BRAIN
A JOURNAL OF NEUROLOGY

BOOK REVIEW
The fantastic organ

The Age of Insight by Eric Kandel is a brutal and beautiful book. Its pages burgeon with beautiful images and beautiful ideas. The ideas are pursued with relentless honesty and diligence for more than 600 pages. There are parts of this book that should only be read in private. I know this after a slightly fraught flight to London, trying to read the book while covering its pictures of masturbating women with my boarding card—I did not want to offend the sensibilities of an Indian gentleman, who showed a polite interest in my reading material. I am not quite sure why I did this, perhaps because I am English; or perhaps because I found the pictures curiously arousing—despite their innocent titles {e.g. 'Seated Woman in Armchair' by Gustav Klimt [(1862–1918): c.1913], pencil and white chalk}. Happily, the book explained why I found them so alluring.

Kandel's treatment is a book of two parts. The first deals with innovative thinking and insights in science and art in turn of the century Vienna. The second overviews the neurobiology of perception and emotion, bringing the reader back to deep questions about perspective taking and neuroaesthetics. The premise that underpins both parts is that the brain is an inference machine, generating hypotheses and fantasies that are tested against sensory data. Put simply, the brain is—literally—a fantastic organ (fantastic: from Greek *phantastikos*, able to create mental images, from *phantazesthai*). For me, the story starts with Hermann von Helmholtz (1821–94) and the notion of unconscious inference (Helmholtz, 1866/1962). Kandel places this story in the context of history and art, in an illuminating and compelling way.

In summarizing the context that art provides for understanding functional brain architectures (and vice versa) Kandel concludes

> 'we now know that one of the main reasons expressionist art appeals to us so strongly is that we have evolved a remarkably large, social brain. Moreover, the brain's mirror neuron systems, theory of mind system and biological modulators of emotions and empathy endow us with a great capacity for understanding other people's minds and emotions.'

The early part of the book is concerned with the creative achievements of people like Oskar Kokoschka [(1886–1980): see Fig. 1] and Egon Schiele (1890–1918), whose expressionism engages unconscious inference, while Gustav Klimt's

> 'intuitive grasp of the power of implied line, contour and top-down processing enabled him to create some of the most subtle and sensual
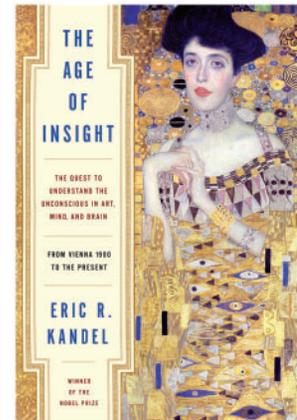
works in the history of modern art. With these new insights into the unconscious empathic, emotional and perceptual apparatuses of the brain, the Austrian modernists were indeed cognitive psychologists in their own right. In parallel with Sigmund Freud (1856–1939), they knew how to enter the private theatre of another's mind, to understand its nature, mood and emotion, and to convey that understanding to the viewer.' (p. 500).

The fact that art affords insights into functional anatomy was not new to me: my first brain imaging experiment (Lueck *et al*., 1989)—in which Semir Zeki established the colour centre in the human brain—was inspired by Piet Mondrian (1872–1944). Mondrian articulated a fundamental truth about perceptual inference and how it depends on causal structure in the sensorium (p. 498):

> 'For there are made laws, discovered laws, but also laws–a truth for all time. These are more or less hidden in the reality which surrounds us and do not change. Not only science but art also, shows us that reality, at first incomprehensible, gradually reveals itself, by the mutual relations that are inherent in things.'

In other words, there is causal structure in our world that the brain distils and embodies in its inferential machinery. Modern versions of Helmholtz's ideas are now among the most popular explanations for message passing in the brain—generally cast in terms of the Bayesian brain hypothesis or predictive coding. These are not abstract or hand waving schemes; nearly everywhere one

Figure 1 (*Left*) Oskar Kokoschka, 'Self-portrait as warrior' (1909). Unfired clay painted with tempera. This self-portrait is remarkably similar to the self-portrait of Sigmund Freud's grandson (*Right*) painted nearly a century later: Lucian Freud (1922–2011), 'Reflection (Self-portrait)', (1985). Lucian Freud was widely considered the pre-eminent British artist of his time. His early work is often associated with (German) Expressionism; although interestingly, this was an influence he tended to deny.

looks, the anatomical and physiological evidence points towards predictive coding as the organizing principle for cortical microcircuits and hierarchical brain systems: see Bastos *et al.* (2012) for a recent review of canonical microcircuits and predictive coding in sensory systems, and Adams *et al.* (2012) for a related treatment of motor systems. In these schemes, neuronal representations in higher levels of sensory cortical hierarchies entail hypotheses that provide predictions for lower levels. These top-down predictions are compared with representations at the lower level to form a prediction error (usually associated with the activity of superficial pyramidal cells). This prediction error is then passed back up the hierarchy, to change higher representations (usually associated with the activity of deep pyramidal cells). These changes provide better predictions and thereby reduce prediction error at each and every level of the hierarchy.

Predictive coding rests upon hierarchical (generative) models of how sensory inputs are caused and instantiate Helmholtzian inference in a Bayes optimal fashion. The intellectual pedigree of these ideas can be traced directly from Helmholtz to Viennese expressionism through Sigmund Freud:

'Perhaps the most direct influence on Freud was the thinking of Hermann von Helmholtz...one of the most remarkable scientists of the 19th century, Helmholtz helped bring physiology together with physics and chemistry. In his work on visual perception, Helmholtz came to see psychology as fundamental to an understanding of brain physiology.' (p. 59).

The implications of Helmholtzian inference for perception, and particularly the perception of art, are clear and deconstructed nicely in Kandel's discussion of 'the beholder's share' (Fig. 2).

'The insight that the beholder's perception involves a top-down inference convinced Gombrich that there is no 'innocent eye': that is, all visual perception is based on classifying concepts and interpreting visual information. One cannot perceive that which one cannot classify.' (p. 204). 'He appreciated the role of cognitive schemata, or internal

representations of the visual world in the brain, arguing that every painting owes more to other paintings the viewer has seen than it does to the world actually being portrayed.' (p. 212)

Kandel further emphasizes the role of history and culture in shaping an otherwise solipsistic mental world of hypotheses—particularly through the writings of Alios Riegl (1858–1905):

'In his emphasis on the historical context in which all art emerged and importance of the beholder's participation for the completion of a painting, Riegl stripped art of its pretension to achieve a universal truth.' (p. 104).

Riegl was a member of the Venetian School of Art History and figures prominently in Kandel's treatment of the beholder's contribution:

'Just as the artist creates a work of art, so the viewer recreates it by responding to its inherent ambiguity. The extent of the beholder's contribution depends upon the degree of ambiguity in the work of art.' (p. 192).

Ambiguity (or perhaps its resolution) gets to the heart of perceptual inference: there is no point—or pleasure—in making statistical inferences about sure bets. The *raison d'être* for inference is to disambiguate among plausible and competing hypotheses.

Kandel's subsequent neurobiological overview of visual perception follows the usual themes in a balanced way, carefully nuanced to emphasize the role of generative models and mental images:

'Influenced in part by Freud's intuitive insights and Kris and Gombrich's systematic attempts, neuroscientists have begun a more rigorous, cellular analysis of human sensory systems. Specifically, Richard Gregory and David Marr began to analyse human perception in terms of bottom-up Gestalt psychology and top-down hypothesis testing and information processing. By confirming on a cellular level that our sensory systems are creative—that they generate hypotheses about what faces, facial expressions, hand positions and bodily movements are important and what distinguishes biological from nonbiological motion—these neuroscientists have taken us backstage into the private theatre of the mind.' (p. 303)

However, to my mind, Kandel lets the side down a little here by over-emphasizing the contributions of (bottom-up) people like David Marr (1945–80), who cast perceptual synthesis as the serial construction or extraction of visual information—invoking notions of primal sketches and other heuristics that I (perhaps deliberately) keep forgetting. From my perspective, this bottom-up era represents an ugly hiatus in—what should have been—a seamless progression of Helmholtz's ideas to our current appreciation of the brain as an inference machine (Dayan *et al.*, 1995)—an organ that continually generates predictions and hypotheses in a top-down fashion. Although there are hints at an underlying generative model in Marrian formulations, the quintessentially predictive architecture of the visual hierarchy actually goes all the way down the eyes. For example, the number of backward connections from the visual cortex to the lateral geniculate far exceeds the number of forward connections (Sillito and Jones, 2002). Furthermore, the notion of predictive coding in the brain was

Figure 2 Giuseppe Arcimboldo, The Vegetable Gardener (c.1590). Oil on panel. Our percepts are constrained by what we expect to see and the portfolio of hypotheses that can be called upon to explain sensory impressions (Gregory, 1968). As Ernst Gombrich (1909–2001) predicted, only specific (predictable) stimuli can 'pick the locks on the neural gateways to vision.' (p. 254). Guiseppe Arcimboldo (1527–93), 'a 16th century Milanese artist who was a favourite of the Viennese, illustrates this dramatically by using fruits and vegetables to create faces in his paintings. When viewed right side up, the paintings are readily recognisable faces' (p. 288).

first introduced as an explanation for retinal processing (Srinivasan et al., 1982). It did strike me that the detour from unconscious inference (Helmholtz, 1866/1962) to conscious inference (as considered by Kandel) was based on the seminal work of several visual neuroscientists from the eastern seaboard of America—where Kandel has worked since 1952.

In the denouement of his book Kandel considers some of the more challenging aspects of the inferential brain; in particular, how we make inferences about other people, ourselves and our emotional states. He cleverly conflates the mirror neuron system with reflections in a mirror. This captures the essence of 'perspective taking', which is unpacked in terms of second order representations (representations of representations) as they relate to theory of mind and how artists use reflections (Fig. 3). These high order aspects of inference in the brain represent the frontiers of theoretical neurobiology—frontiers that may hold the key for a systemic understanding of several neuropsychiatric syndromes.

It is self evident that if our brains entail generative models of our world, then much of the brain must be devoted to modelling entities that populate our world; namely, other people. In other words, we spend much of our time generating hypotheses and predictions about the behaviour of people—including ourselves. As noted by Kandel 'the brain also needs a model of itself' (p. 406). This places the mirror neuron system centre stage in generating both proprioceptive and exteroceptive predictions about how you (and I) will behave. To appreciate fully the bilateral nature of predictions provided by the mirror neuron system, we have to take unconscious inference to the next level and consider it in an embodied context. Put simply, one can regard perception as the suppression of exteroceptive prediction errors by selecting predictions that are best able to explain sensations. However, exactly

the same argument can be applied to action that minimizes proprioceptive prediction errors via the classical reflex arcs. In other words, we can reduce prediction errors in one of two ways: we can either change predictions so that they match (exteroceptive) sensory samples, or we can change the samples through action, to make them match (proprioceptive) predictions. This is active inference. So what has this got to do with mirror neurons?

If mirror neurons provide top-down predictions of both the kinaesthetic (proprioceptive) and exteroceptive consequences of moving—and thereby cause movements through motor reflexes, then they provide a ready-made set of hypotheses for inferring the motor intentions of other people. This is because the exteroceptive (e.g. visual) consequences of movements are the same and all we have to do is to suppress the proprioceptive predictions. This provides a nice perspective on why mirror neurons respond both to self-made acts and during action observation (Kilner et al., 2007). More generally:

'just as the visual brain constructs models of reality from figural primitives, so our social brain is innately wired to function as a psychologist, forming models of other people's motivations, desires and thoughts.' (p. 406)

However, there is an important twist here. To harness the mirror neuron system during action observation, we have to suppress proprioceptive prediction errors that would otherwise elicit movements and cause us to mimic (mirror) the subject of our observation. This suppression rests on (mathematically speaking) reducing the precision of—or confidence in—proprioceptive prediction errors. This speaks to fundamental aspect of inference in the brain; namely the encoding of precision or confidence through neuromodulation. In other words, not only do we have to infer
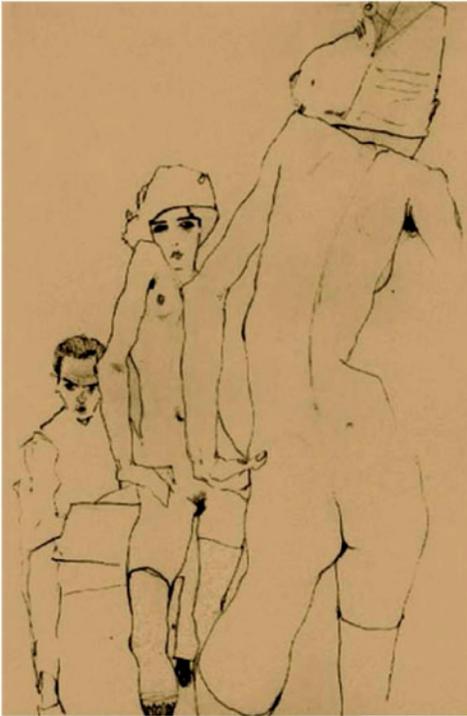
Figure 3 Egon Schiele, 'Nude in front of the mirror' (1910). Pencil on paper. 'By covering parts of the body that are not usually considered necessary to cover, Schiele emphasises those that are uncovered. Moreover, the models posture is a perfect caricature: it exaggerates one aspect a woman's body that is inherently erotic, her hip. The hips of Schiele's model are brilliantly drawn. By placing her before a mirror he is also, indirectly, drawing her nude body from the front and depicting himself sketching her. Seen in the mirror, it is unclear whether the woman is unself-consciously posing for the artist or whether she is engaging in seduction. Schiele's gaze is voyeuristic but the model seems oblivious to his intensity and playfully poses while looking at herself in the mirror. Schiele uses a mirror to express his interest in the direct and indirect image, the outward appearance and the private theatre of the mind, the demure and the sensual'. (p. 402).

the *content* of our sensorium but also the *context*, in terms of the precision or certainty about the content. This represents a subtle but ubiquitous problem that the brain has to solve—and the solution rests on modulating the gain or post-synaptic sensitivity of neuronal populations encoding prediction error.

I have deliberately taken this essay slightly beyond Kandel's synthesis by introducing neuromodulation as the neuronal basis of precision (the encoding or representation of uncertainty). I do this because it ties inference in the brain to synaptic processes that may be compromised in syndromes like schizophrenia, Parkinson's disease, autism, hysterical disorders, and so on. The basic idea, which is gaining increasing traction in the literature, is that many disorders of active inference can be understood as a failure of neuromodulation. A recent theoretical paper in *Brain* (Edwards *et al*., 2012) provides a nice example of this: it describes how one can understand functional (hysterical) symptoms as aberrant inference that follows from a failure to optimize precision (dopaminergic

neuromodulation). Wherever one looks in the theoretical literature, the same theme is emerging: from false inference as an explanation for the positive symptoms (hallucinations and delusions) of schizophrenia (Fletcher and Frith, 2009), to the loss of central coherence in autism (Pellicano and Burr, 2012). Many of these formulations rest on understanding behaviour and action as part of the (active) inferential process that underlies Kandel's premise. So can we explain emotion with precision and active inference?

Perhaps the most prescient challenge to formal descriptions of the brain as an inference machine is how one can accommodate emotions, self-awareness and disorders thereof. There are already exciting ideas in the recent literature that provide a simple perspective on emotional processing from an inferential (predictive coding) perspective; e.g. Seth *et al*. (2011). The basic idea is as follows: recall from above that a simple explanation for motor behaviour appeals to motor reflexes that cancel proprioceptive prediction errors. These prediction errors are formed by comparing primary afferent input from stretch receptors with descending proprioceptive predictions to alpha motor neurons in the spinal-cord (and cranial nerve nuclei). This view replaces descending motor commands with motor predictions that are fulfilled by peripheral reflexes (Adams *et al*., 2012). The predictions themselves are elaborated on the basis of deep hierarchical inference about states of the world, including the trajectories of our own bodies. Exactly the same mechanism can be applied to interoceptive signals. This means that the internal milieu is controlled by autonomic reflexes that transcribe descending interoceptive predictions into a physiological homoeostasis. As with the mirror neuron system (and sensorimotor representations in general), these interoceptive predictions are just one—among many—of multimodal predictions that emanate from high-level hypotheses about our embodied state. For example, the best explanation for the myriad of sensory inputs I was experiencing on the aeroplane was my situated state of mind. A situated state that comprises an internally consistent hierarchical model of the world, with multiple levels of description; from the (alluring) visual objects that predict the sensory impressions upon my retinal epithelia to the (oculomotor) proprioceptive predictions producing saccadic eye movements (Fig. 4). Crucially, these hierarchal representations also predict my interoceptive state; including sympathetic and parasympathetic outflow—literally, my gut feelings. In this view, interoceptive information does not cause our self-awareness, or vice versa. There is a circular causality in which neuronal representations cause changes in autonomic status by enslaving autonomic reflexes. At the same time, interoceptive signals entrain hierarchical representations so that they provide the best prediction. Emotional valence is therefore a necessary aspect of any representation in the brain that includes interoceptive predictions. This means that—in terms of the brain's computational anatomy—the influence of *gut feelings* (interoceptive signals) is inherently contextualized by concomitant exteroceptive and proprioceptive input. As Kandel observes:

'As with visual perception, where we have learned that the brain is not a camera but a Homeric storyteller, so with emotion: the brain actively interprets the world using top-down inferences that depend upon context. As James pointed out, feelings do not exist until the brain interprets the cause of the body's physiological signals and assembles an
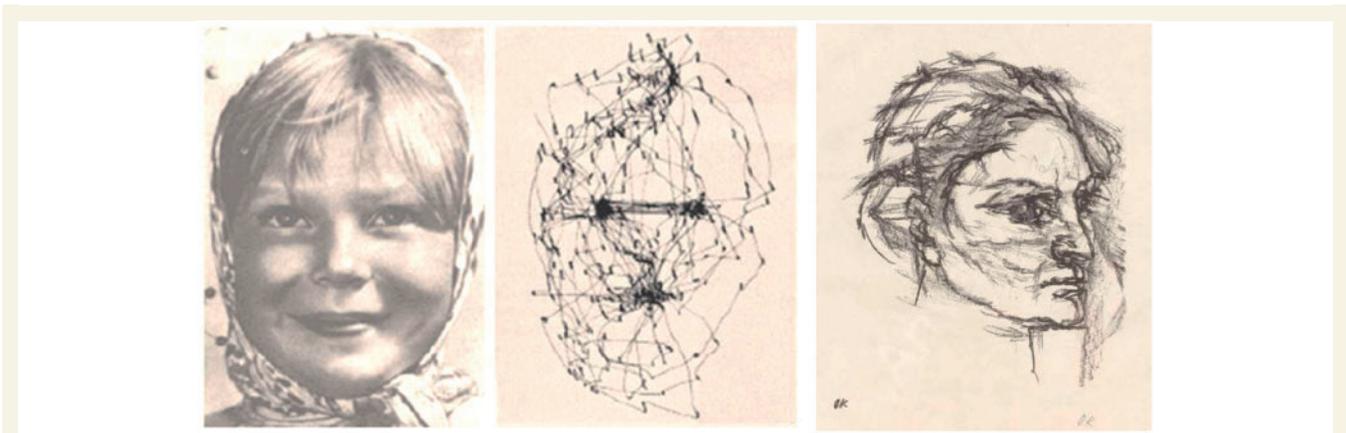
Figure 4 (*Left*) Eye movements elicited by a face stimulus, as studied by the Russian psychophysicist Alfred Yarbus (1914–86) in the 1960s. (*Right*) Oskar Kokoschka. 'Der gefesselte Columbus' (1921). Lithograph. 'Some aspects of Yarbus's findings are echoed in Kokoschka's lithographs. The images look as if the artist were retracing his own eye movements as he observed the subject. Once again, Kokoschka seems to be bringing to the unconscious surface of his art the unconscious processes of his mind – in this case, the mechanisms by which the eye actively explores and interprets the physical and human world, particularly the face.' (p. 339).

appropriate, creative response that is consistent with our expectations and the immediate context.' (p. 351)

Extending the active inference framework to include autonomic reflexes and interoceptive predictions raises a whole series of interesting questions. For example, what role do neuromodulators like dopamine and oxytocin play in nuancing the precision of interoceptive prediction errors? What is the relationship between exteroception and interoception during self observation (Ainley *et al*., 2012)? Do von Economo neurons have a privileged role in communicating top-down interoceptive predictions from the insular cortex to the amygdala, and hypothalamic systems (Critchley and Seth, 2012)? The theme of hierarchal inference and circular causality is now also emerging in the context of neuropsychoanalysis (Panksepp and Solms, 2012); can one cast Freudian constructs in terms of conscious and unconscious inference in a biological brain? And what roles do art and neuroaesthetics play in disclosing mental models? Are generative models endowed epigenetically with canonical (Platonic) forms or, as Riegl and Gombrich would have it, engraved by culture and experience? Are we driven to search out—or create—their sensory correlates, as evolution's epistemic engineers?

These questions speak to the very issues that exercised the scientists and artists that Kandel celebrates in his scholarly and thought-provoking book. I hope this conclusion reflects the depth of Kandel's synthesis, his encyclopaedic knowledge and writing style, which is (nearly) as engaging as his iconic laughter.

Funding

Karl J. Friston
Wellcome Principal Fellow and Scientific Director
Wellcome Trust Centre for Neuroimaging, Institute of Neurology, UCL, UK
k.friston@ucl.ac.uk

References

Adams RA, Shipp S, Friston KJ. Predictions not commands: active inference in the motor system. Brain Struct Funct 2012, doi: 10.1007/s00429-012-0475-5. [Epub ahead of print].

Ainley V, Tajadura-Jiménez A, Fotopoulou A, Tsakiris M. Looking into myself: changes in interoceptive sensitivity during mirror self-observation. Psychophysiology 2012; 49: 1504–8.

Bastos AM, Usrey WM, Adams RA, Mangun GR, Fries P, Friston KJ. Canonical microcircuits for predictive coding. Neuron 2012; 76: 695–711.

Critchley H, Seth A. Will studies of macaque insula reveal the neural mechanisms of self-awareness? Neuron 2012; 74: 423–6.

Dayan P, Hinton GE, Neal RM, Zemel RS. The helmholtz machine. Neural Comput 1995; 7: 889–904.

Edwards MJ, Adams RA, Brown H, Pareés I, Friston KJ. A Bayesian account of 'hysteria'. Brain 2012; 135: 3495–512.

Fletcher PC, Frith CD. Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. Nat Rev Neurosci 2009; 10: 48–58.

Gregory RL. Perceptual illusions and brain models. Proc R Soc Lond B 1968; 171: 179–96.

Helmholtz H. Concerning the perceptions in general. In: Treatise on physiological optics. 3rd edn. New York: Dover; 1866/1962.

Kilner JM, Friston KJ, Frith CD. Predictive coding: an account of the mirror neuron system. Cogn Proc 2007; 8: 159–66.

Lueck CJ, Zeki S, Friston KJ, Deiber M, Cope P, Cunningham VJ, et al. The colour centre in the cerebral cortex of man. Nature 1989; 340: 386–9.

Panksepp J, Solms M. What is neuropsychoanalysis? Clinically relevant studies of the minded brain. Trends Cogn Sci 2012; 16: 6–8.

Pellicano E, Burr D. When the world becomes 'too real': a Bayesian explanation of autistic perception. Trends Cogn Sci 2012; 16: 504–10.

Seth AK, Suzuki K, Critchley HD. An interoceptive predictive coding model of conscious presence. Front Psychol 2011; 2: 395.

Sillito AM, Jones HE. Corticothalamic interactions in the transfer of visual information. Philos Trans R Soc Lond B Biol Sci 2002; 357: 1739–52.

Srinivasan MV, Laughlin SB, Dubs A. Predictive coding: a fresh view of inhibition in the retina. Proc R Soc Lond B Biol Sci 1982; 216: 427–59.