, 2014).

The pruning hypothesis accounts for developmental phenotypes within ASD (early-onset, late-onset, and regressive-recovering phenotypes). It posits that and initial exuberant formation of neuronal connections is followed by a period of synaptic pruning. This process has been modeled in supervised neural networks (that learnt the past tense of English). The basic idea is pruning is too aggressive in ASD, leading to behavioral deficits, followed by some recovery as the system self organizes. Alternatively, Davis and Plaisted-Grant (2014) 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 would therefore render 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 changes in dispersion at the level of neuronal populations (using something called the Fokker Planck equation). This means 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 point of contact illustrate the discourse that is enabled by a formal approach – and computationally informed studies of the sort offered by Sevgi et al.

Acknowledgements: KF is funded by the Wellcome Trust (Principal Research Fellowship 088130/Z/09/Z). I would like to 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, Michael Thomas, Andrew Oliver, Liz Pellicano, and Lisa Quattrocki Knight.

References

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*.56:348-53.

Davis G, Plaisted-Grant K (2014): Low endogenous neural noise in autism. *Autism*. Published online. DOI: 10.1177/1362361314552198.

Happé F, Frith U (2006): The weak coherence account: detail focused cognitive style in autism spectrum disorders. *J. Autism Dev. Disord*., vol. 36, p. 5–25.

Helmholtz H (1866/1962): Concerning the perceptions in general. In: *Treatise on physiological optics*, vol. III New York: Dover.

Lawson RP, Rees G, Friston K (2014): An aberrant precision account of autism. *Front Hum Neurosci*. 8:302.

Pellicano E, Burr D (2012): When the world becomes too real: a Bayesian explanation of autistic perception. *Trends in Cognitive Sciences*, vol. 16, pp. 504-510.

Quattrocki E, Friston K (2014): Autism, oxytocin and interoception. *Neurosci Biobehav Rev*. 47:410-30.

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* XX:XXX-XXX.

Thomas MSC, Davis R, Karmiloff-Smith A, Knowland VCP, Charman T (2015): The over-pruning hypothesis of autism. *Developmental Science*. Article first published online: DOI: 10.1111/desc.12303.

Van de Cruys S, de-Wit L, Evers K, Boets B, Wagemans, J (2013): Weak priors versus overfitting of predictions in autism: Reply to Pellicano and Burr (TICS, 2012). *Percept.* 4, 95.

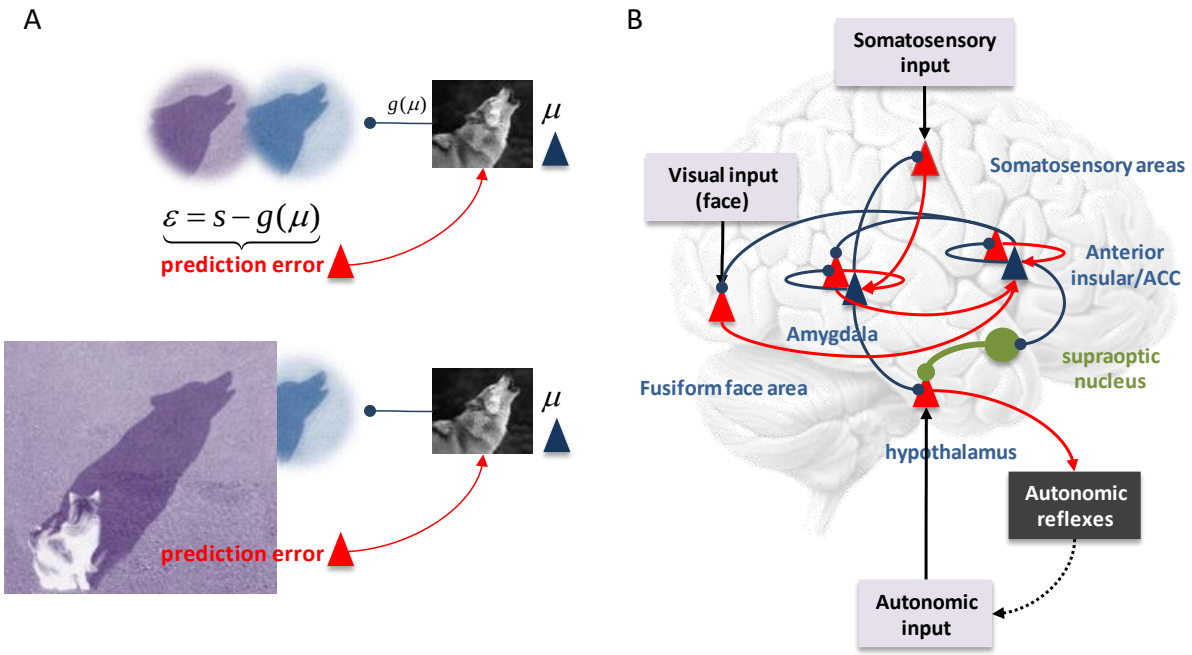

Figure Legend

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 canine. 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 is then used to update or revise the expectation, until prediction error is 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 of sensations 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 simple, somatosensory and autonomic signals. The anatomical 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, while 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 and). Arrows denote excitatory connections, while 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 upon high-level representations predicting both the 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 prediction error units. One hypothesis for autism rests on a failure to attenuate the precision of autonomic prediction errors; thereby precluding inference about somatosensory and visual information (e.g., a mother's face or affiliative touch) that does not elicit autonomic reflexes.